LIMDEP

Version 10

Econometric Modeling Guide

by

William H. Greene
Econometric Software, Inc.

Econometric Software, Inc.
15 Gloria Place
Plainview, NY 11803
USA
Tel: +1 516-938-5254
Fax: +1 516-938-2441
Email: sales@limdep.com
Websites: www.limdep.com and www.nlogit.com.

Econometric Software, Australia
215 Excelsior Avenue
Castle Hill, NSW 2154
Australia
Tel: +61 (0)4-1843-3057
Fax: +61 (0)2-9899-6674
Email: hgroup@optusnet.com.au

End-User License Agreement

This is a contract between you and Econometric Software, Inc. The *software product* refers to the computer software and documentation as well as any upgrades, modified versions, copies or supplements supplied by Econometric Software. By installing, downloading, accessing or otherwise using the software product, you agree to be bound by the terms and conditions of this Agreement.

Subject to the terms and conditions of this Agreement, Econometric Software, Inc. grants you a non-assignable, non-transferable license, without the right to sublicense, to use the licensed software and documentation in object-code form only, solely for your internal business, research, or educational purposes.

Copyright, Trademark, and Intellectual Property

This software product is copyrighted by, and all rights are reserved by Econometric Software, Inc. No part of this software product, either the software or the documentation, may be reproduced, distributed, downloaded, stored in a retrieval system, transmitted in any form or by any means, sold or transferred without prior written permission of Econometric Software. You may not, or permit any person, to: (i) modify, adapt, translate, or change the software product; (ii) reverse engineer, decompile, disassemble, or otherwise attempt to discover the source code of the software product; (iii) sublicense, resell, rent, lease, distribute, commercialize, or otherwise transfer rights or usage to the software product; (iv) remove, modify, or obscure any copyright, registered trademark, or other proprietary notices; (v) embed the software product in any third-party applications; or (vi) make the software product, either the software or the documentation, available on any website.

LIMDEP® and *NLOGIT*® are registered trademarks of Econometric Software, Inc. The software product is licensed, not sold. Your possession, installation and use of the software product does not transfer to you any title and intellectual property rights, nor does this license grant you any rights in connection with software product registered trademarks.

Use of the Software Product

You have only the non-exclusive right to use this software product. A single user license is registered to one specific individual as the sole authorized user, and is not for multiple users on one machine or for installation on a network, in a computer laboratory or on a public access computer. For a single user license only, the registered user may install the software on a primary stand alone computer and one home or portable secondary computer for his or her exclusive use. However, the software may not be used on the primary computer by another person while the secondary computer is in use. For a multi-user site license, the specific terms of the site license agreement apply for scope of use and installation.

Limited Warranty

Econometric Software warrants that the software product will perform substantially in accordance with the documentation for a period of ninety (90) days from the date of the original purchase. To make a warranty claim, you must notify Econometric Software in writing within ninety (90) days from the date of the original purchase and return the defective software to Econometric Software. If the software does not perform substantially in accordance with the documentation, the entire liability and your exclusive remedy shall be limited to, at Econometric Software's option, the replacement of the software product or refund of the license fee paid to Econometric Software for the software product. Proof of purchase from an authorized source is required. This limited warranty is void if failure of the software product has resulted from accident, abuse, or misapplication. Some states and jurisdictions do not allow limitations on the duration of an implied warranty, so the above limitation may not apply to you. To the extent permissible, any implied warranties on the software product are limited to ninety (90) days.

Econometric Software does not warrant the performance or results you may obtain by using the software product. To the maximum extent permitted by applicable law, Econometric Software disclaims all other warranties and conditions, either expressed or implied, including, but not limited to, implied warranties of merchantability, fitness for a particular purpose, title, and non-infringement with respect to the software product. This limited warranty gives you specific legal rights. You may have others, which vary from state to state and jurisdiction to jurisdiction.

Limitation of Liability

Under no circumstances will Econometric Software be liable to you or any other person for any indirect, special, incidental, or consequential damages whatsoever (including, without limitation, damages for loss of business profits, business interruption, computer failure or malfunction, loss of business information, or any other pecuniary loss) arising out of the use or inability to use the software product, even if Econometric Software has been advised of the possibility of such damages. In any case, Econometric Software's entire liability under any provision of this agreement shall not exceed the amount paid to Econometric Software for the software product. Some states or jurisdictions do not allow the exclusion or limitation of liability for incidental or consequential damages, so the above limitation may not apply to you.

Table of Contents

E1: Econometric Model Estimation

E1.1 Introduction

The primary function carried out by *LIMDEP* is the estimation of econometric models. The first part of the documentation, the *Reference Guide* describes how to use *LIMDEP* to read a data set, establish the current sample, compute transformations of variables, and carry out other functions that get your data ready to use for estimation purposes. Several important tools, such as the matrix algebra program, scientific calculator and program editor are described there as well. This second part, the *Econometric Modeling Guide*, will describe specific modeling frameworks and instructions to be used for fitting these models. A large part of this documentation is devoted to descriptions of the models, themselves, including mathematical background. However, the presentation is (of necessity) not complete, and users are urged to supplement this documentation with the necessary background material for the models they are using.

The organization of this manual is by estimation framework, not by model command. We have found that users prefer that the program documentation be oriented toward the types of functions they want to perform, not to an alphabetical listing of commands. As such, you will find the arrangement of topics in this manual rather similar to the arrangement of topics in treatises in econometrics, such as Greene (2011). We begin with descriptive statistics in Chapters E2-E4, various linear regression models in Chapters E5-E8, and so on. To some degree, the complexity of the models deepens as this manual proceeds.

E1.2 Econometric Models

This manual is devoted primarily to the methods by which you can use *LIMDEP* to fit equations to data, to test hypotheses about the relationships implied by that estimation process, and to use the models for simulation and computation of useful partial effects. For purposes of documenting the program, we use the term 'model estimation' broadly, to encompass all those functions that involve manipulation of data to produce statistics to summarize the information the data contain. Thus, this manual begins with several chapters about computing descriptive statistics, which one might not normally consider model building. However, as data summaries, for program purposes, we consider these part of the model building functions in *LIMDEP*.

The definition of a 'model' in *LIMDEP* consists of the modeling framework, the statement of the variables in the model, and what role the variables will play in that model. The remainder of this chapter will describe in general terms how to use this format to construct model estimation commands in *LIMDEP*.

E1.3 Model Commands

LIMDEP's model commands all use the same format. The essential parts are as follows:

Model name ; model variables specification
; essential specifications for some models
; optional specifications $

The '**Model name**' designates the modeling framework. In most cases, this defines a broad class of models, such as **POISSON**, which indicates that the command is for one of the twenty or so different models for count data, most of which are extensions of the basic Poisson regression model.

The 'model variables specification' generally defines the dependent and independent variables in a model. In almost all cases, the model will include one or more dependent variables, denoted a Lhs, or 'left hand side' variable in *LIMDEP*'s command structure. Independent variables usually appear on the Rhs, or 'right hand side,' of a model specification. To continue our example, a Poisson model might be specified using

POISSON ; Lhs = patents ; Rhs = one,r_and_d $

which specifies one of the most well known applications of this model in economics. (The variable '*one*' is the constant term. We'll return to this below.) Some 'model' commands will have only one of these two specifications, such as

DSTAT ; Rhs = patents $

which requests descriptive statistics for the variable *patents*. As can be seen here, we use the term 'model command' broadly to indicate analysis of a set of data, whether for description or parameter estimation. Other model commands might have only a Lhs variable, such as

SURVIVAL ; Lhs = failtime $

which requests a nonparametric (life table) analysis of a variable named *failtime*. There are also many other types of variable specifications, such as

; Inst = a set of variable names

which will be used to specify the set of instrumental variables in the **2SLS** or **LIML** command.

Most models can be specified with nothing more than the model name and the identification of the essential variables. But, some models require additional specifications in order to be identified. For example, the specific model you want may be a particular case of a broad class of models and in order to specify it, you must provide the 'essential' specifications. For example, the basic command for survival modeling (with covariates to provide the 'model') would be

SURVIVAL ; Lhs = failtime ; Rhs = one,usehours $

This form of the command is for Cox's proportional hazard model. In order to fit a parametric model, such as the Weibull model, you would use

SURVIVAL ; Lhs = failtime ; Rhs = one,usehours
; Model = Weibull $

Note the last specification. This is the only way to specify a Weibull model, so for this model, this specification is essential. The Weibull model is requested as a type of survival model by this command. Obviously, not all models have mandatory specifications – the examples above do not. But, many do. The documentation in the chapters to follow will identify these.

Finally, all model frameworks have options which either extend the model itself or control how the model is estimated or how the estimation results are displayed. For example, the following fits a linear regression model and requests a robust estimator of the covariance matrix of the estimates:

REGRESS ; Lhs = profit ; Rhs = one,sales
; Heteroscedasticity consistent $

The latter specification does not change the model specification, it requests an additional computation, the White heteroscedasticity consistent estimator. For another example,

REGRESS ; Lhs = profit ; Rhs = one,sales
; Plot residuals $

fits a linear model and then plots the residuals. If the latter specification is omitted, the residuals will not be plotted.

In writing commands, there is a shortcut you may use either to shorten your commands or, in other cases, to include documentation in your commands. The form is as follows: Certain specifications are simply 'switches' in commands. Thus, in the two examples immediately above, the optional specifications merely request certain computations – the switch is 'off' until the specification turns it on. In specifications such as these, only the first three or more characters are sufficient to make the switch unique. Thus, the two examples above could be

REGRESS ; Lhs = profit ; Rhs = one,sales ; Het $
REGRESS ; Lhs = profit ; Rhs = one,sales ; Plot $

This applies to all 'switch – type' specifications, whether essential or optional.

Other specifications provide information as part of the sentence. In such a case, the provision will always be in the form

; Specification = information

For example,

; Wts = weighting variable name

will be used to specify a weighting variable for estimation. When a specification provides information after an equals sign, then the string must be in exactly the form shown for that command – you may not include superfluous text in this case. *LIMDEP* will always be looking for the equals sign in a specific place, and will issue a diagnostic when it does not find it. Thus, for the example shown,

; Wts variable = weighting variable

will produce an error message.

Model commands for *LIMDEP*'s models may become very long and complicated mixtures of many specifications. The language is fairly terse so as to be concise, but bear in mind that it is being used to specify several hundred different variations on over 50 broad model categories.

E1.4 The Command Builders

Nearly all of the documentation to follow, and most discussions, will assume that commands are being issued from a text editor (editing window), such as shown in the example below. The command is issued, that is, actually carried out, by highlighting it in the editing window and submitting it with the GO button. (See Chapter R2 for further discussion.)

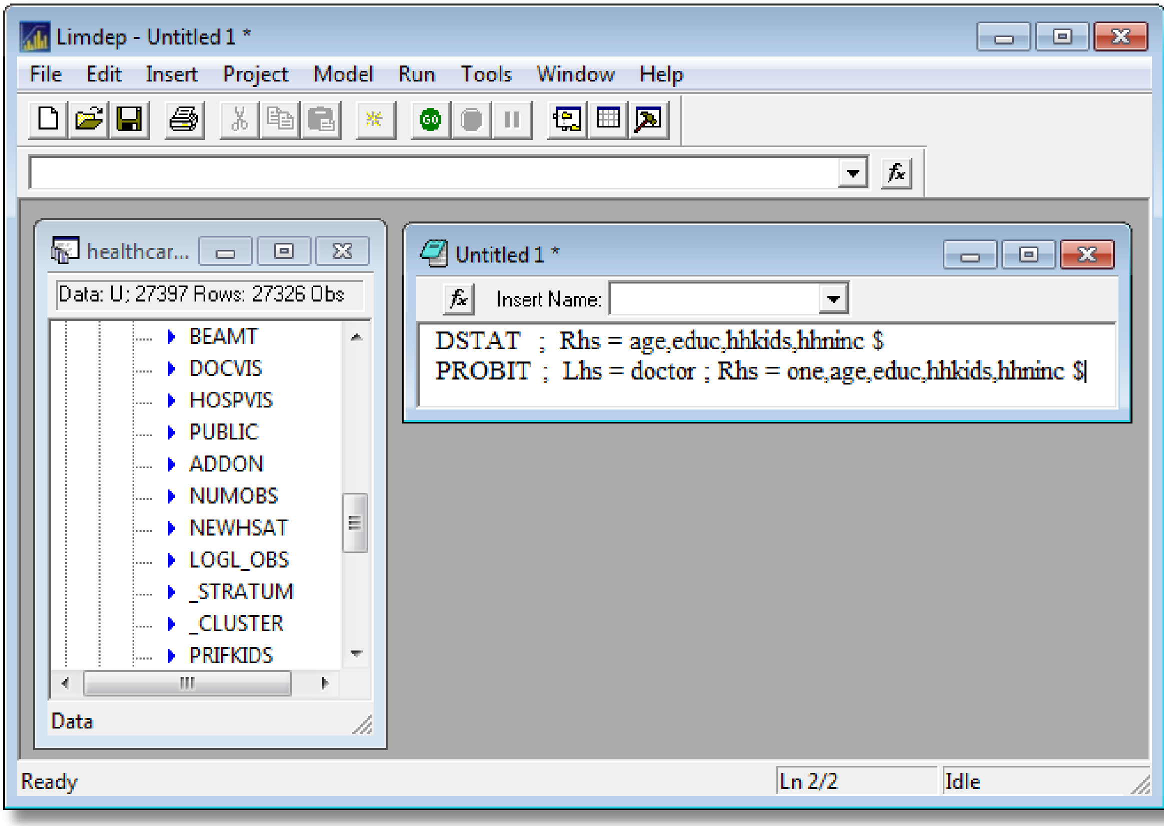


Figure E1.1 Desktop with Project and Editing Windows

An alternative method of submitting commands is to use the interactive dialog boxes, which, for reasons that will be evident shortly, we call the command builders. Command builders for model commands are produced by selecting Model in the main menu above the toolbar. This brings down the menu shown in Figure E1.2 which offers a number of groups of model frameworks. You may then select one of the groupings of models shown, to open a subsidiary menu of specific models. An example for the binary choice models is shown in Figure E1.2. You may then click a model name to open the command builder dialog box for that specific command. An example for the **PROBIT** command is shown in Figure E1.3.

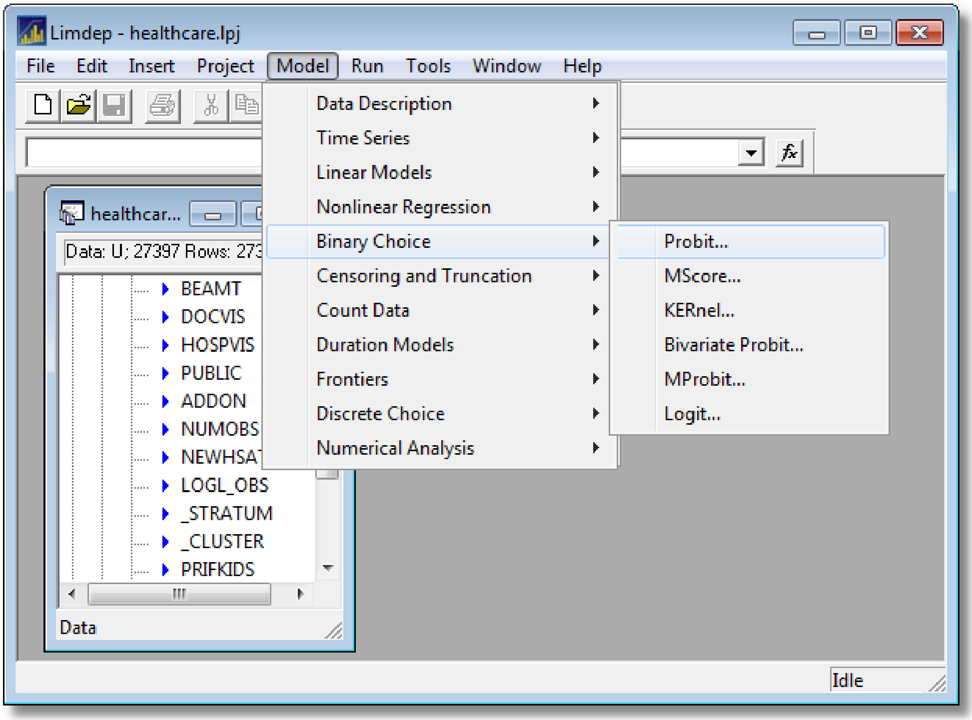


Figure E1.2 Model Choice from the Model Menu

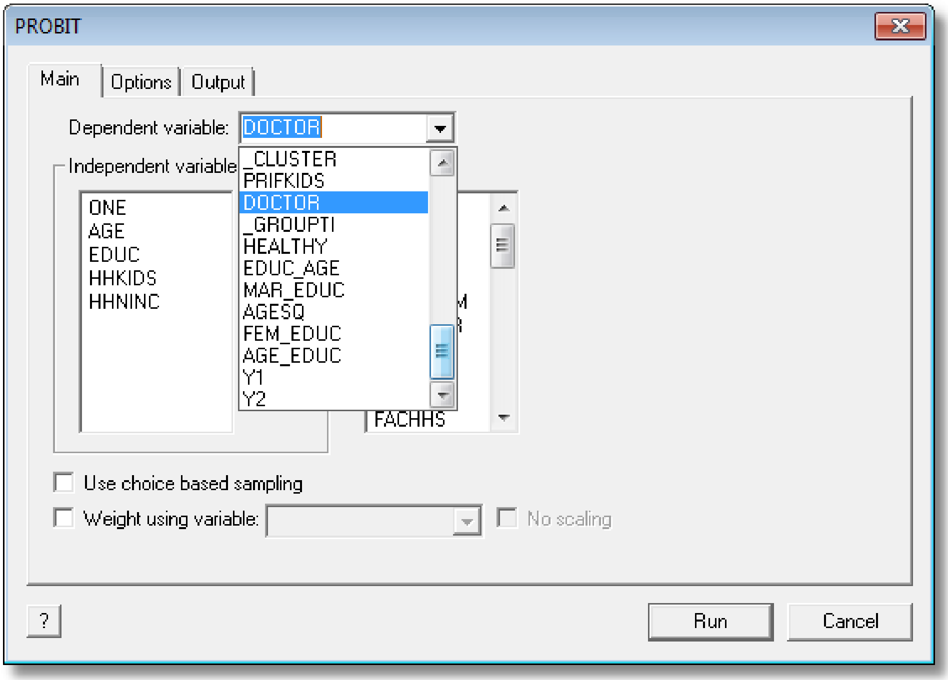


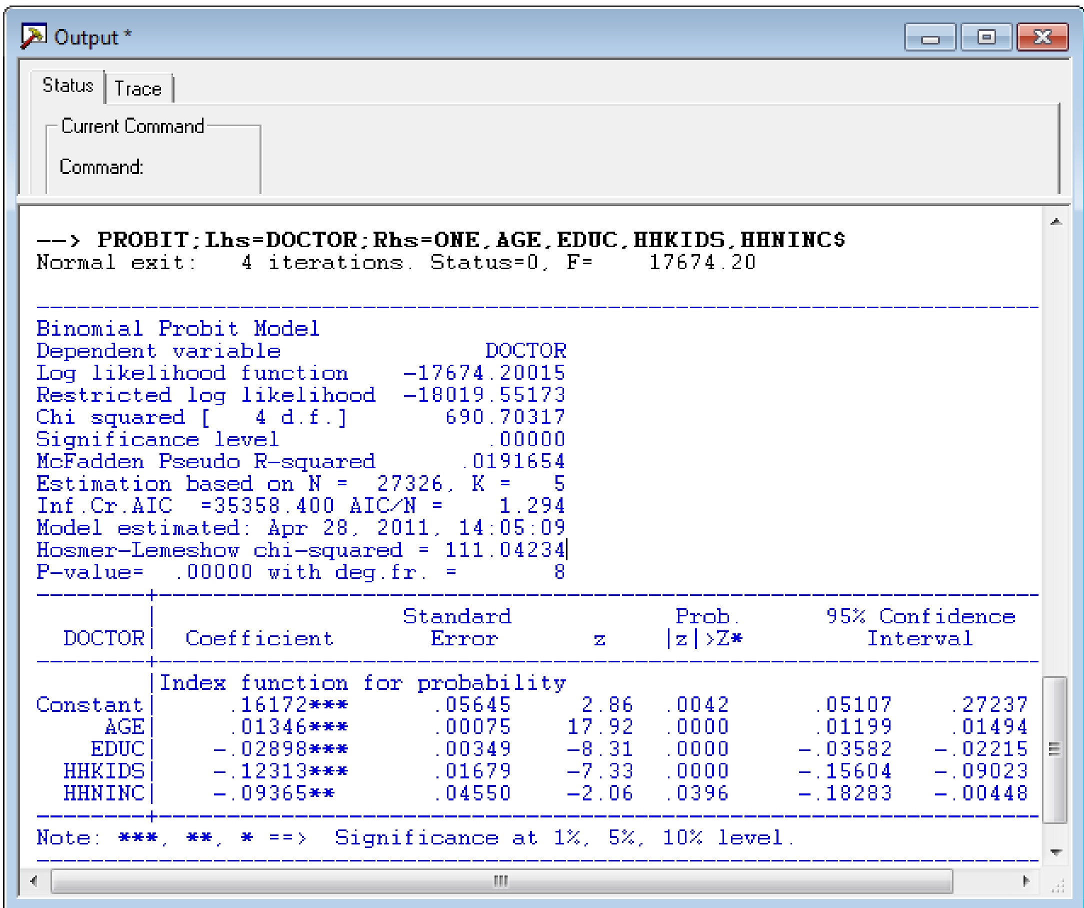
Figure E1.3 Main Page for Command Builder (PROBIT)

NOTE: The '?' button at the lower left of the command builder dialog box is a link to a context sensitive Help file that contains a large amount of information about the command.

The Main tab (or page) in the command builder dialog box requests the variables part of the commands. Note in Figure E1.3, we have selected the Lhs and Rhs variables that will appear in the probit model to be estimated. A few of the optional features will usually appear here as well, including, for example, a weighting variable. Other optional specifications are provided on the other pages of the command builder window. As can be seen in Figure E1.3, the probit model command builder has two additional pages. *Note, you must provide the essential variable parts of a command before you may enter the* Options *page. The command builder will insist on this.*

Once you have selected the model specification in the command builder window, click the Run to submit the command to *LIMDEP* for processing. This produces two results: First, the command is carried out, and the results appear in the output window, as would result in general when a model command is issued. Second, as its name implies, the command builder 'builds' the model command, and places a copy of it in the output window with the results. (See Figure E1.4.)

The first line of text above the output is the command generated by this selection in the window. You can copy these commands from the output window and paste them into the editing window, as we have done in our example in Figure E1.5. You might find this useful if you wish to modify the model and reuse the command. The editor will usually be more convenient. Note, as well, that the command interpreter will ignore the leading '-->' so there is no absolute need to edit these characters out of the editing window.




```
--> PROBIT;Lhs=DOCTOR;Rhs=ONE,AGE,EDUC,HHKIDS,HHNINC$
Normal exit:    4 iterations. Status=0, F=     17674.20

-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                DOCTOR
Log likelihood function     -17674.20015
Restricted log likelihood   -18019.55173
Chi squared [    4 d.f.]       690.70317
Significance level               .00000
McFadden Pseudo R-squared      .0191654
Estimation based on N =  27326, K =    5
Inf.Cr.AIC  =35358.400 AIC/N =    1.294
Model estimated: Apr 28, 2011, 14:05:09
Hosmer-Lemeshow chi-squared = 111.04234
P-value=  .00000 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  DOCTOR|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|      .16172***     .05645     2.86  .0042      .05107     .27237
     AGE|      .01346***     .00075    17.92  .0000      .01199     .01494
    EDUC|    -.02898***      .00349    -8.31  .0000    -.03582    -.02215
  HHKIDS|    -.12313***      .01679    -7.33  .0000    -.15604    -.09023
  HHNINC|    -.09365**       .04550    -2.06  .0396    -.18283    -.00448
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Figure E1.4 Output Window

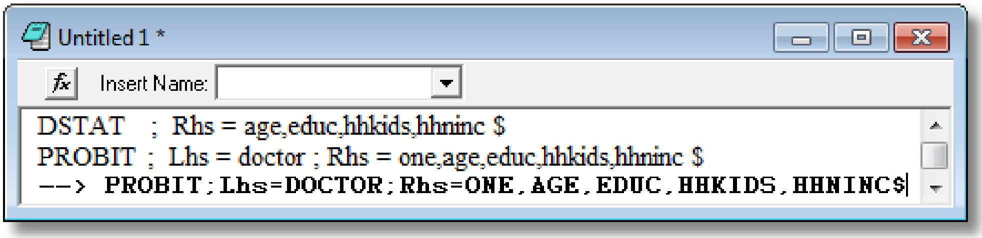


Figure E1.5 Detail from Editing Window

NOTE: The command builders are not complete. Some options and model forms must be specified with commands formed in the text editor. The command builders are intended generally for development of the more basic forms of the models and for relatively uncomplicated models. Not all optional features in all models are present in the command builder. Moreover, many of the model frameworks are not contained in the command builder menu. We anticipate that the command builders will be used by those who are becoming accustomed to using *LIMDEP*. After a relatively short introductory period, you will probably find the text editor more convenient than the command builders.

E1.5 Model Groups

The various model commands and modeling frameworks supported by *LIMDEP* and *NLOGIT* are discussed in the chapters to follow. The following are the model names for the different classes of estimators. Do note, some, such as **HISTOGRAM** and **BURR** are quite narrow, single purpose instructions, while others, such as **POISSON**, call for large classes of models that may (as in this case) contain a large number of different variants.

Data Setup and Model Preparation

NAMELIST	defines lists of variables for model commands (and matrices).
REVIEW	reviews model commands and create tables.
SORT	sorts variables.
SETPANEL	establishes parameters for panel data analysis.
IMPUTE	estimates imputation model for multiple imputation procedures.

Descriptive Statistics

CLASSIFY	discriminant analysis – classification into latent groups.
DSTAT	descriptive statistics.
TABLES	descriptive statistics for stratified data.

Cross Section

CROSSTAB	cross tabulations for discrete data.
HISTOGRAM	histograms for discrete and continuous data.
KERNEL	kernel density estimation of the density for a variable.

Time Series

IDENTIFY	descriptive statistics (ACF, PACF) for time series data.
SPECTRAL	spectral analysis of a time series.

Plotting

FPLOT	function plot for user specified function.
MPLOT	scatter plot of matrices.
PLOT	scatter or time plots of variables against each other.
SPLOT	simultaneous scatter plots for several variables.

Linear Regressions and Variants

Single Equation

FRONTIER	stochastic frontier models.
HREG	heteroscedastic linear regression based on Harvey's exponential model.
QREG	quantile regression.
REGRESS	linear regression models (also **OLSQ** and **CRMODEL**).
TSCS	time series/cross section, covariance structure models.
2SLS	two stage (instrumental variable) estimation of linear models.
LIML	limited information maximum likelihood estimation.
LOWESS	locally weighted nonparametric regression.

Multiple Linear Equation Models

SURE	linear seemingly unrelated regression models.
3SLS	three stage (IV, GLS) estimator for systems of linear equations.

Sample Selection Models

MATCH	propensity score matching to analyze treatment effects.
SELECT	sample selection models with linear and tobit models.
INCIDENTAL	incidental truncation (selection) model.
SWITCH	switching regression models.

Nonlinear Regression, Optimization, Manipulation of Nonlinear Functions

ARMAX	Box-Jenkins ARMA and dynamic linear equations.
BOXCOX	regression based on the Box-Cox transformation of variables.
NLSQ	nonlinear least squares for nonlinear regression models.
NLSURE	nonlinear systems of equations, SURE or GMM estimation.

Analysis of Nonlinear Functions

FINTEGRATE	function integration for user specified nonlinear function.
GMME	GMM estimation of model parameters.
MAXIMIZE	maximization of user specified functions.
MINIMIZE	user defined minimization command.
WALD	standard errors and Wald tests for user specified nonlinear functions.
SOLVE	finds roots of nonlinear functions.
FUNCTION	computes and displays function values.

Single Equation Models for Binary, Ordered and Multiple Discrete Choices

ARCTANGENT	arctangent model for binary choice.
BINARY CHOICE	simulation program for all binary choice estimators.
BIVARIATE	bivariate probit models, partial observability models.
BURR	Burr model for binary choice.
CLOGIT	multinomial logit model for discrete choice among multiple alternatives (*LIMDEP only – not used in NLOGIT*).
COMPLOG	complementary log log model for binary choice.
FRACRESP	fractional response model for panel data.
GOMPERTZ	Gompertz model for binary choice.
LOGIT	binary and multinomial choice models based on the logistic distribution.
MLOGIT	multinomial logit model.
MPROBIT	multivariate probit model.
MSCORE	maximum score semiparametric estimation for binary dependent variable.
NPREG	nonparametric regression models.
ORDERED	ordered probability models for ordered discrete choice.
PROBIT	several forms of binary choice models.
SEMIPAR	Klein and Spady semiparametric estimator for binary choice.

Models for Count Data

GAMMA	gamma model for count data.
NEGBIN	negative binomial regression model.
POISSON	models for count data.

Models for Censored Variables

BTOBIT	bivariate tobit models.
GROUPED	regression models for categorical censored data.
MIMIC	multiple indicators and multiple causes for a latent variable.
NTOBIT	nested tobit models.
TOBIT	censored regression models.

Models for Variables with Limited Range of Variation

LOGLINEAR	loglinear models, beta, gamma, Weibull, exponential, geometric, inverse Gaussian, arctangent, binomial.
LOGNORMAL	lognormal regression model.
TRUNCATE	truncated regression models.

Models for Survival Times and Hazard Functions

SURVIVAL	survival (hazard function) models.

Post Estimation Commands for Estimated Models

PARTIAL EFFECTS	analyzes average partial effects.
DECOMPOSE	Oaxaca-Blinder decompositions.
SIMULATE	simulation of outcome variables with estimated models.

E1.6 General Model Specifications

The preceding section listed the model commands that are used for estimation and data analysis. The various specifications that accompany the command are used to specify the basic model and to add certain optional features or model variations. Some of these are extremely general. For example, nearly every model command will contain a **; Lhs = variable(s)** specification to identify the dependent variable(s). In contrast, **; Cost** is used only by the frontier model command to request a cost (as opposed to a production) stochastic frontier model. Altogether, there are several hundred different specifications that attend the various model commands. The following list gives some of the most frequently used model specifications, in decreasing level of generality. Specialized codes, such as **; Cost** and **; DEA** for the frontier models are omitted here, and are detailed in the particular chapters for the specific models.

E1.6.1 Variable Specifications in Model Commands

These essential parts of model commands are described in Chapter R8.

; Lhs = names	specifies model dependent variable(s).
; Rhs = names	specifies model independent variable(s).
; Rh1 = names	provides first list of variables in two equation model.
; Rh2 = names	provides second list of variables in two equation model.
; Inst = names	provides list of instrumental variables.
; Wts = name	specifies a weighting variable; the optional parameter, [**,Noscale**] prevents scaling to sum to sample size.
; Hfn = names	provides a list of variables for variance in heteroscedasticity model.
; Hf1 = names,	**; Hf2 = names**, **; Hfu = names**, **; Hfe = names**, **; Hfr = names** are all used to provide lists of variables that appear in variance (heteroscedastic) functions.
; Eqn = names	is used in the SURE/3SLS, multivariate probit models to provide the lists of variables that appear in the set of equations. The '**n**' will be the number of the equation, as in **; Eq1 = list of variables**.

NOTE: The variable *one* is a program created variable that always equals 1.0. Use *one* to indicate a constant term in a model.

E1.6.2 Controlling Output from Model Commands

These optional features are described in the Chapter R9.

; Par	requests the program to keep ancillary parameters such as a correlation coefficient in the main results vector *b*.
; Partial Effects	requests display of marginal effects (same as **; Marginal Effects**).
; OLS	requests display of least squares starting values when (and if) they are computed.

; Clevel = value requests use of value for confidence level in confidence intervals in model results tables.
; Table = name requests the estimator to save model results to be combined later in output tables.
; Covariance Matrix requests display of the estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Matrix includes matrix forms as embedded objects in the output window.
; Quietly requests that model output not be displayed for this command.

E1.6.3 Robust Asymptotic Covariance Matrices

The clustering computation for robust covariance matrices is described in Sections R10.2 and E17.5. Choice based sampling is described at several points; a somewhat detailed discussion appears in Section E39.4. Robust estimation also appears in the discussion of several models. General discussion appears in Section E17.5.

; Choice requests the choice based sampling (sandwich with weighting) estimated matrix.
; Cluster = spec requests computation of the cluster form of corrected covariance estimator.
; Stratum = spec is used with **; Cluster** to specify a stratified, two level form of data clustering.
; Robust requests a 'sandwich' estimator or robust covariance matrix for TSCS and several discrete choice models.

E1.6.4 Optimization Controls for Nonlinear Optimization

These optional features are described in detail in Chapter R26.

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets the convergence value for convergence on the gradient.
; Tlf [= value] sets the convergence value for function convergence.
; Tlb[= value] sets the convergence value for convergence on change in parameters.
; Tln = value sets the convergence tolerance for nonlinear least squares.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Lpt = n sets the number of points to use for Laguerre quadrature.
; Hpt = n sets the number of points to use for Hermite quadrature.
; Set keeps current setting of optimization parameters as permanent.

E1.6.5 Predictions and Residuals

Fitted values (predictions) and residuals are described in Section R12.2.

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps the fitted values as a new (or replacement) variable in data set.
; Res = name	keeps the residuals as a new (or replacement) variable.
; Prob = name	saves the probabilities as a new (or replacement) variable for discrete choice models such as probit or logit.
; Fill	requests that missing values or values outside the estimating sample be replaced by fitted values based on the estimated model.

E1.6.6 Hypothesis Tests and Restrictions

These features are described in Chapter R13.

; CML: spec	defines a constrained maximum likelihood estimator.
; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; Rst = list	specifies equality and fixed value restrictions.

E1.6.7 Setup for Panel Data Models

LIMDEP contains an extremely large menu of panel data estimators. The set of controls listed below are used primarily with the nonlinear estimators for panel data. The data arrangement is described in Chapter R5 and in Section E15.2. Models may also have a two way structure, in which there is a time specific effect. Time effects are described in Sections E17.3 and E30.4. The controls listed below are discussed in numerous chapters and summarized with the estimators in Chapter E30.

Data Specification for Panel Data

SETPANEL	is the general command used to set up a panel.
; Pds = spec	is the general specification for panel data, either a fixed number of periods or a variable number given by the named variable.
; Time = spec	specifies the time dimension for two way fixed effects models.
; Periods = t	specifies a length of time (number of periods) for panel estimators.
; Str = name	specifies a stratification variable for **DSTAT, REGRESS, SURVIVAL**.

Panel Data Specifications in Nonlinear Modeling Frameworks

; FEM	specifies a fixed effects model.
; Fixed	when used, requests fixed effects. This is used by only two models, **LOGIT** and **REGRESS**. **; FEM** is used more generally. In **LOGIT**, **; Fixed** and **; FEM** request different estimators. Elsewhere, **; Fixed** and **; FEM** will be synonyms.
; Random	is the general request for random effects models.
; RPM	indicates a random parameters model used throughout *LIMDEP*. (Note, **; RPL** is a random parameters counterpart – random parameters logit model – that is used only in *NLOGIT*)
; LCM	requests a latent class model. It appears with **; Pts = number of classes**.
; Halton	is used with **; RPM** and **; RPL** to request Halton sequences.
; Cor	is used with **; RPM** and **; RPL** to request correlated random parameters.

E2: Descriptive Statistics for Cross Section and Panel Data

E2.1 Introduction

This chapter describes methods of obtaining univariate descriptive statistics for one or more variables in your data set. Procedures are given for cross sections and for panel data.

E2.2 Univariate Summary Statistics

The primary command for descriptive statistics is

DSTAT ; Rhs = list of variables $

This produces a table which lists for each variable, x_k, $k = 1,...,K$, the basic statistics:

Sample mean $= \bar{x}_k = \frac{1}{n}\sum_{i=1}^{N_k} x_{ik}$,

Standard deviation $= s_k = \sqrt{\frac{1}{N_k - 1}\sum_{i=1}^{N_k}\left(x_{ik} - \bar{x}_k\right)^2}$,

Maximum value,

Minimum value,

Number of valid (nonmissing) cases.

Standard deviations are computed in two steps, computing the means first, then the sums of squared deviations (rather than in the less accurate one step using the mean square minus the square of the mean).

E2.2.1 Weights

Weights may be used in computing all of the sums above by specifying

; Wts = name of weighting variable

Weights are always scaled so they sum to the current sample size. Thus, for example, the weighted mean would be

$$\bar{x}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} w_{ik} x_{ik}$$

where $$w_{ik} = \frac{N_k z_{ik}}{\sum_{i=1}^{N_k} z_{ik}}$$

and z_i = the weighting variable.

E2.2.2 Missing Observations in Descriptive Statistics

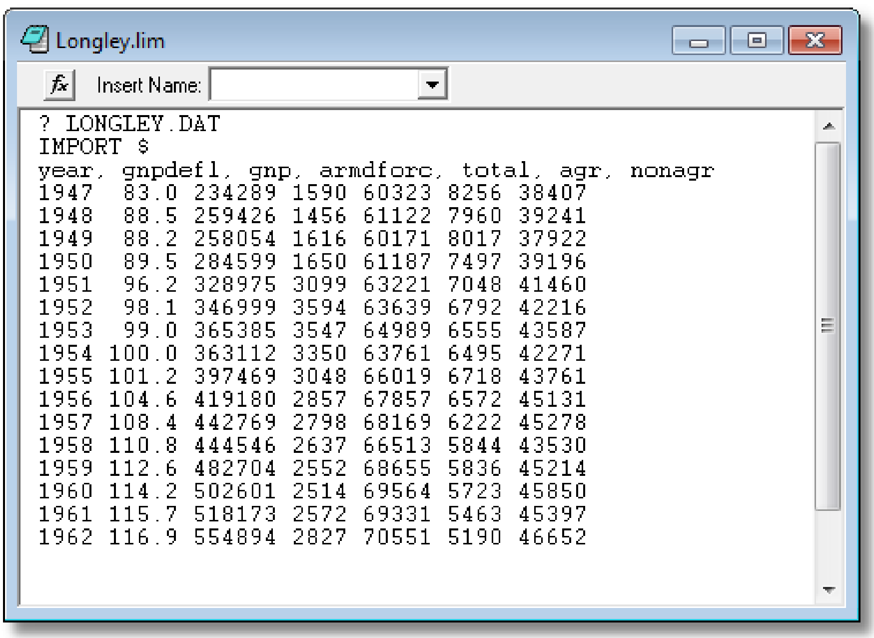
In all cases, weighted or otherwise, sums are based on the valid observations. **DSTAT** automatically selects out the missing data. Most other models (save for those in *NLOGIT* 5 and most of the panel data estimators) do not routinely do so unless you have the **SKIP** switch set. (See Section R7.5.5.) Each variable may have a different number of valid cases, so the table of results gives the number for each one.

NOTE: The covariance and correlation matrices are based on the subset of observations for which there were no missing data for any variables. Each row in the table of results will list the number of valid cases used for that particular variable. Unfortunately, if different observations are missing for the various variables used in a covariance or correlation matrix, the union of the observations for which all variables are present can contain very few observations. For better or worse, this union is the set of observations used in computing the matrices.

If your data contain missing values, the scaling described in the previous section is automatically adjusted for each variable. Moments are scaled by the number of valid observations or sum of weights for that variable.

E2.2.3 Display of Descriptive Statistics

The standard display of results for descriptive statistics is shown in the example below for the Longley data (Longley.dat) displayed in Figure E2.1 as they are ready to be read into *LIMDEP*. The Longley data as well as the other sample data sets are located in the Data Sets book of the Help file and also in the C:\LIMDEP10\Data Files folder created with program installation.




```
? LONGLEY.DAT
IMPORT $
year, gnpdefl, gnp, armdforc, total, agr, nonagr
1947  83.0 234289 1590 60323 8256 38407
1948  88.5 259426 1456 61122 7960 39241
1949  88.2 258054 1616 60171 8017 37922
1950  89.5 284599 1650 61187 7497 39196
1951  96.2 328975 3099 63221 7048 41460
1952  98.1 346999 3594 63639 6792 42216
1953  99.0 365385 3547 64989 6555 43587
1954 100.0 363112 3350 63761 6495 42271
1955 101.2 397469 3048 66019 6718 43761
1956 104.6 419180 2857 67857 6572 45131
1957 108.4 442769 2798 68169 6222 45278
1958 110.8 444546 2637 66513 5844 43530
1959 112.6 482704 2552 68655 5836 45214
1960 114.2 502601 2514 69564 5723 45850
1961 115.7 518173 2572 69331 5463 45397
1962 116.9 554894 2827 70551 5190 46652
```

Figure E2.1 Longley Data to Be Read from Text Editor

The command is

DSTAT ; Rhs = year,gnpdefl,gnp,armdforc,total,agr,nonagr $

Descriptive Statistics

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
YEAR	1954.50	4.76095	1947	1962	16	0
GNPDEFL	101.681	10.79155	83	116.900	16	0
GNP	387698.	99394.9	234289	554894	16	0
ARMDFORC	2606.69	695.920	1456	3594	16	0
TOTAL	65317.0	3511.97	60171	70551	16	0
AGR	6636.75	930.816	5190	8256	16	0
NONAGR	42819.6	2846.30	37922	46652	16	0

If weights have been specified with **; Sts = variable**, the title line of the table will declare the name of the weighting variable, for example ‘Descriptive Statistics (Weighted by POPULATN).’

E2.2.4 Command Builder Dialog Box

Select Model:Data Description/Descriptive Statistics to invoke the dialog boxes for this program. The Main page is shown in Figure E2.2. The various optional specifications available for the **DSTAT** command are provided on the Options page, as shown in Figure E2.3.

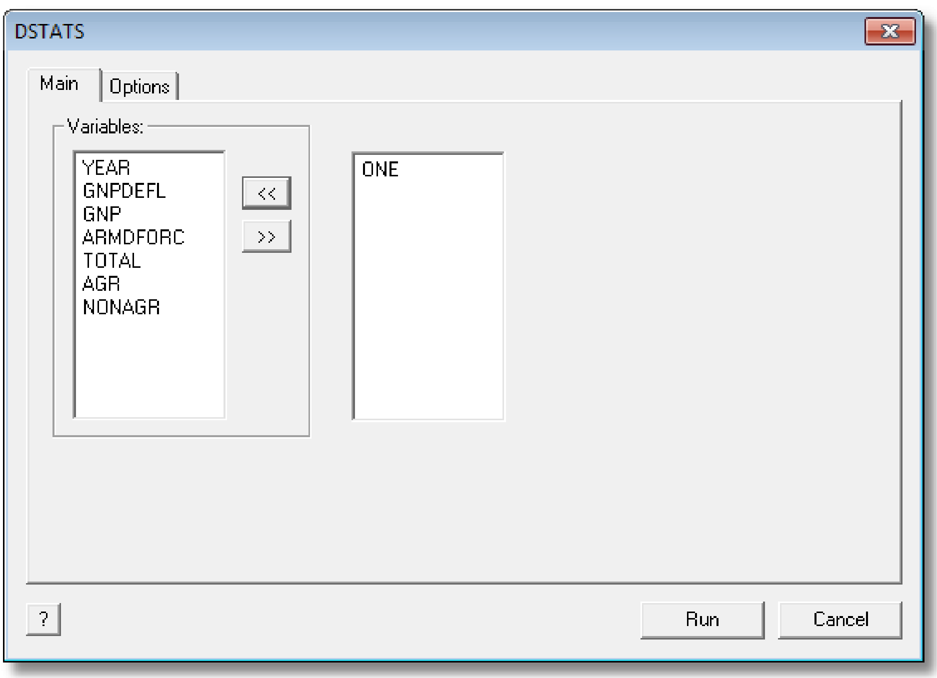


Figure E2.2 Command Builder Main Page for DSTAT

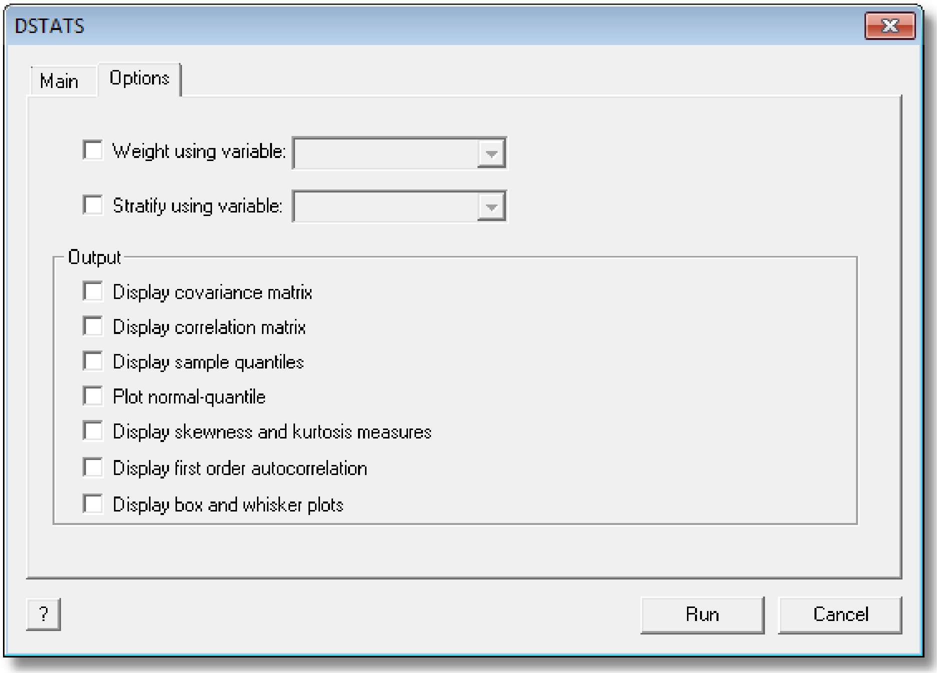


Figure E2.3 Command Builder Options Page for DSTAT

E2.3 Standard Error of the Mean

The sample mean reported in the standard table is purely descriptive. When the sample mean is viewed as an estimator of a population mean, it is customary to report the 'standard error,' as well. The estimate of the standard error of the mean is estimated with

$$s_{\bar{x}}(k) = \frac{s_x(k)}{\sqrt{N_k}}$$

where $s_x(k)$ is the standard deviation for x_k and N_k is the sample size. Request a display of the standard error of the mean with each variable by adding

; Sem

to the **DSTAT** command. This will also produce a listing of a confidence interval,

$$CI(k) = \bar{x}_k \pm z^*_{1-\alpha/2} s_{\bar{x}}(k)\ ,$$

where $z^*_{1-\alpha/2}$ is the critical value from the standard normal distribution. We do not assume that data are drawn from a normal population, so the normal rather than the t distribution is used for the confidence interval. If you wish to produce confidence intervals based on the t distribution with $N_k - 1$ degrees of freedom, rather than the normal, use

; Sem (t).

(Note that if the sample size exceeds 50, the *t* and normal distributions will be indistinguishable.) By default, the estimator produces a 95% confidence interval. You can change this with

; Clevel = value.

(See Section R9.1.1.) Partial results from the earlier example appear below.

```
Descriptive Statistics
--------+--------------------------------------------------------------------
Variable|     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+--------------------------------------------------------------------
    YEAR|  1954.50      4.76095         1947         1962            16       0
        | SE(mean) =    1.19024   95% CI = [  1952.16718,1956.83282  ]
 GNPDEFL|  101.681     10.79155           83      116.900            16       0
        | SE(mean) =    2.69789   95% CI = [    96.39349,106.96901   ]
  NONAGR|  42819.6      2846.30        37922        46652            16       0
        | SE(mean) =  711.57410   95% CI = [ 41424.90289,44214.22211 ]
--------+--------------------------------------------------------------------
```

E2.4 Clustered Data

When the sample mean is used as an estimator and the data are clustered as in a panel or sometimes in a stratified data set, then the standard error of the mean computed in the previous section will generally underestimate the true standard error of the estimator. The ‘cluster’ estimator is often used to produce a more robust estimator of the standard error. The alternative formulation used in this case is (after a bit of algebra)

$$s_{\bar{x}}^{c}(k) = \sqrt{\frac{1}{N_k}\sum_{c=1}^{C}\left(\sum_{i=1}^{N_c}\left(x_{ik,c} - \bar{x}_k\right)\right)^2}$$

where C is the total number of clusters indexed $c = 1,...,C$, N_c is the number of observations in cluster c. (There are no corrections for degrees of freedom). This calculation does not change the basic statistics; it modifies the computation of the standard error of the mean. This computation is requested by adding

; Cluster = specification

to the **DSTAT** command. (See Section R10.1 for details.) If the data set is such that the full population sizes are known and not (assumed to be) infinite, then one may specify a ‘finite population correction’ with

; FPC = the fixed number of clusters in the population from which the sample is drawn.

When you specify a finite population correction, you provide the known value for the total number of clusters in the population. The reported standard error of the mean is computed as the square root of

$$Corrected\ Variance[\overline{x}] = \left(1 - \frac{C}{C^*}\right)\left(\frac{C}{C-1}\right) s_{\overline{x}}^{c}(k)$$

The finite population correction is (1 – C/C^*). The population number of clusters is assumed to be infinite if you do not specify **; FPC =** ***C*** ***** for a particular value of C^*.

The example below is based on the German health care data used in numerous applications throughout the documentation. The data are from Riphahn, Wambach, and Million (2003). The raw data were downloaded from the journal's data archive website, http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million, and are provided in the sample data file German-healthcare.dat. (The data files can be found in the resource folder created with installation: C:\LIMDEP10\Data Files.) The results show a comparison of the corrected and uncorrected estimates based on the health care panel data, which has 7,293 clusters ranging in size from one to seven.

DSTAT ; Rhs = hhninc ; Sem $

```
Descriptive Statistics
--------+-------------------------------------------------------------------
Variable|     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+-------------------------------------------------------------------
  HHNINC|    .35208       .17691            0      3.06710        27326       0
        | SE(mean) =     .00107   95% CI = [      .34999,.35418      ]
--------+-------------------------------------------------------------------
```

DSTAT ; Rhs = hhninc ; Sem ; Cluster = id $

```
Descriptive Statistics
--------+-------------------------------------------------------------------
Variable|     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+-------------------------------------------------------------------
  HHNINC|    .35208       .17691            0      3.06710        27326       0
        | SE(mean) =     .00186   95% CI = [      .34845,.35572      ]
--------+-------------------------------------------------------------------
Clusters|  Cluster corrected std. deviations:    7293 clusters,       1 strata
      ID|  Numbers of clusters and strata may vary if there are missing values
--------+-------------------------------------------------------------------
```

The computation above can also be done in the linear regression model; regression on only a constant computes the sample mean. The commands

REGRESS ; Lhs = hhninc ; Rhs = one $

and **REGRESS ; Lhs = hhninc ; Rhs = one ; Cluster = id $**

produce the regression results below, which it can be seen replicate the computations in **DSTAT**.

```
+----------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.     |
| Sample of  27326 observations contained   7293 clusters defined by   |
| variable ID       which identifies by a value a cluster ID.          |
+----------------------------------------------------------------------+
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                 =          .35208
             Standard deviation   =          .17691
             Number of observs.   =           27326
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .35208***      .00186   189.72  .0000       .34845    .35572
--------+--------------------------------------------------------------------
Constant|     .35208***      .00107   328.99  .0000       .34999    .35418
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E2.5 Skewness and Kurtosis

The third and fourth moments for the variables may be obtained with

DSTAT ; Rhs = list of variables
; All $

This requests the skewness and kurtosis measures,

$$\text{Sample skewness} \quad = m_3 = \frac{\sum_{i=1}^{N_k}(x_{ik} - \overline{x}_k)^3 / (N_k - 1)}{s_k^3},$$

$$\text{Sample kurtosis} \quad = m_4 = \frac{\sum_{i=1}^{N_k}(x_{ik} - \overline{x}_k)^4 / (N_k - 1)}{s_k^4}.$$

(This option may be combined with **; Sem** described in the preceding section.) Note that the higher moments are normalized by the standard deviations. This produces the comparison to the values for the normal distribution of zero and three, respectively. For the earlier example, we have

DSTAT ; Rhs = year,gnpdefl,gnp,armdforc,total,agr,nonagr
; All $

```
Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|     Mean        Std.Dev.     Minimum      Maximum        Cases Missing
--------+---------------------------------------------------------------------
    YEAR|  1954.50       4.76095        1947         1962           16       0
        | Skewness  .0000  Kurtosis=  1.6787
 GNPDEFL|  101.681      10.79155          83      116.900           16       0
        | Skewness -.1418  Kurtosis=  1.7117
     GNP|  387698.       99394.9      234289       554894           16       0
        | Skewness  .0245  Kurtosis=  1.7643
ARMDFORC|  2606.69       695.920        1456         3594           16       0
        | Skewness -.3917  Kurtosis=  1.9226
   TOTAL|  65317.0       3511.97       60171        70551           16       0
        | Skewness -.0913  Kurtosis=  1.5455
     AGR|  6636.75       930.816        5190         8256           16       0
        | Skewness  .2817  Kurtosis=  1.9502
  NONAGR|  42819.6       2846.30       37922        46652           16       0
        | Skewness -.4506  Kurtosis=  1.7654
--------+---------------------------------------------------------------------
```

E2.6 Display Format

Three different formats are provided for display of descriptive statistics.

E2.6.1 Fixed Width Format

The default output display shown in the earlier examples is in a floating point format with integers displayed for the minimum and maximum if the variable is an integer. If your data contain extremely large or small values, you may prefer to change the display to scientific notation. Add

; Fixed

to the command to request a fixed with decimal format. For example,

DSTAT ; Rhs = year,gnpdefl,gnp,armdforc,total,agr,nonagr
; Fixed $

produces the following:

```
Descriptive Statistics
--------+--------------------------------------------------------------------
Variable|     Mean        Std.Dev.     Minimum      Maximum        Cases Missing
--------+--------------------------------------------------------------------
    YEAR|  .195450D+04  .476095D+01     1947         1962           16       0
 GNPDEFL|  .101681D+03  .107916D+02       83      .116900D+03       16       0
     GNP|  .387698D+06  .993949D+05   234289       554894           16       0
ARMDFORC|  .260669D+04  .695920D+03     1456         3594           16       0
   TOTAL|  .653170D+05  .351197D+04    60171        70551           16       0
     AGR|  .663675D+04  .930816D+03     5190         8256           16       0
  NONAGR|  .428196D+05  .284630D+04    37922        46652           16       0
--------+--------------------------------------------------------------------
```

E2.6.2 Matrix Output

The results of this procedure may be embedded in the output window in a matrix that can be exported to other programs such as *Excel*. Add

; Matrix

to the **DSTAT** command to request this. Figure E2.4 shows the result for the previous example. Double clicking the embedded object displays the matrix containing the results as shown in Figure E2.5.

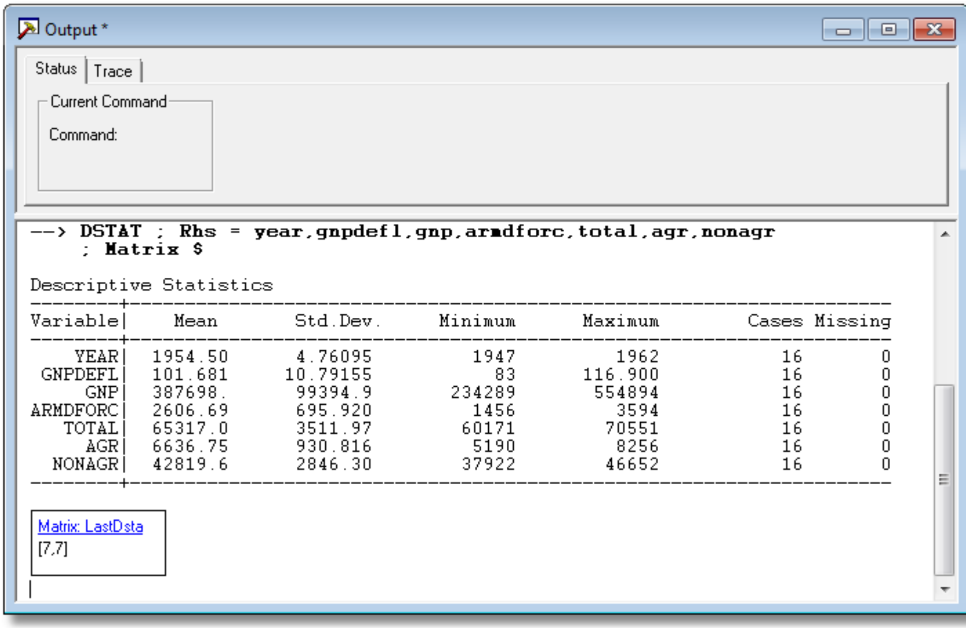
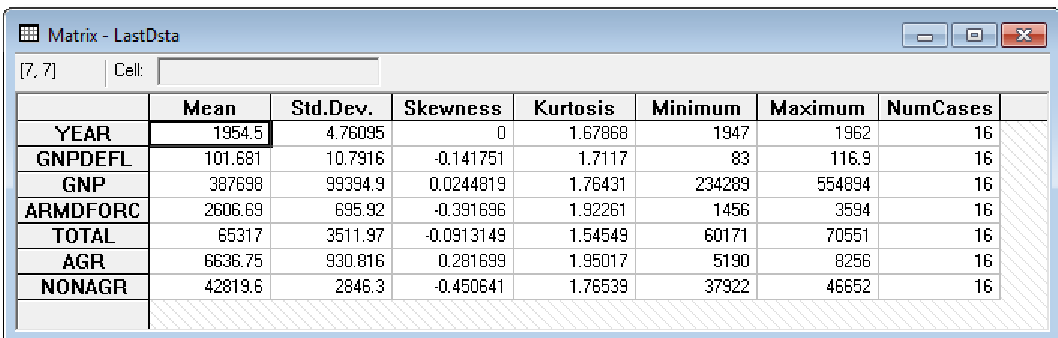


Figure E2.4 Descriptive Statistics with Results Matrix



	Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum	NumCases
YEAR	1954.5	4.76095	0	1.67868	1947	1962	16
GNPDEFL	101.681	10.7916	-0.141751	1.7117	83	116.9	16
GNP	387698	99394.9	0.0244819	1.76431	234289	554894	16
ARMDFORC	2606.69	695.92	-0.391696	1.92261	1456	3594	16
TOTAL	65317	3511.97	-0.0913149	1.54549	60171	70551	16
AGR	6636.75	930.816	0.281699	1.95017	5190	8256	16
NONAGR	42819.6	2846.3	-0.450641	1.76539	37922	46652	16

Figure E2.5 Embedded Descriptive Statistics Matrix

E2.7 Stratified Data

DSTAT allows a stratification variable to be used to provide descriptive statistics for subgroups in the sample. A second command, **TABLES**, described in the next section, is provided to allow a convenient format for the results.

One full set of results is produced for each value of the stratification variable. Use

```
DSTAT       ; Rhs = list of variables
            ; Str = the stratification variable ... $
```

The variable must take values 1,2,... Up to 50 strata may be defined by the variable, which is assumed to be discrete. If a continuous variable is given instead, too many strata (each with one observation) will result, and an error will follow.

To specify ranges of a continuous variable, use

```
CREATE      ; Copy = the stratification variable $
RECODE      ; ... to create the discrete variable ... $
DSTAT       ; ... ; Str = the recoded variable $
```

The health care data provides an example. The following produces descriptive statistics for several variables for men, women, and the full sample.

```
CREATE      ; gender = female + 1 $
DSTAT       ; Rhs = age,educ,hhninc,married
            ; Str  = gender $
```

```
--------------------------------------------------------------------------
Descriptive Statistics for AGE
Stratification is based on GENDER
Subsample        |        Mean     Std.Dev.    Cases  Sum of wts  Missing
-----------------+--------------------------------------------------------
GENDER      =  1 |   42.652812    11.270394    14243    14243.00        0
GENDER      =  2 |   44.475961    11.319204    13083    13083.00        0
Full Sample      |   43.525690    11.330248    27326    27326.00        0
-----------------+--------------------------------------------------------
Descriptive Statistics for EDUC
Stratification is based on GENDER
-----------------+--------------------------------------------------------
Subsample        |        Mean     Std.Dev.    Cases  Sum of wts  Missing
-----------------+--------------------------------------------------------
GENDER      =  1 |   11.728700     2.436490    14243    14243.00        0
GENDER      =  2 |   10.876381     2.109105    13083    13083.00        0
Full Sample      |   11.320631     2.324885    27326    27326.00        0
-----------------+--------------------------------------------------------
Descriptive Statistics for HHNINC
Stratification is based on GENDER
-----------------+--------------------------------------------------------
Subsample        |        Mean     Std.Dev.    Cases  Sum of wts  Missing
-----------------+--------------------------------------------------------
GENDER      =  1 |     .359054      .173564    14243    14243.00        0
GENDER      =  2 |     .344495      .180179    13083    13083.00        0
Full Sample      |     .352084      .176908    27326    27326.00        0
-----------------+--------------------------------------------------------
Descriptive Statistics for MARRIED
Stratification is based on GENDER
```

```
-----------------+-----------------------------------------------------
Subsample        |        Mean     Std.Dev.    Cases  Sum of wts  Missing
-----------------+-----------------------------------------------------
GENDER      =  1 |     .765148      .423921    14243    14243.00        0
GENDER      =  2 |     .751510      .432154    13083    13083.00        0
Full Sample      |     .758618      .427929    27326    27326.00        0
-----------------+-----------------------------------------------------
```

E2.8 Tables for Stratified Samples

The command for descriptive statistics for stratified data arranged in separate tables is

TABLES ; Rhs = list of variables
; Str = stratification $

This procedure is used to compute means and standard deviations for a stratified sample. As shown below, it differs from the **DSTAT** command described above in the format of the tables that it produces. You can use this procedure for stratified data in the same manner as described in the previous section. The command can specify any of three types of sample partitioning:

TABLES ; Rhs = up to 10 variables
; Pds = specification for groups in a panel data set
or **; Str = specification for strata as defined above $**

One of the two setups for a partitioned sample is used. The **; Pds** or **; Str** sets up the sample specification as if it were a panel data set (which it could, but need not be). The results will include for the full sample, then for each stratum or group in the sample, means, standard deviations, and sample sizes for each variable specified. The data set may contain up to 10,000 strata and up to 10 variables for this processor.

The data used in the preceding example are an unbalanced panel observed for seven years. To illustrate the **TABLES** command, we will produce a table for the seven years using the default format of the command.

TABLES ; Rhs = hhninc ; Str = year $

```
---------+------------------------------------------------------------
         | Means and Standard Deviations for Clustered  or Stratified Data
         | Variable  =  HHNINC       Weights for observations are 1.000000
         | Full Sample =   27326 data rows. Valid rows =    27326
         | Rows skipped (bad stratum or weight)       =        0
         | Sum of weights for all valid observations   =    27326.000
 Overall | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
         |      .352       .177    27326.000      27326        0    27326
 Stratum | There were    7 strata   found in the sample
         | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
---------+------------------------------------------------------------
Stratum01|      .297       .148     3874.000       3874        0     3874
Stratum02|      .309       .140     3794.000       3794        0     3794
Stratum03|      .325       .165     3792.000       3792        0     3792
Stratum04|      .349       .164     4483.000       4483        0     4483
Stratum05|      .336       .158     3666.000       3666        0     3666
Stratum06|      .407       .191     4340.000       4340        0     4340
Stratum07|      .445       .217     3377.000       3377        0     3377
---------+------------------------------------------------------------
```

The default names are *stratum*00 to *stratum*99 then *strat*0100 to *strat*9999. If you have a relatively small number of strata, you may wish to provide names for them in the results. You may provide names for the strata with the specification

; Labels = labels for strata.

If you provide fewer labels than there are strata, the names you provide are used for the first group of strata, and the default names are used for the remainder. For the preceding example, we could use

```
TABLES      ; Rhs = hhninc ; Str = year
            ; Labels = 1984,1985,1986,1987,1998,1991,1994 $
```

```
---------+------------------------------------------------------------
         | Means and Standard Deviations for Clustered  or Stratified Data
         | Variable  =  HHNINC      Weights for observations are 1.000000
         | Full Sample =   27326 data rows. Valid rows =    27326
         | Rows skipped (bad stratum or weight)        =        0
         | Sum of weights for all valid observations   =    27326.000
 Overall | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
         |      .352       .177    27326.000       27326        0    27326
 Stratum | There were    7 strata   found in the sample
         | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
---------+------------------------------------------------------------
     1984|      .297       .148      3874.000        3874        0     3874
     1985|      .309       .140      3794.000        3794        0     3794
     1986|      .325       .165      3792.000        3792        0     3792
     1987|      .349       .164      4483.000        4483        0     4483
     1998|      .336       .158      3666.000        3666        0     3666
     1991|      .407       .191      4340.000        4340        0     4340
     1994|      .445       .217      3377.000        3377        0     3377
---------+------------------------------------------------------------
```

E2.8.1 Groups in the Sample

The general form of model commands allows you to partition the sample during the processing. See Section R8.7.3. To use that feature here to produce separate analyses for male and female headed households, we could use.

```
TABLES      ; For[female = 0,1] ; Rhs = age
            ; Str = year
            ; Labels = 1984,1985,1986,1987,1988,1991,1994 $
```

The results are as follows:

```
+------------------------------------------------------+
| Setting up an iteration over the values of FEMALE    |
| The model command will be executed for   2 values    |
| of this variable.  In the current sample of   27326 |
| observations, the following counts were found:       |
| Subsample   Observations    Subsample  Observations |
| FEMALE   =   0      14243    FEMALE  =   1     13083 |
+------------------------------------------------------+
| Actual subsamples may be smaller if missing values   |
| are being bypassed.  Subsamples with 0 observations |
| will be bypassed.                                    |
+------------------------------------------------------+
```

```
******************************************************************
*       Subsample analyzed for this command is FEMALE   =       0 *
*****************************************************************
---------+------------------------------------------------------------
         | Means and Standard Deviations for Clustered  or Stratified Data
         | Variable  =  AGE         Weights for observations are 1.000000
         | Full Sample =   14243 data rows. Valid rows =    14243
         | Rows skipped (bad stratum or weight)       =       0
         | Sum of weights for all valid observations   =    14243.000
 Overall | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
         |    42.653     11.270    14243.000       14243        0    14243
 Stratum | There were    7 strata   found in the sample
         | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
---------+------------------------------------------------------------
     1984|    42.992     11.060     2017.000        2017        0     2017
     1985|    42.916     11.057     1978.000        1978        0     1978
     1986|    43.014     11.120     1968.000        1968        0     1968
     1987|    42.892     11.124     1911.000        1911        0     1911
     1988|    42.729     11.297     2313.000        2313        0     2313
     1991|    42.325     11.571     2244.000        2244        0     2244
     1994|    41.653     11.583     1812.000        1812        0     1812
---------+------------------------------------------------------------
******************************************************************
*       Subsample analyzed for this command is FEMALE   =       1 *
******************************************************************
---------+------------------------------------------------------------
         | Means and Standard Deviations for Clustered  or Stratified Data
         | Variable  =  AGE         Weights for observations are 1.000000
         | Full Sample =   13083 data rows. Valid rows =    13083
         | Rows skipped (bad stratum or weight)       =       0
         | Sum of weights for all valid observations   =    13083.000
 Overall | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
         |    44.476     11.319    13083.000       13083        0    13083
 Stratum | There were    7 strata   found in the sample
         | Mean       Std Dev.   Sum of Weights   Sample  Missing    Total
---------+------------------------------------------------------------
     1984|    45.086     11.335     1857.000        1857        0     1857
     1985|    44.814     11.202     1816.000        1816        0     1816
     1986|    44.908     11.157     1824.000        1824        0     1824
     1987|    44.198     11.232     2170.000        2170        0     2170
     1988|    44.793     11.324     1755.000        1755        0     1755
     1991|    43.828     11.450     2096.000        2096        0     2096
     1994|    43.754     11.490     1565.000        1565        0     1565
---------+------------------------------------------------------------
```

E2.8.2 Weights

The moments may be weighted with

; Wts = any kind of weights

Weights are handled the same as described earlier, but are now scaled appropriately for each stratum as well as by variable, again to account properly for missing observations.

E2.9 Sample Quantiles

You may obtain more detailed statistics about variables by requesting the sample quantiles. This feature produces sample order statistics and the deciles and quartiles of the sample of values for each variable. The keyword in the command is

; Quantiles

NOTE: The quantiles feature in **DSTAT** is limited to samples of 200,000 observations.

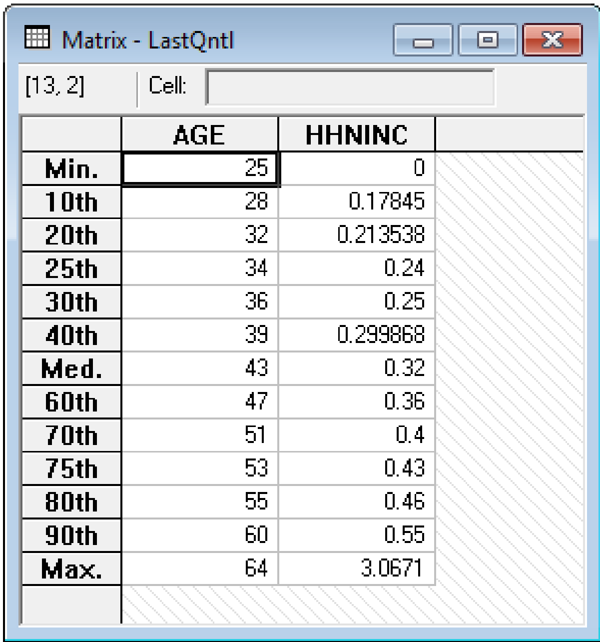
For an example based on our earlier results:

DSTAT ; Rhs = age,hhninc
; Quantiles $

```
Descriptive Statistics
--------+-------------------------------------------------------------------------
Variable|     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+-------------------------------------------------------------------------
     AGE| 43.52569     11.33025          25           64        27326       0
  HHNINC|   .35208       .17691           0      3.06710        27326       0
--------+-------------------------------------------------------------------------
Quantiles
---------------------------------------
Percentile     AGE          HHNINC
---------------------------------------
Min.          25.000       .00000
10th          28.000       .17845
20th          32.000       .21354
25th          34.000       .24000
30th          36.000       .25000
40th          39.000       .29987
Med.          43.000       .32000
60th          47.000       .36000
70th          51.000       .40000
75th          53.000       .43000
80th          55.000       .46000
90th          60.000       .55000
Max.          64.000       3.0671
---------------------------------------
Partition of Range Minimum to Maximum
---------------------------------------
Range of X     AGE          HHNINC
---------------------------------------
Minimum       25.000       .00000
1st.Qrtl      34.750       .76678
Midpoint      44.500       1.5336
3rd.Qrtl      54.250       2.3003
Maximum       64.000       3.0671
---------------------------------------
```

In addition to the displayed output, a matrix object, Matrix:LastQntl, has been embedded in the output window. A new matrix, *lastqntl* has also been placed in the matrix work area. You can double click the object matrix to display it, as shown in Figure E2.6. Using edit/copy and edit/paste, you can export the contents of this matrix to a spreadsheet program, such as *Excel*.

Matrix: LastQntl
[13,2]



	AGE	HHNINC
Min.	25	0
10th	28	0.17845
20th	32	0.213538
25th	34	0.24
30th	36	0.25
40th	39	0.299868
Med.	43	0.32
60th	47	0.36
70th	51	0.4
75th	53	0.43
80th	55	0.46
90th	60	0.55
Max.	64	3.0671

Figure E2.6 Quantiles Embedded Matrix

E2.9.1 Box and Whisker Plots

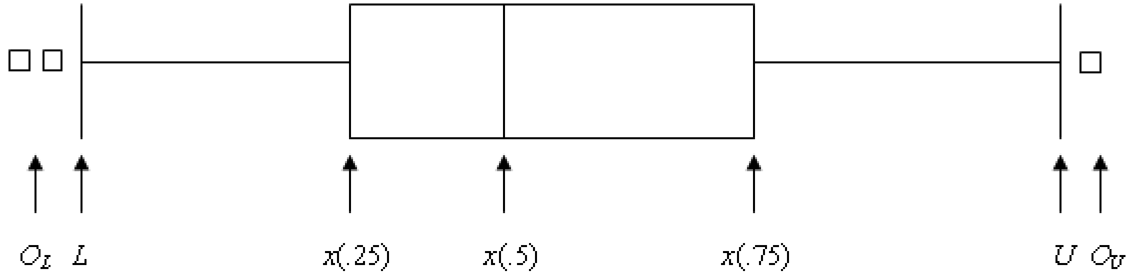
The box and whisker plot is a device used crudely to describe the location, range and skewness of a variable. *LIMDEP* will place up to five such plots in a figure. The function is requested simply by adding

; Box Plots

to the **DSTAT** command.

NOTE: Box and whisker plots can produce less than helpful, even absurd results when variables with very different scales are forced into the same figure. Figure E2.7 below suggests how the problem arises. Users are cautioned about this problem. There is no practical fix, other than to be sure that variables that are placed in the same figure have similar locations and scales.

Box and whisker plots are constructed (vertically in *LIMDEP*'s plots) for the variable x as follows:



$x(.5)$	=	the median of the sample values,
$x(.25)$	=	the 25^{th} sample percentile of the sample values,
$x(.75)$	=	the 75^{th} sample percentile of the sample values,
$x(.75) - x(.25)$	=	the interquartile range of x = the *IQR*,
L	=	the smallest sample value larger than $x(.25)$ - 1.5×*IQR*,
U	=	the largest sample value smaller than $x(.75)$ + 1.5×*IQR*,
O_L	=	sample values less than L marked as 'outliers,'
O_U	=	sample values greater than U marked as 'outliers.'

In the example below, box and whisker plots are produced for *age* and *educ* in the health care sample using

DSTAT ; Rhs = age,educ ; Box plots $

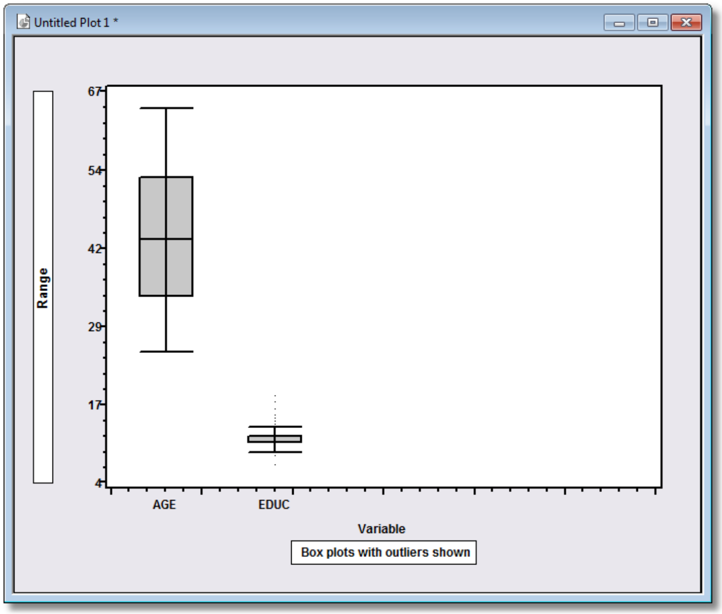


Figure E2.7 Box and Whisker Plots

E2.9.2 Related Procedures for Quantiles

There are four **CALC** functions provided for computing specific quantiles for a variable: Min(*variable*), Max(*variable*), Med(*variable*), and Qnt(*quantile*,*variable*). For example,

```
CALC        ; List
            ; Med(x) ; Qnt(0.50, x) $
```

will display the median of the sample of values on variable x (twice).

Two regression features based on quantiles are also available. The median regression is estimated by least absolute deviations (LAD). The LAD estimator is requested with

```
REGRESS     ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Alg = LAD $
```

A more general program for quantile regressions is available with the **QREG** command,

```
QREG        ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Qnt = the specific quantile $
```

The LAD results are reproduced with **; Qnt = .5**, but any other quantile may be specified instead. The LAD or median and quantile regressions are detailed in Section E9.3.

E2.10 Analysis of Variance and Panel Data

LIMDEP contains a wide variety of routines for analysis of panel data. Most of these are estimation programs for models that involve fixed or random effects specifications. Listed below is a set of functions and routines that can be used to produce descriptive statistics for a panel. In all cases, it is assumed that you have a stratification variable that takes values 1, 2, ..., G, where G is the number of groups of observations. The number of observations in each group is T_g. This is never required to be the same across groups. (See Section R5.3 for discussion of group indicators for panel data.)

E2.10.1 Analysis of Variance

One way analysis of variance is computed by regression of the variable to be analyzed on a complete set of group dummy variables. The command is

```
REGRESS     ; Lhs = the variable
            ; Rhs = one
            ; Str = the stratification variable
            ; Panel   $
```

If the number of observations is the same in every group, then you may dispense with the stratification variable and use

```
            ; Pds = the fixed number of observations
```

There are no other optional specifications for this command. The results produced are shown in the example below, where we analyze the income in the health care data set.

REGRESS ; Lhs = hhninc ; Rhs = one
; Pds = _groutppi
; Panel $ (We could use **; Str = id**.)

```
--------------------------------------------------------------------------
Analysis of Variance for         HHNINC
Stratification Variable          _STRATUM
Total Sample Size                                27326      Group Sizes
Number of Groups                                  7293      Max =     1
Number of groups with no data                        0      Min =     1
Overall Sample Mean                           .3520836      Avg =   1.0
Total Sample Minimum                          .0000000
Total Sample Maximum                         3.0671000
Sample Standard Deviation                     .1769083
Total Sample Variance                         .0312965
Source of Variation                 Variation     Deg.Fr.      Mean Square
Between Groups                .5757920920D+03       7292  .7896216292D-01
Within  Groups                .2793856074D+03      20033  .1394626903D-01
Total                         .8551776995D+03      27325  .3129653063D-01
Residual S.D.                 .1180943226D+00
R-squared                         .6733011074
F ratio                          5.6618843915     P value           .00000
--------------------------------------------------------------------------
```

The preceding analysis of variance can also be obtained with

SETPANEL ; … to set up the panel $
DSTAT ; Rhs = the variable ; Panel $

E2.10.2 Matrix Functions for Describing Panel Data

These functions are used to compute statistics for columns of data which are stratified. For example, the variable *income* might contain time series of 10 yearly observations on average family income in each of the 50 states. The number of observations would then be 500. In order to use these commands, you must provide a stratification variable. Stratification variables are described at length in Section R5.3.

Each of the functions listed below creates a matrix with number of rows equal to the largest value found for your stratification indicator. For our example above, that would be 50 since there are 50 states and our indicator would (presumably) take values 1,...,50. However, these functions do not require that all values be present in the indicator. For example, suppose our statewide data did not include states 14, 21-29, and 36. Our indicator would take 39 distinct values, but the highest value would be 50. The matrices created here would have 50 rows, but 11 rows in each one would contain zeros (not -999s). For purposes of this discussion, we will call this maximum *G*, emphasizing that *G* is only the exact number of groups if your indicator takes all of the values 1,...,*G*.

NOTE: These functions automatically bypass missing data. If any variable shows -999, the observation is omitted from any sum. If the stratification indicator is missing, the entire group is bypassed.

The results are limited to the maximum size of a matrix, 50,000 cells. The commands produce different numbers of columns, so the number of groups which can be accommodated by these commands will differ somewhat. Thus, **MATRIX ; gs = Grps(i) $** could create a 4,500×5 matrix. This function creates a single column of length *G*.

Gsiz(*indicator*) = G×1 matrix of group sizes.

These functions require a namelist or a list of *K* variables. The matrices they create each have one column for each variable and *G* rows, one for each group.

Gxbr(*list, indicator*) = G×K matrix of group means,
Gsdv(*list, indicator*) = G×K matrix of group standard deviations,
Gmax(*list, indicator*) = G×K matrix of group maxima,
Gmin(*list, indicator*) = G×K matrix of group minima,
Gsum (*list, weight, indicator*) = G×K matrix of sums, weighted by the weight variable.

You must use a namelist in the Gsum(*list, weight, indicator*) function. You may follow the namelist with the names of some variables which are also to be summed, but not to be multiplied by the weighting variable. The following function requires a single variable and the indicator, and produces a five column matrix:

Grps(*variable, indicator*) = G×5 matrix,
Column 1 = group sizes,
Column 2 = group means,
Column 3 = group standard deviations,
Column 4 = group maxima,
Column 5 = group minima.

The Gxbr function can be used to compress a panel data set into a data set of group means that you can analyze with other statistical commands. You would do so as follows:

Step 1. Define the list of variables.

NAMELIST ; old = list of variables to be compacted $

Step 2. Give the replacement list. This step is not needed if the original data can be overwritten.

NAMELIST ; new = namelist for variables to be created $

Step 3. Set up the stratification indicator if it is not already in the data set.

CREATE ; i = whatever is appropriate $

Step 4. Get the matrix of means.

MATRIX ; means = Gxbr(old, i) $

Step 5. Move group means into the data area, and pick up the number of rows.

```
CALC      ; g = Row(means) $
CREATE    ; new =  means $
SAMPLE    ; 1 - g $
```

If necessary, you might want to drop observations with empty cells. You cannot do this by selecting on zero values for the group means, since 0.0 is a valid value for this variable. But, you can do this by computing the group sizes with the Gsiz function. For a small number of groups, you can look at the matrix directly to find the empty cells. For a large number of cells, you can use the following after you have set up the sample as shown above:

```
MATRIX    ; gsize  =  Gsiz(indicator) $
CREATE    ; newg =  gsize $
REJECT    ; newg =  0 $
```

Another common operation is to create a variable which repeats the group means of that variable for each observation in a group. You can easily to this with **MATRIX**, with

```
MATRIX    ; grpmeans= Gxbr(variable, i) $
CREATE    ; means = grpmeans(i) $
```

There is also a **CREATE** command that does the same thing,

```
CREATE    ; means = Group Mean(variable, Str = i ) $
```

E2.11 Discriminant Analysis

The command for carrying out a linear discriminant analysis is

```
CLASSIFY  ; Lhs = class stratification variable = 0 for out of sample
          ; Rhs = covariates $
```

This procedure carries out a 'discriminant analysis' for a set of observations on variables $x_1,...,x_K$. The sample is divided into G+1 groups of observations identified with known classification 1,...,G or classification unknown, the 'G+1 group.' The objective is to use the data with known classification to develop a rule which is then used to make a best guess as to the appropriate classification of the unclassified (class G+1) observations. Analysis is carried out as follows: For the data in the G groups, we have prior assignment probabilities

$$\Pi^0 = (\pi_1,..., \pi_G)^0$$

These represent the *prior* classification probabilities for each observation in the sample (i.e., given no information about the covariates is used). Under most circumstances, π_g will equal 1/G, indicating no specific prior information, though we allow for others if the user has specific values to provide. If the group sizes are unequal and not randomly so, then the group proportions, N_g/N, may be a preferable prior. Each observation also has an assignment to its specific group, y_{ig} which equals either 'g' or 0 if the assignment is unknown (and will be estimated here).

For the data in the *G* groups, we first compute mean vectors and covariance matrices,

$$\overline{\mathbf{x}}_g = \frac{1}{N_g}\sum_{i=1}^{N_g}\mathbf{x}_{ig}$$

$$\mathbf{S}_g = \frac{1}{N_g}\sum_{i=1}^{N_g}(\mathbf{x}_{ig} - \overline{\mathbf{x}}_g)(x_{ig} - \overline{\mathbf{x}}_g)'$$

For each observation *i* in group *g*, $\mathbf{x}_{ig}$, the 'distance measure' (Mahalanobis distance) from the center of each group *m* is

$$d_{ig|m} = (\mathbf{x}_{ig} - \overline{\mathbf{x}}_m)'\mathbf{S}_m^{-1}(\mathbf{x}_{ig} - \overline{\mathbf{x}}_m)\,.$$

The predicted assignment for all observations, including those without prior known classification, is the one with the smallest distance;

$$\hat{y}_{ig|m} = m \text{ such that } d_{ig|m} < d_{ig|j} \text{ for } m \neq j.$$

(Note that for observations with known classification, the rule can predict incorrectly.)

When unequal prior probabilities are provided, a refinement of the prediction rule uses the ex post (posterior) probabilities:

$$\hat{\pi}_{ig|m} = K\frac{exp(\text{-}.5(d_{ig|m} + |\mathbf{S}_g|))}{\pi_m^0}$$

(The leading scalar *K* is used to make the probabilities sum to one.) Then, the predicted classification is the one with the maximum posterior probability. Note that since the computations require the inverses of the group specific covariance matrices, the entire procedure breaks down if this cannot be computed for any group.

This calculation is requested with the command

CLASSIFY ; Lhs = class stratification variable = 0 for out of sample
; Rhs = covariates

with optional specifications:

; Wts = weighting variable, used as replications
; Keep = variable to use for classification result
; List = switch to request observation specific listing
; Labels = list of labels for groups to use in displays

The default calculation allows the covariance matrix to differ among the groups. Use

; Pooled

to specify that the distance measures should based on the single full sample covariance matrix, **S**, rather than $\mathbf{S}_g$ for the specific groups. You can also control the way the distance measure is computed, with

; Var = identity

to ignore the covariance matrix, and use, simply, the distance of $\mathbf{x}_{ig|m}$ from the mean in group m. In this case,

$$d_{ig|m} = (\mathbf{x}_{ig} - \overline{\mathbf{x}}_m)'(\mathbf{x}_{ig} - \overline{\mathbf{x}}_m) = \sum_{k=1}^{K} (x_{ig,k} - \overline{x}_{m,k})^2 .$$

You may, instead, just scale the variables without rotating them – in this instance, you would use only the diagonal elements of the covariance matrix. Use

; Var = diagonal

to employ

$$d_{ig|m} = (\mathbf{x}_{ig} - \overline{\mathbf{x}}_m)'[diag(\mathbf{S}_m)]^{-1}(\mathbf{x}_{ig} - \overline{\mathbf{x}}_m) = \sum_{k=1}^{K} \frac{(x_{ig,k} - \overline{x}_{m,k})^2}{S_{m,kk}^2} .$$

Finally, the priors are formulated as follows: If no prior probabilities are specified, then

$\pi_g^0 = 1/G$ (this is the default).

If you specify simply

; Priors

then $\pi_g^0 = N_g / \Sigma_g N_g$.

Finally, you may specify your own group of priors with

; Prior = a list of G values that must sum to one.

In the example below, there are four groups of 20 observations. We specified that the firms in the last group were unidentified. There are three variables analyzed, *i,f,c*. The command is

CLASSIFY ; Lhs = j ; Rhs = i,f,c ; List $

In the listing (which is abbreviated), a ‘*’ indicates a correctly predicted group identifier. Otherwise, ‘=P’ and ‘=A’ indicates the predicted and actual group, respectively. The prior and posterior probabilities are listed as well.

```
+------------------------------------------------+
| Linear Discriminant Analysis                   |
| Full sample number of observations   =     60 |
| Sum of frequencies                   =     60 |
| Number of Classes in the sample      =      3 |
| Number of out of sample observations =     20 |
| Sum of frequencies for out of sample =     20 |
+------------------------------------------------+
   Class  Sample  Sum of wts  Proportion  Prior P
J=1           20          20       .3333    .3333
J=2           20          20       .3333    .3333
J=3           20          20       .3333    .3333
```

```
Analysis by observation             Posterior (6 - ... not shown)
Indiv  Actual   Predicted   Prior    J=1     J=2      J=3
-----  -------- --------- ------   -------- -------- -------- -------- --------
    1  J=1       J=1       .3333  .9991=*  .0000    .0009
    2  J=1       J=1       .3333  .9996=*  .0000    .0004
    3  J=1       J=1       .3333  .9967=*  .0000    .0033
    4  J=1       J=1       .3333  .9993=*  .0000    .0007
   ...
   20  J=1       J=1       .3333  .9850=*  .0000    .0150
   21  J=2       J=1       .3333  .7500=P  .0000=A  .2500
   22  J=2       J=3       .3333  .4780    .0000=A  .5220=P
   23  J=2       J=1       .3333  .9354=P  .0000=A  .0646
   24  J=2       J=1       .3333  .9804=P  .0000=A  .0196
   ...
   40  J=2       J=1       .3333  .9852=P  .0000=A  .0148
   41  J=3       J=1       .3333  .9127=P  .0000    .0873=A
   42  J=3       J=1       .3333  .9928=P  .0000    .0072=A
   43  J=3       J=1       .3333  .9969=P  .0000    .0031=A
   44  J=3       J=1       .3333  .9908=P  .0000    .0092=A
   ...
   60  J=3       J=3       .3333  .4062    .0000    .5938=*
   61  =>none<=  J=3       .0000  .0871    .0000    .9129=P
   62  =>none<=  J=1       .0000  .5106=P  .0000    .4894
   63  =>none<=  J=1       .0000  .6288=P  .0000    .3712
   64  =>none<=  J=3       .0000  .0999    .0000    .9001=P
   ...
   76  =>none<=  J=3       .0000  .3261    .0000    .6739=P
   80  =>none<=  J=3       .0000  .2765    .0000    .7235=P

Classification Results.  Total Frequencies Based on Weights if Any
              Actual                                              Out of
Predicted       J=1       J=2       J=3  =>none<=  Not Used Total   Sample
      J=1       20.       15.       19.        5.        0.    59.      0.
      J=2        0.        0.        0.        0.        0.     0.      0.
      J=3        0.        5.        1.       15.        0.    21.      0.
    Total       20.       20.       20.       20.        0.    80.      0.
```

E2.12 Accuracy and the NIST Benchmarks

The National Institute of Standards and Technology (NIST) has compiled a set of accuracy benchmarks for statistical software, the *Statistical Reference Datasets* (StRD at http://www.itl.nist.gov/div898/strd/), which can be used for testing the accuracy of programs such as *LIMDEP*. There are (as of this writing) five sets of problems: univariate summary statistics, analysis of variance, linear regression, nonlinear regression and Markov Chain Monte Carlo estimation. The problems are designed to test different aspects of computation and present varying levels of difficulty. We will be presenting some of the test problems in this manual, primarily to verify the program accuracy, but also to demonstrate the variety of problems that they present. McCullough (1999) presents a detailed analysis of the datasets with several programs, including *LIMDEP*, and suggests a routine method of measuring accuracy. For the first three suites, as will be seen below, *LIMDEP* matches the NIST standard for all visible digits, so accuracy is not a consideration. Some of McCullough's analysis will be presented with the nonlinear least squares suite, where there is much more variation.

E2.12.1 NIST Benchmarks for Univariate Statistics

In the examples below, the source level NIST problem statement is presented with the *LIMDEP* solution to the problem. In some cases, only a few of the data observations are listed with the problem, so as to suggest their appearance. Many of the NIST datasets are included with the *LIMDEP* program, and can be found in the NIST Benchmarks book of the Help file and also in the C:\LIMDEP10\Command Files folder created with program installation. The full NIST datasets can be downloaded from the NIST website. The following shows two of the datasets from the univariate summary statistics suite. The first, the Maryland lottery problem is rated 'lower level of difficulty.' One of the problem sets in this suite, NumAcc4 (Numerical Accuracy #4) is rated 'Higher Level of Difficulty.' (This is the top of three levels in the suites.) This problem involves 1,001 observations (actually 500 occurrences each of 1.00000001 and 1.00000003, and one of 1.00000002. The certified values and *LIMDEP*'s results for this data set are shown below.

```
Dataset Name:  Maryland Pick-3 Lottery
Description:   This is an observed/"real world" data set
               consisting of 218 Maryland Pick-3 Lottery values
               from September 3, 1989 to April 14, 1990 (32 weeks).
               One 3-digit random number (from 000 to 999)
               is drawn per day, 7 days per week for most
               weeks, but fewer days per week for some weeks.
               We use these data here to test accuracy
               in summary statistics calculations.
Stat Category: Univariate: Summary Statistics
Reference:     None
Data:          "Real World"
               1    Response          : y = 3-digit random number
               0    Predictors
               218  Observations
Model:         Lower Level of Difficulty
               2    Parameters        : mu, sigma
               1    Response Variable : y
               0    Predictor Variables
               y    = mu + e
                                                  Certified Values
Sample Mean                                 ybar:  518.958715596330
Sample Standard Deviation (denom. = n-1)       s:  291.699727470969
Sample Autocorrelation Coefficient (lag 1) r(1):  -0.120948622967393
Number of Observations:                              218
Data: Y
```

READ ; Nobs = 218 ; Nvar = 1 ; Names = y ; By Variables $

```
162 671 933 414 788 730 817  33 536 875 670 236 473 167 877 980 316 950
456  92 517 557 956 954 104 178 794 278 147 773 437 435 502 610 582 780
689 562 964 791  28  97 848 281 858 538 660 972 671 613 867 448 738 966
139 636 847 659 754 243 122 455 195 968 793  59 730 361 574 522  97 762
431 158 429 414  22 629 788 999 187 215 810 782  47  34 108 986  25 644
829 630 315 567 919 331 207 412 242 607 668 944 749 168 864 442 533 805
372  63 458 777 416 340 436 140 919 350 510 572 905 900  85 389 473 758
444 169 625 692 140 897 672 288 312 860 724 226 884 508 976 741 476 417
831  15 318 432 241 114 799 955 833 358 935 146 630 830 440 642 356 373
271 715 367 393 190 669   8 861 108 795 269 590 326 866  64 523 862 840
219 382 998   4 628 305 747 247  34 747 729 645 856 974  24 568 24  694
608 480 410 729 947 293  53 930 223 203 677 227  62 455 387 318 562 242
428 968
```

DSTAT ; Rhs = y ; AR1 $

Maryland Lottery Results

```
Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|    Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+---------------------------------------------------------------------
       Y|  518.959      291.700            4          999          218       0
        |  Autocorrelation -.120948623
--------+---------------------------------------------------------------------
```

NumAcc4 Results

```
The NIST Certified True (exact values) are
          10000000.2      .100000000 and  -.999 for the autocorrelation.
Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|    Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+---------------------------------------------------------------------
       Y|   10000000.2   .100000001  .100000D+08  .100000D+08     1001       0
        |   Autocorrelation -.998999999
--------+---------------------------------------------------------------------
```

LIMDEP's results agree with NIST to the visible digits in the results. (The AR1 specification computes a first order autocorrelation coefficient. This is discussed in Chapter E5.)

E2.12.2 Accuracy in ANOVA Computations – The NIST Benchmarks

The listing below displays the analysis of variance for one of the 11 datasets in the NIST/StRD suite to be used for analysis of panel data. Several additional datasets are included in the NIST Benchmarks book of the Help file and also in the C:\LIMDEP10\Command Files folder created with program installation.

```
File Name:      NIST-ANOVA-atomic.lim
Dataset Name:   Atomic Weight of Silver   (NIST-atomic.dat)
File Format:    ASCII
                Certified Values   (lines 41 to 47)
                Data               (lines 61 to 108)
Procedure:      Analysis of Variance
Reference:      Powell, L.J., Murphy, T.J. and Gramlich, J.W. (1982).
                "The Absolute Isotopic Abundance & Atomic Weight
                of a Reference Sample of Silver".
                NBS Journal of Research, 87, pp. 9-19.
Data:           1 Factor
                2 Treatments
                24 Replicates/Cell
                48 Observations
                7 Constant Leading Digits
                Average Level of Difficulty
                Observed Data
Model:          3 Parameters (mu, tau_1, tau_2)
                y_{ij} = mu + tau_i + epsilon_{ij}
Certified Values:
Source of                 Sums of                Mean
Variation     df      Squares                Squares              F Statistic
Between       1    3.63834187500000E-09 3.63834187500000E-09 1.59467335677930E+01
Within       46    1.04951729166667E-08 2.28155932971014E-10
                   Certified R-Squared 2.57426544538321E-01
                   Certified Residual Standard Deviation     1.51048314446410E-05
Data:  Instrument           AgWt
```

Read the data first. The ‘By Variables’ form is convenient when the data set is a single variable.

READ ; Nobs = 48 ; Nvar = 2 ; Names = i, y ; By Variables $

```
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2
107.8681568 107.8681465 107.8681572 107.8681785 107.8681446 107.8681903
107.8681526 107.8681494 107.8681616 107.8681587 107.8681519 107.8681486
107.8681419 107.8681569 107.8681508 107.8681672 107.8681385 107.8681518
107.8681662 107.8681424 107.8681360 107.8681333 107.8681610 107.8681477
107.8681079 107.8681344 107.8681513 107.8681197 107.8681604 107.8681385
107.8681642 107.8681365 107.8681151 107.8681082 107.8681517 107.8681448
107.8681198 107.8681482 107.8681334 107.8681609 107.8681101 107.8681512
107.8681469 107.8681360 107.8681254 107.8681261 107.8681450 107.8681368
```

One way analysis of variance is computed by ‘regressing’ a variable on a constant term using the panel data format.

REGRESS ; Lhs = y ; Rhs = one ; Str = i ; Panel $

```
--------------------------------------------------------------------------
Analysis of Variance for         Y
Stratification Variable          _STRATUM
Total Sample Size                                    48      Group Sizes
Number of Groups                                      2      Max =    24
Number of groups with no data                         0      Min =    24
Overall Sample Mean                         107.8681451      Avg =  24.0
Total Sample Minimum                        107.8681079
Total Sample Maximum                        107.8681903
Sample Standard Deviation                      .0000173
Total Sample Variance                          .0000000
Source of Variation              Variation     Deg.Fr.       Mean Square
Between Groups               .3638341875D-08          1  .3638341875D-08
Within  Groups               .1049517292D-07         46  .2281559330D-09
Total                        .1413351479D-07         47  .3007130807D-09
Residual S.D.                .1510483144D-04
R-squared                        .2574265445
F ratio                        15.9467335680     P value          .00001
--------------------------------------------------------------------------
```

The *LIMDEP* results are accurate to the 10 digits displayed in the results.

E3: Histograms and Kernel Density Estimators

E3.1 Introduction

This chapter describes methods of describing the empirical distribution of a variable. The tools provided are:

- Normal quantile plots to compare the empirical CDF to the normal distribution,
- Histograms for continuous data to describe the distribution,
- Histograms for discrete data to provide frequency counts and characterize the distribution,
- Kernel density estimators,
- Tests for normality based on moments (Bowman-Shenton) and on the CDF (Kolmogorov-Smirnov).

E3.2 Normal-Quantile Plots

A normal-quantile, or N-Q plot compares the within sample cumulative distribution of a variable to what would have been expected if the data were drawn from a normal population. The calculations are as follows:

1. Data on $x_1,...,x_n$ are sorted in ascending order into $x_{(i)}$.
2. For $i = 1,...,n$, compute $c_i = (i - .5)/n$, $t_i = \Phi^{-1}(c_i)$, $z_i = s_k t_i + \bar{x}_k$.

Thus, z_i is the counterpart to $x_{(i)}$ from the normal distribution with the same mean and standard deviation as the sample of *x*s. We then plot *x* and *z* against *z*. This produces a scatter plot plus a straight line (*z* vs. *z*). The larger the deviation of the scatter from the line, the greater the departure from normality. To obtain this additional output, add

; Plot

to the **DSTAT** command. An example based on the income variable in the health care data used earlier is shown in Figure E3.1. Income is highly skewed, so the large departure from normality in the left figure is to be expected. Taking logs of income or wealth variables usually renders them normally distributed, or at least approximately so.

DSTAT ; Rhs = hhninc, Log(hhninc) ; Plot $

The figure on the right suggests that the lower tail of the log of *hhninc* does not approximate normality very well.

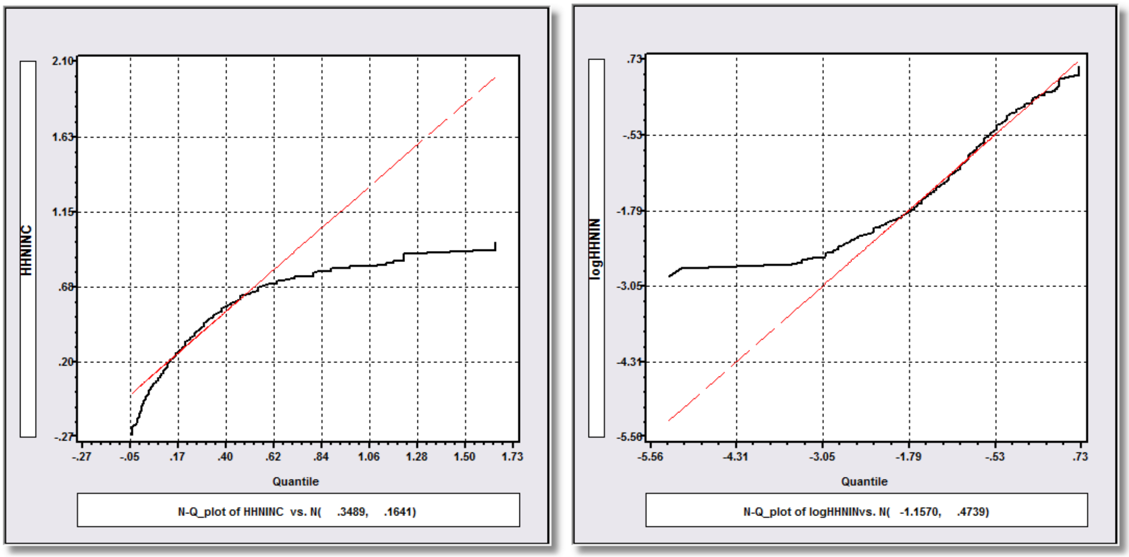


Figure E3.1 Normal-Quantile Plots

Figure E3.2 shows the appearance of an N-Q plot for a variable that is precisely normally distributed.

```
CREATE    ; z = Rnn(0,1) $
DSTAT     ; Rhs = z ; Plot $
```

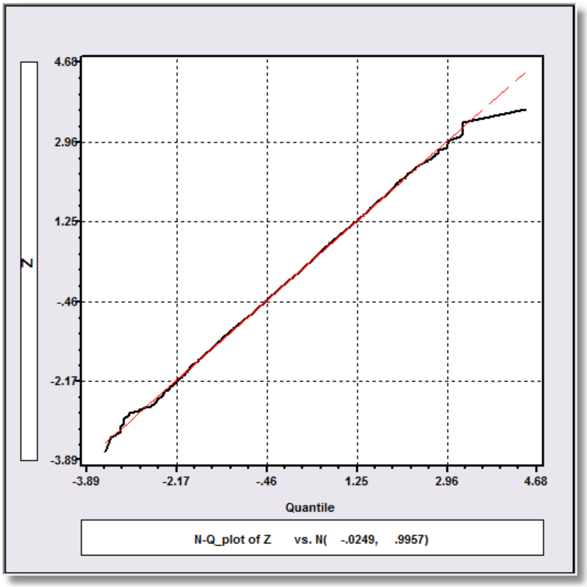


Figure E3.2 N-Q Plot for Normally Distributed Variable

E3.3 HISTOGRAM Command

The basic command for computing and plotting a histogram for a variable is

HISTOGRAM ; Rhs = the variable $

A title for the figure may be provided by using

; Title = ...< the desired title, up to 60 characters> ...

To use the command builder for the **HISTOGRAM** command, select Data Description from the Model menu, then select Histogram. The Main page of the **HISTOGRAM** command builder is shown in Figure E3.3.

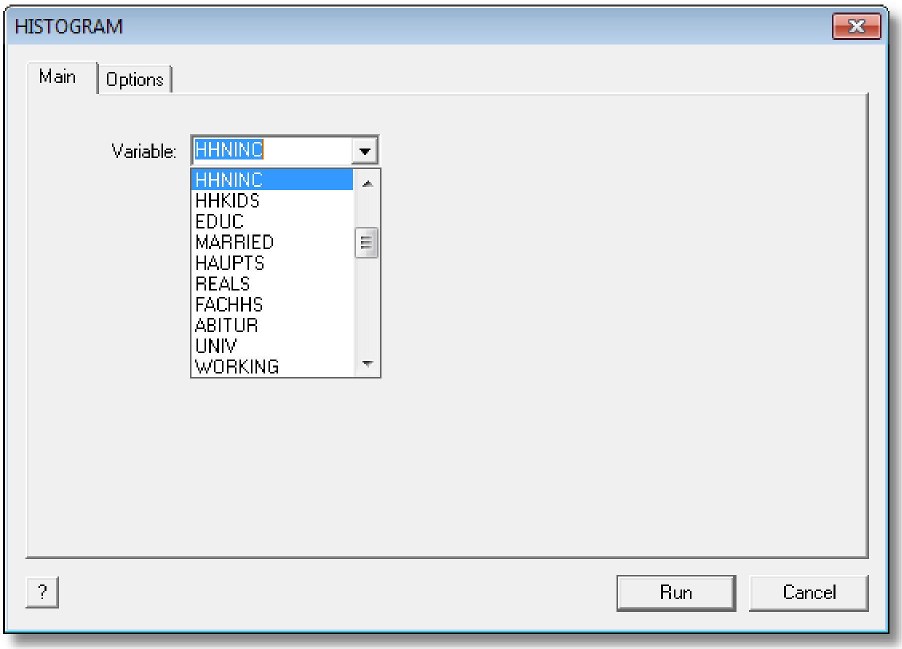


Figure E3.3 Main Page of Command Builder for HISTOGRAM

The Options page of the command builder for **HISTOGRAM** is shown in Figure E3.4.

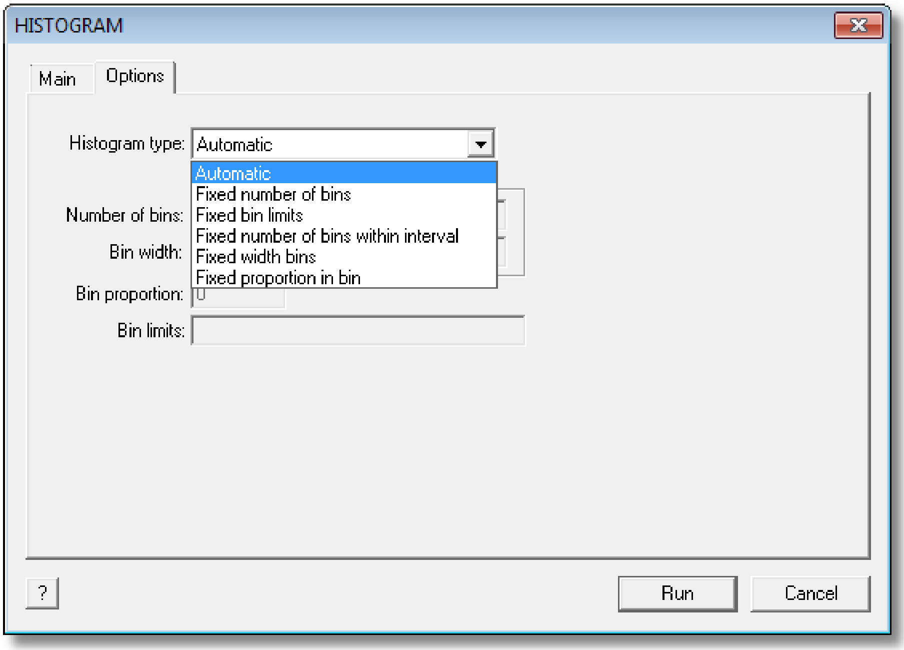


Figure E3.4 Options Page of Command Builder for HISTOGRAM

LIMDEP computes histograms for continuous and discrete (count) data. In the default figure for continuous data, the values are assigned to 40 equal width intervals over the range of the variable. (The number of and widths of the bins in the figure may be changed.) The histogram for continuous data provides a descriptive device to illustrate the distribution of the variable. (The kernel density estimator described in Section E3.6 may be better suited for this purpose.) The default figure for discrete data is a frequency count. Data are assumed to be coded 0,1,...,499. Values less than zero or greater than 499 are treated as out of range. A count of invalid observations is given with the output of the command. The histogram can be accompanied by a table listing the relative and cumulated frequencies.

To illustrate the use of this feature, we use the health care data set. (See Section E2.4.) The data, which will be used in several applications below, are an unbalanced panel of observations on health care utilization by 7,293 individuals. The group sizes in the panel number as follows: T_i: 1=1525, 2=1079, 3=825, 4=926, 5=1051, 6=1000, 7=887. There are altogether 27,326 observations. Some of the variables in the file are

hhninc	=	household nominal monthly net income in German marks / 10000.
hhkids	=	children under age 16 in the household = 1, otherwise = 0,
educ	=	years of schooling
married	=	marital status
female	=	1 for female, 0 for male
docvis	=	number of visits to the doctor
doctor	=	number of doctor visits > 0
hospvis	=	number of visits to the hospital

E3.4 Histograms for Continuous Data

A histogram for the continuous variable *hhninc* would appear as shown in Figure E3.5:

HISTOGRAM ; Rhs = hhninc $

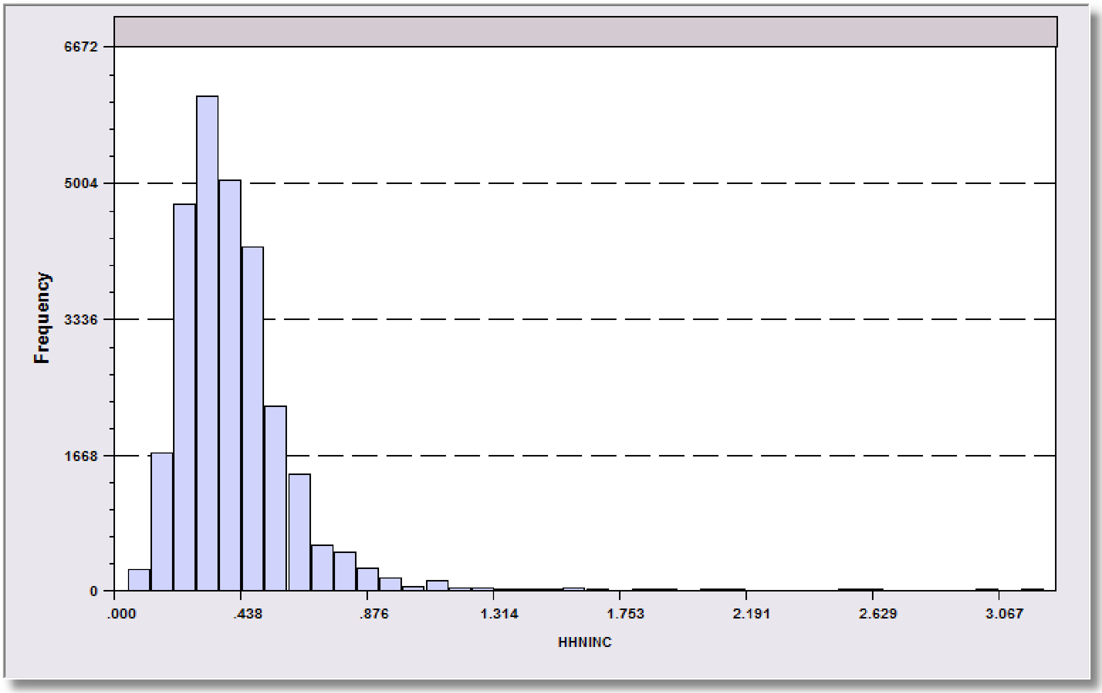


Figure E3.5 Histogram for a Continuous Variable

The displayed output from this command consists only of the figure containing the histogram. You can request a listing of the bin boundaries and frequency counts by adding

; List

to the **HISTOGRAM** command. The listing below would be produced for this histogram.

```
Histogram for HHNINC    NOBS =   27326, Too low:     0, Too high:     0
Bin is defined as Lower <= X < Upper.
Bin  Lower limit   Upper limit       Frequency     Cumulative Frequency
========================================================================
  0        .000          .767        252 ( .0092)         252( .0092)
  1        .767         1.534       1683 ( .0616)        1935( .0708)
  2       1.534         2.300       4725 ( .1729)        6660( .2437)
  3       2.300         3.067       6062 ( .2218)       12722( .4656)
  4       3.067         3.834       5023 ( .1838)       17745( .6494)
  5       3.834         4.601       4208 ( .1540)       21953( .8034)
  6       4.601         5.367       2253 ( .0824)       24206( .8858)
  7       5.367         6.134       1414 ( .0517)       25620( .9376)
  8       6.134         6.901        554 ( .0203)       26174( .9578)
  9       6.901         7.668        454 ( .0166)       26628( .9745)
 10       7.668         8.435        264 ( .0097)       26892( .9841)
 11       8.435         9.201        154 ( .0056)       27046( .9898)
 12       9.201         9.968         37 ( .0014)       27083( .9911)
 13       9.968        10.735        110 ( .0040)       27193( .9951)
 14      10.735        11.502         26 ( .0010)       27219( .9961)
 15      11.502        12.268         34 ( .0012)       27253( .9973)
 16      12.268        13.035         15 ( .0005)       27268( .9979)
 17      13.035        13.802          5 ( .0002)       27273( .9981)
 18      13.802        14.569          9 ( .0003)       27282( .9984)
 19      14.569        15.336         21 ( .0008)       27303( .9992)
 20      15.336        16.102          2 ( .0001)       27305( .9992)
 21      16.102        16.869          0 ( .0000)       27305( .9992)
 22      16.869        17.636          1 ( .0000)       27306( .9993)
 23      17.636        18.403          2 ( .0001)       27308( .9993)
 24      18.403        19.169          0 ( .0000)       27308( .9993)
 25      19.169        19.936          2 ( .0001)       27310( .9994)
 26      19.936        20.703          8 ( .0003)       27318( .9997)
 27      20.703        21.470          0 ( .0000)       27318( .9997)
 28      21.470        22.236          0 ( .0000)       27318( .9997)
 29      22.236        23.003          0 ( .0000)       27318( .9997)
 30      23.003        23.770          0 ( .0000)       27318( .9997)
 31      23.770        24.537          2 ( .0001)       27320( .9998)
 32      24.537        25.304          1 ( .0000)       27321( .9998)
 33      25.304        26.070          0 ( .0000)       27321( .9998)
 34      26.070        26.837          0 ( .0000)       27321( .9998)
 35      26.837        27.604          0 ( .0000)       27321( .9998)
 36      27.604        28.371          0 ( .0000)       27321( .9998)
 37      28.371        29.137          1 ( .0000)       27322( .9999)
 38      29.137        29.904          0 ( .0000)       27322( .9999)
 39      29.904        30.671          4 ( .0001)       27326(1.0000)
```

E3.4.1 Fixed Number of Bins

You can select the number of bars to plot with

; Int = k

This can produce less than satisfactory results, however. For the example above, we add

; Int = 10

to obtain the following:

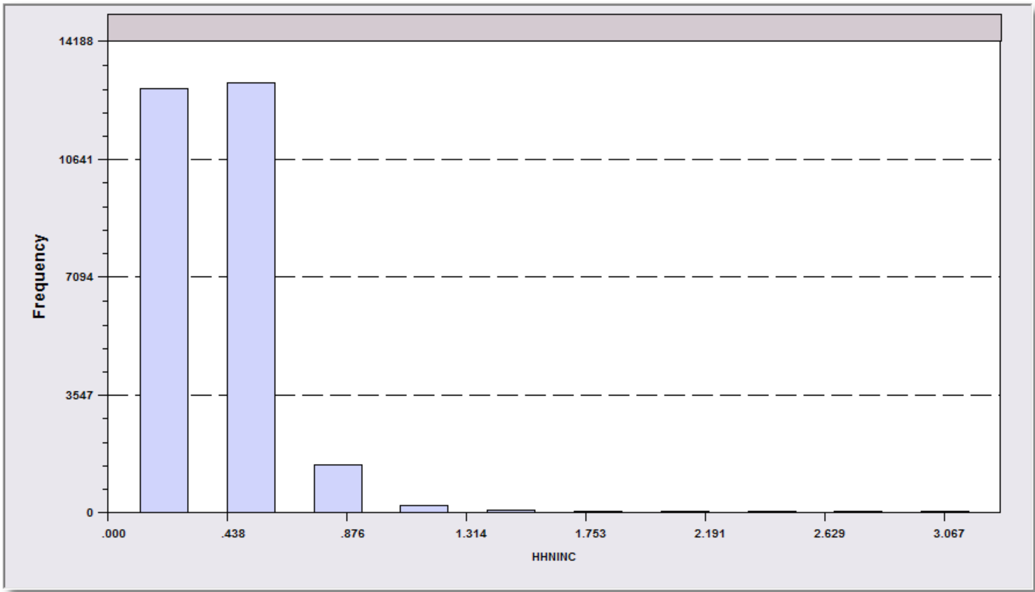


Figure E3.6 Histogram with Fixed Number of Bins

The problem is that the skewness of the income distribution has placed many observations out in the right tail. In order to accommodate them, *LIMDEP* has created several bins that have few observations in them. Fixing the number of bins may also cause some observations of a discrete variable to be out of range if you have values that exceed *K*-1. (The values for *K* intervals are 0,1,...,*K*-1.) For continuous variables, this specification requests that the range of the variable be divided into *K* equal length parts. Obtaining a satisfactory representation of the distribution may take some experimentation. For the preceding application, using 100 bins instead of 10 produces the results in Figure E3.7, which seems reasonable.

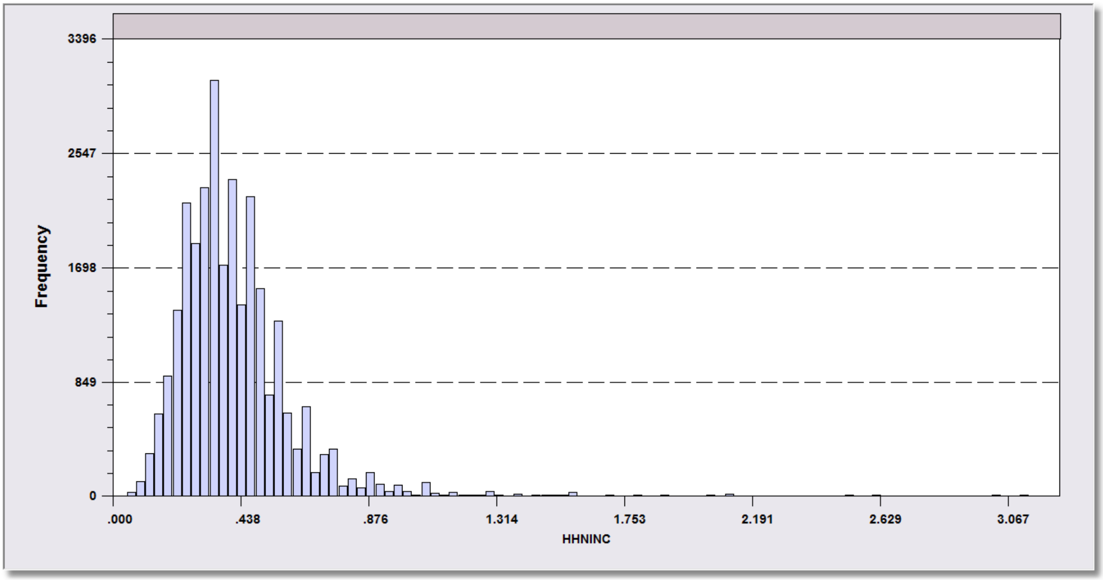


Figure E3.7 Histogram for Household Income

E3.4.2 Trimming Data for Histograms

The figure is still being substantially influenced by the long thin tail of the distribution. By trimming the extreme observations, the figure can sometimes be improved. Figure E3.8 shows the result. We have also added a title to the figure

HISTOGRAM ; If [hhninc <= 1.6] ; Rhs = hhninc
; Int = 50
; Title = Household Income (Trimmed: Less than 1.6) $

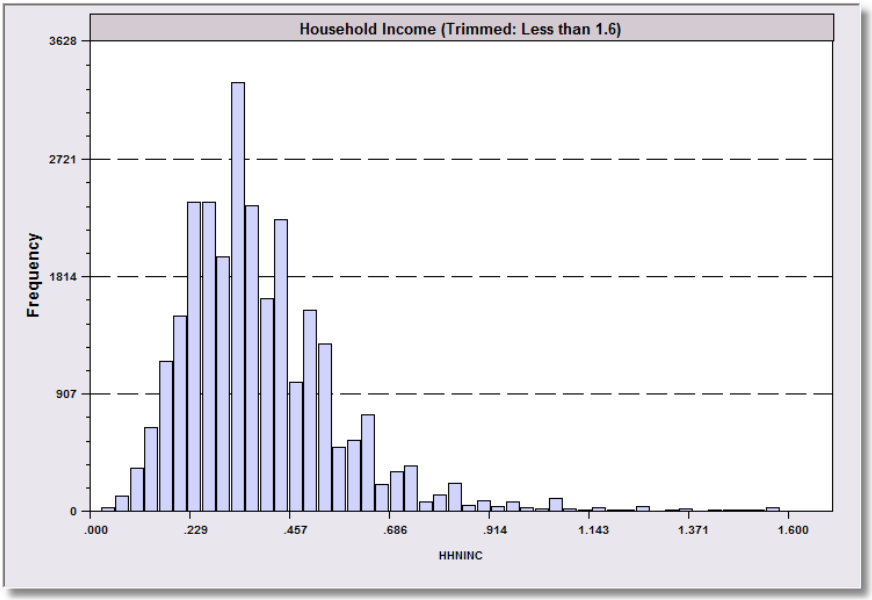


Figure E3.8 Histogram for Trimmed Data

E3.4.3 Fixed Bin Limits

There are other ways to examine continuous data. One way is to use **RECODE** to change your continuous variable into a discrete one. Alternatively, you may provide a set of interval limits and request a count of the observations in the intervals you define. The command would be

HISTOGRAM ; Rhs = variable ; Limits = l0,l1,...,lk $

where the limits you give are the left boundaries of the intervals. Thus, the number of limits you provide gives the number of intervals. Intervals are defined as 'greater than or equal to lower' and 'less than upper.' With this specification, the rightmost upper limit is +∞. For example, still using our income data,

HISTOGRAM ; Rhs = educ ; Limits = 8,10,12,14,16,18 $

defines six bins for the histogram, with the rightmost containing all values greater than or equal to 18. The result appears in Figure E3.9.

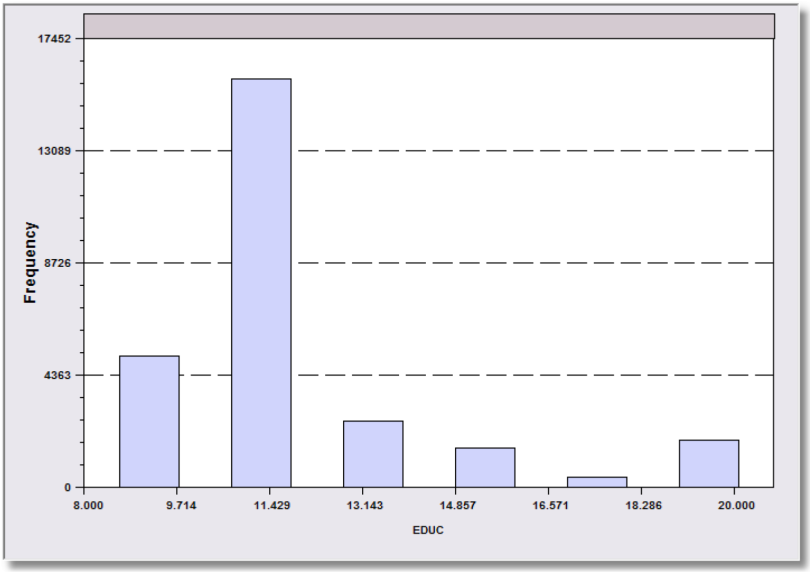


Figure E3.9 Histogram with Fixed Bin Limits

If you wish to avoid having data discarded, provide a very large negative value (-1.D10) for the lowest limit – the highest interval is assumed to be open. Otherwise, observations lower than the lowest value in the list are treated as out of range.

E3.4.4 Fixed Number of Bins in a Range

To request *K* equal length intervals in the range *lower* to *upper*, use

HISTOGRAM ; Rhs = variable ; Int = k ; Limits = lower,upper $

We'll use this device to remove some of the long tail of the income distribution.

HISTOGRAM ; Rhs = hhninc ; Int = 20 ; Limits = 0, 1.2 $

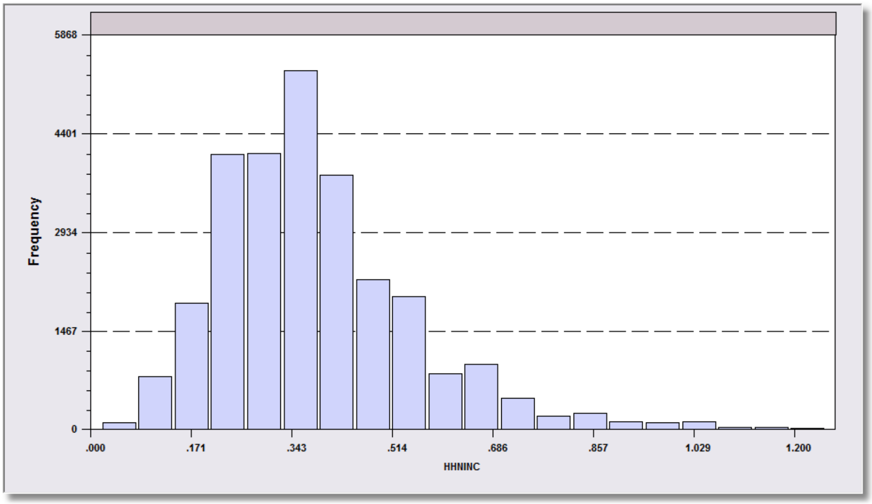


Figure E3.10 Histogram with Fixed Number of Bins in a Specified Range

E3.4.5 Fixed Width Bins in a Range

To specify both the range of variation and the fixed width of the bins, use

HISTOGRAM ; Rhs = variable ; Limits = lower (width) upper $

For example,

HISTOGRAM ; Rhs = hhninc ; Limits = 0 (.05) 3 $

specifies bins [0,.05), [.05,.10), … [2.95,3.00].

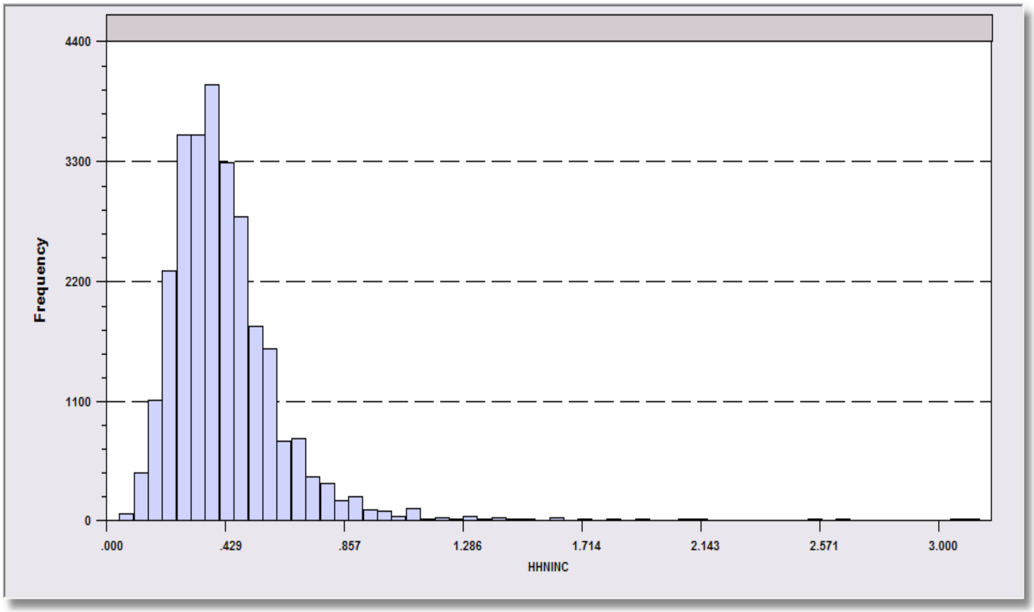


Figure E3.11 Histogram with Fixed Width Bins in a Specified Range

It is possible that the range given does not produce an even number of fixed width bins. In this case, the rightmost bin is shortened to use the remaining range. For example, in the preceding, if the .05 were .07, then there would be a narrow bin at the right; the 43rd bin would be from 2.94 to 3.00

E3.4.6 Fixed Interval Widths

Alternatively, you can specify the width of the interval, and allow the program to determine the number. Do so with

HISTOGRAM ; Rhs = variable ; Width = the desired value $

For example, the histogram for *hhninc* computed earlier could be obtained with

HISTOGRAM ; Rhs = hhninc ; Width = 0.05 $

E3.4.7 Fixed Proportion of Observations in Each Bin

NOTE: If this specification results in more than 499 intervals, a diagnostic will result.

Finally, you can use **HISTOGRAM** to search for the interval limits instead of the frequency counts. The command

HISTOGRAM ; Rhs = variable ; Bin = p $

where '*p*' is a sample fraction (proportion), will obtain the interval boundaries such that each bin contains the specified proportion of the observations.

NOTE: If the specified proportion does not lead to an even set of bins, then an extra, smaller bin is created if the remaining proportion is more than $p/2$. For example, if p is .22, there will be four bins with .22 and one at the right end with .12. But, if the extra mass is less than $p/2$, it is simply added to the rightmost bin, as for p = .16, for which the sixth bin will contain .2 of the observations.

E3.4.8 Comparison to a Normal Distribution

A common exercise is to compare the distribution of a sample to a normal distribution. Add

; Normal

to the **HISTOGRAM** command to produce a normal density superimposed on the histogram. The normal distribution plotted has the same mean and standard deviation of the data. The normal density is plotted within the range of the data. If the sample data are skewed or have extreme observations, this may result in a truncation of the normal distribution. Figure E3.12 shows an example. The data are trimmed to produce the figure. The command is

```
HISTOGRAM ; If [hhninc <= 1.6]
          ; Rhs = hhninc
          ; Int = 50
          ; Title = Household Income (Trimmed: Less than 1.6)
          ; Normal $
```

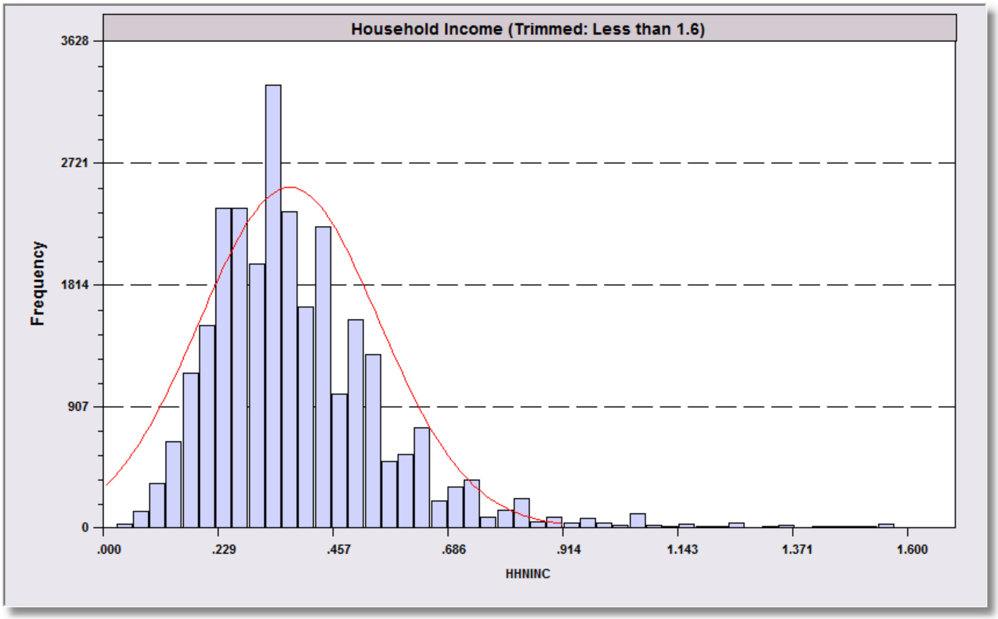


Figure E3.12 Histogram with Normal Distribution

E3.5 Histograms for Discrete Data

The data are first inspected to determine the type and the correct number of bars to plot for a discrete variable. For a discrete variable, the plot can be exact. Up to 500 bars may be displayed: For example, the count of doctor visits in the health care data appear as follows:

HISTOGRAM ; Rhs = docvis
; Title = Number of Doctor Visits $

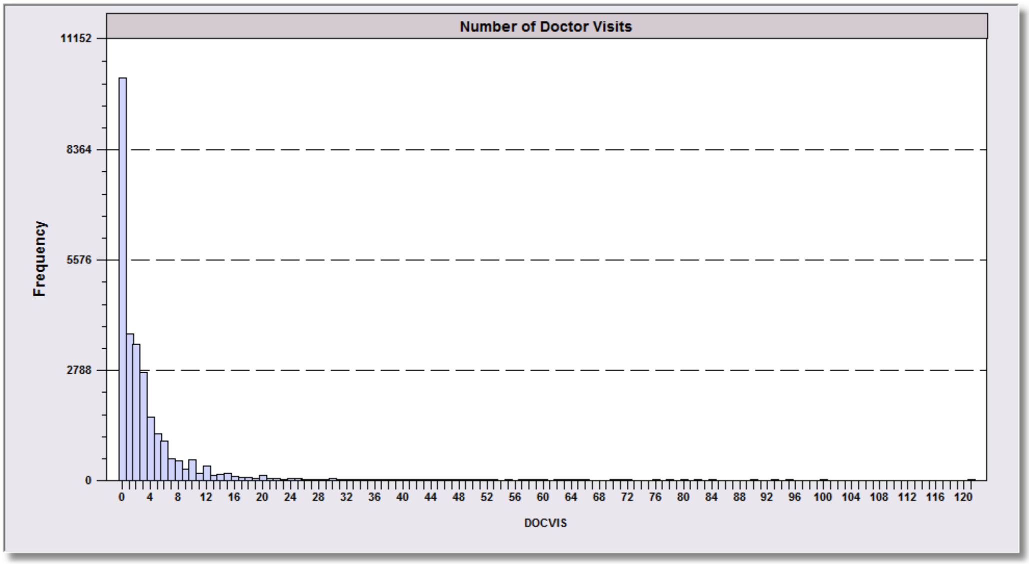


Figure E3.13 Histogram for a Discrete Variable

The long tail of the skewed distribution has rather distorted the figure. The options described earlier can be used to modify the figure. However, those options are assumed to be used for continuous data, which would distort the figure in another way. A more straightforward way to deal with the preceding situation is to operate on the data directly. For example,

```
HISTOGRAM ; If [docvis <= 25]
          ; Rhs = docvis
          ; Title = Number of Doctor Visits $
```

truncates the distribution, but produces a more satisfactory picture of the frequency count.

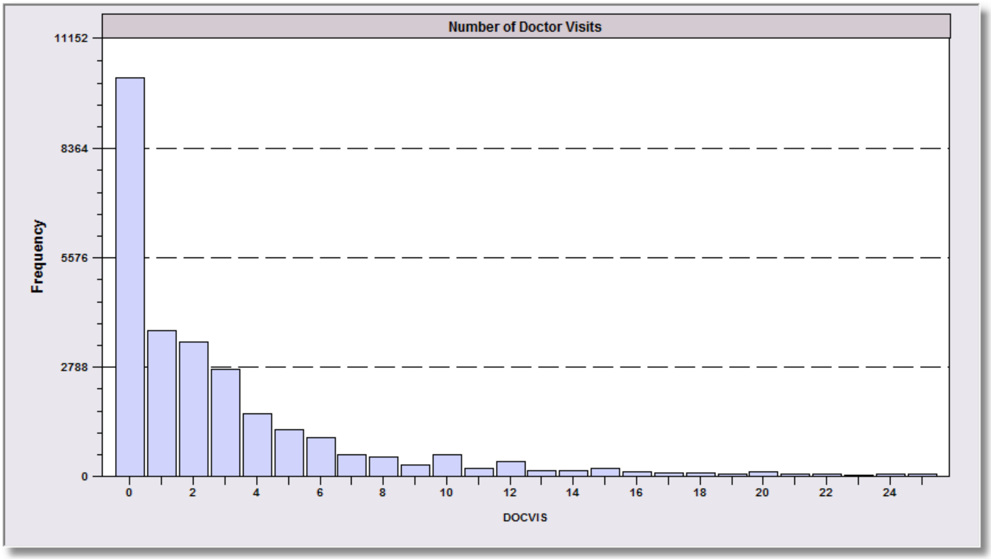


Figure E3.14 Histogram for a Truncated Discrete Distribution

E3.5.1 Bin Labels Scaled to Sample Proportions

Proportions instead of raw frequencies may be plotted by using

; Proportions

The plot is now sample proportions instead of raw frequencies. This affects the labeling of the figure but not its appearance – it will now resemble the density for the variable being plotted. For example, with this option, Figure E3.14 becomes E3.15. Note, the sum of the relative proportions shown in the figure equals 1.0. However, the bars are not scaled to make the areas sum to one.

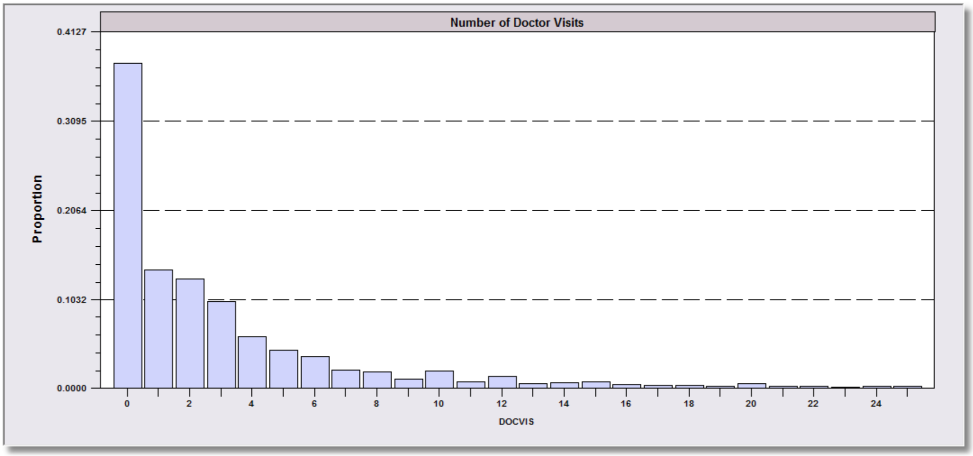


Figure E3.15 Histogram with Relative Frequencies

E3.5.2 Multiple Histograms

To plot up to four histograms in one figure, use

HISTOGRAM ; Rhs = var1, var2 (up to 4) ; All $

An example is shown below in Figure E3.16. Note that if **; All** is omitted, a separate histogram is produced for each variable. Multiple histograms are limited to 40 bins, Figure E3.16 shows a histogram for the two count variables of interest in the health study.

HISTOGRAM ; If [docvis <= 20 & hospvis <= 20] ; Rhs = docvis,hospvis ; All
; Title = Histogram for Hospital Visits and Doctor Visits $

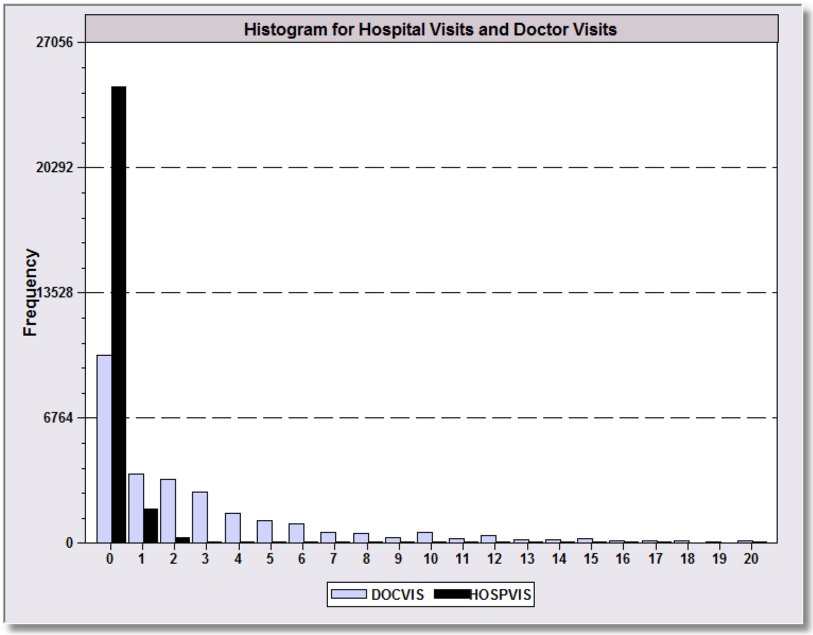


Figure E3.16 Histogram for Two Variables

E3.5.3 Stratification

The specification

; Group = stratification variable
(up to four groups instead of four variables)

may be used to produce the same sort of multiple plot figure, where the separate histograms correspond to different groups. The groups are assigned labels '*name*001,' '*name*002' etc. You may provide your own labels with

; Labels = labels for the groups

The example in Figure E3.17 below is produced using

```
HISTOGRAM ; If [docvis <= 25]
          ; Rhs = docvis
          ; Group = female
          ; Labels = female,male
          ; Title = Hospital Visits Male vs. Female $
```

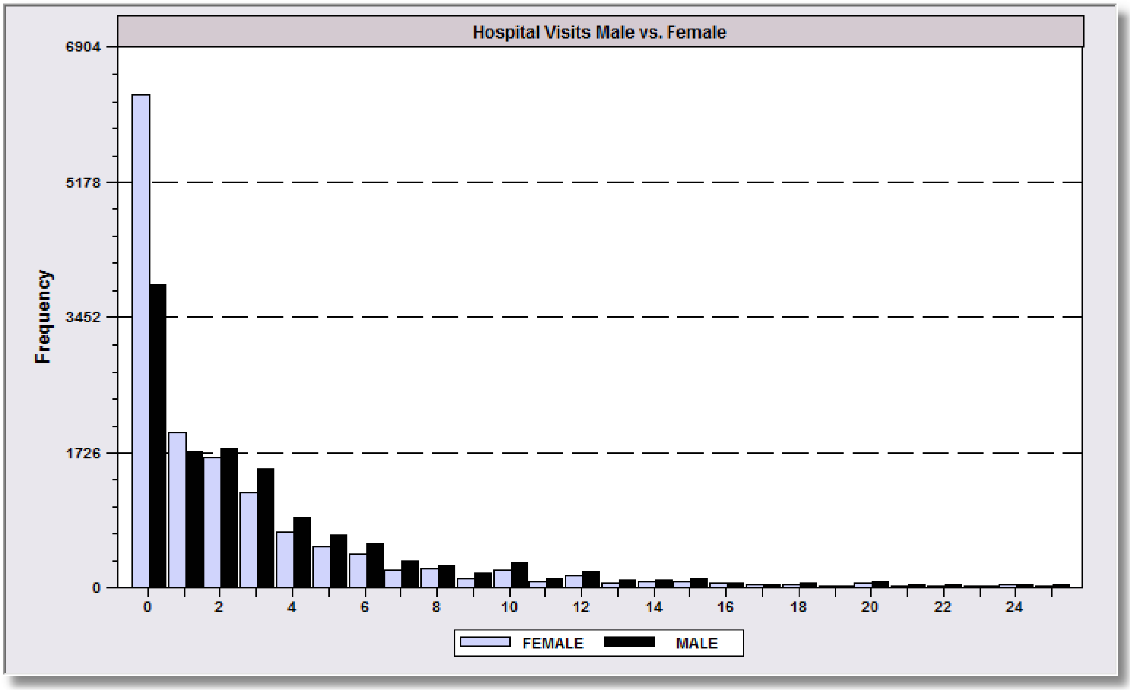


Figure E3.17 Histograms for Two Strata

E3.5.4 Labels for Bins

Bins are generally labeled '1,' '2,' etc. as in the examples above. You may provide your own labels for the bins with

; Choices = labels for the bins, up to 15 bin labels

The following example combines several of the features described above for a data set on mode choice. (The group indicator, *sex*, was simulated for this example – it is not present in the original data set.) The clogit data are used in Chapter E38 to illustrate multinomial choice models.

```
CREATE    ; choice = Trn(-4,0) * mode $
REJECT    ; choice = 0 $
CREATE    ; sex = Rnu(0,1) > .55 $
HISTOGRAM ; Rhs = choice
          ; Choices  = air,train,bus,car
          ; Title = Mode Choice: Sydney-Melbourne Commute
          ; Group = sex
          ; Labels = male,female $
```

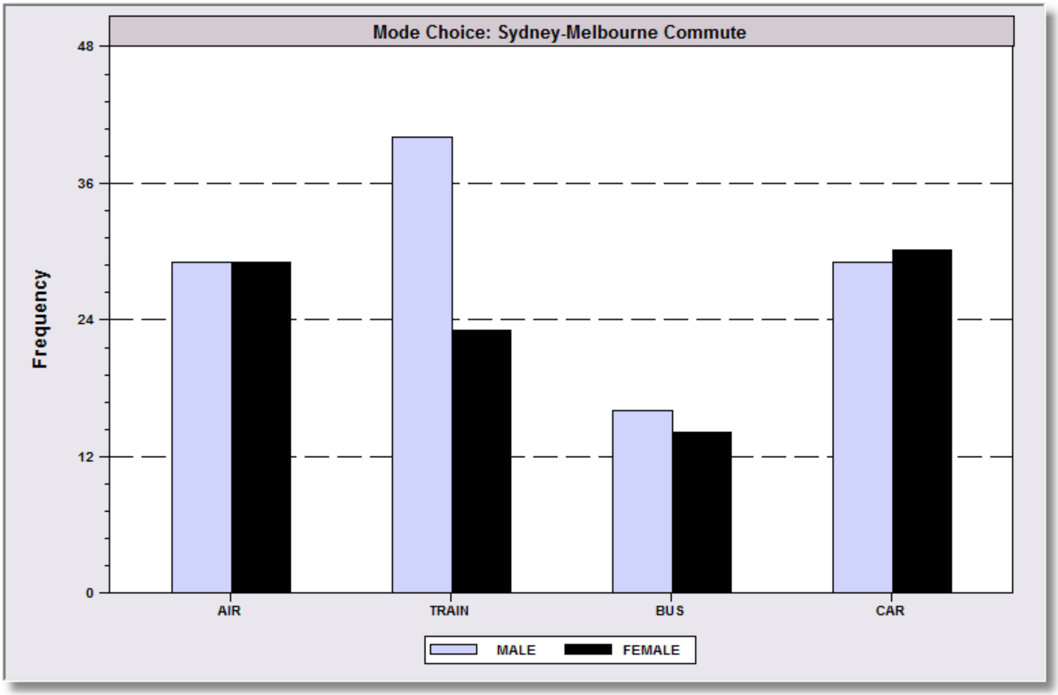


Figure E3.18 Histogram with Strata, Bin Labels and Title

E3.6 Kernel Density Estimation

The basic command for computing and plotting a kernel density estimate for a variable is

KERNEL ; Rhs = the variable $

A title for the figure may be provided by using

; Title = ...< the desired title, up to 60 characters> ...

The kernel density estimator is a device used to describe the distribution of a variable nonparametrically, that is, without any assumption of the underlying distribution. The kernel density function for a single variable is computed using

$$f(z_j) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left[\frac{(z_j - x_i)}{h}\right], j = 1,...,M.$$

The function is computed for a specified set of values z_j, $j = 1,...,M$. Note that each value requires a sum over the full sample of n values. *The default value of M is 100*. The primary component of the computation is the kernel function, K[.], a weighting function that integrates to one. Eight alternatives are provided:

1. Epanechnikov: $K[z] = .75(1 - .2z^2) / \text{Sqr}(5)$ if $|z| \leq 5$, 0 else,
2. Normal: $K[z] = \phi(z)$ (normal density), $-\infty < z < \infty$
3. Logit: $K[z] = \Lambda(z)[1-\Lambda(z)]$ (default) , $-\infty < z < \infty$
4. Uniform: $K[z] = .5$ if $|z| \leq 1$, 0 1 else,
5. Beta: $Z[z] = (1-z)(1+z)/24$ if $|z| < 1$, 0 1 else,
6. Cosine: $K[z] = 1 + \cos(2\pi z)$ if $|z| < .5$, 0 else,
7. Triangle: $K[z] = 1 - |z|$, if $|z| \leq 1$, 0 else.
8. Parzen: $K[z] = 4/3 - 8z^2 + 8|z|^3$ if $|z| \leq .5$, $8(1-|z|)^3$ else.

The other essential part of the computation is the smoothing (bandwidth) parameter, h. Large values of h stabilize the function, but tend to flatten it and reduce the resolution (in the same manner as its discrete analog, the bin width in a histogram). Small values of h produce greater detail, but also cause the estimator to become less stable.

The basic command is

KERNEL ; Rhs = the variable $

With no other options specified, the routine uses the logit kernel function, and uses a data driven bandwidth equal to

$$h = .9Q/n^{0.2} \text{ where } Q = \min(\textit{std.dev.}, \textit{range}/1.5)$$

The command builder for this estimator may be found by selecting either Model:Data Description/Kernel Density or Model:Nonlinear Regression/Kernel. The dialog boxes are the same in both places.

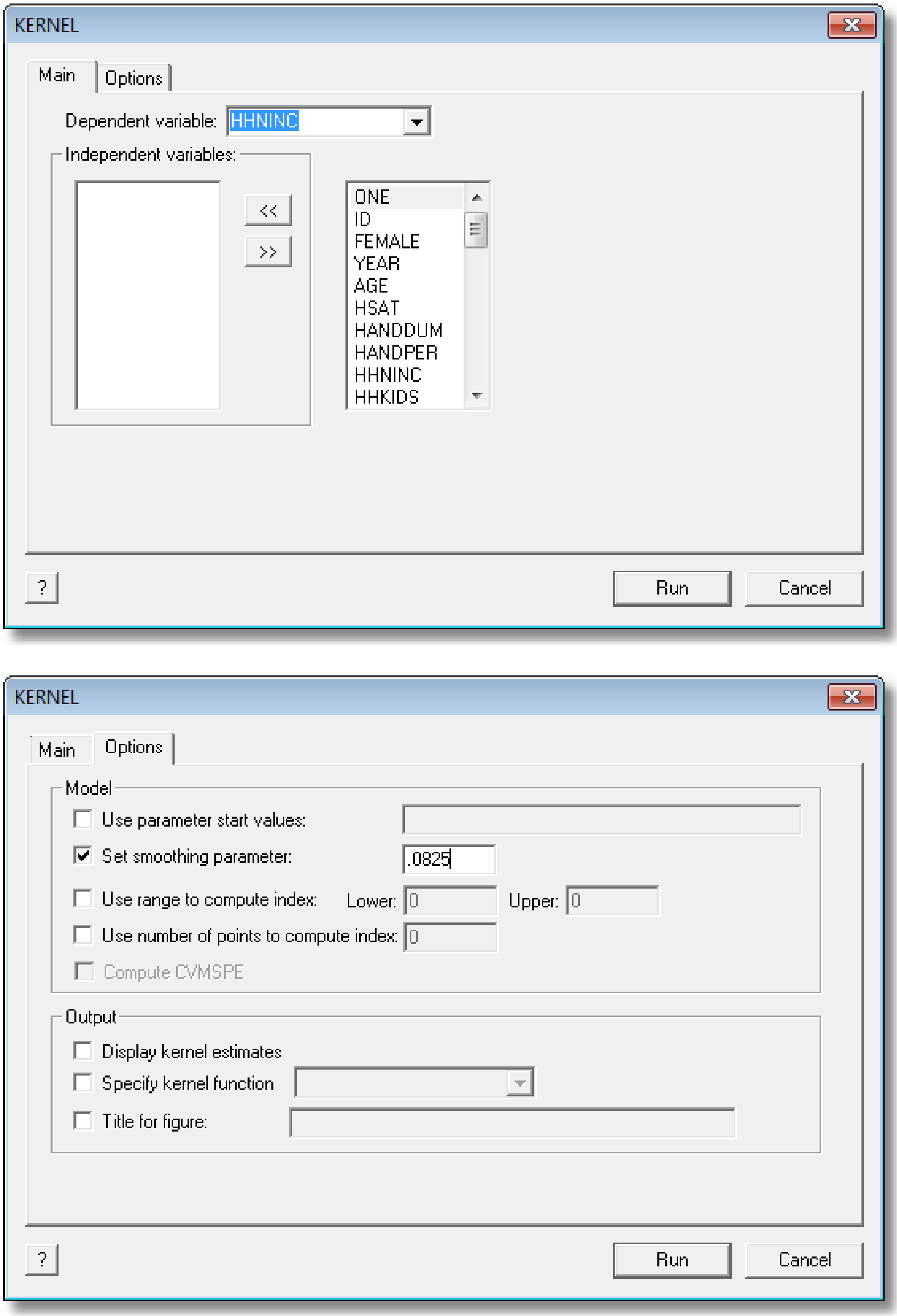
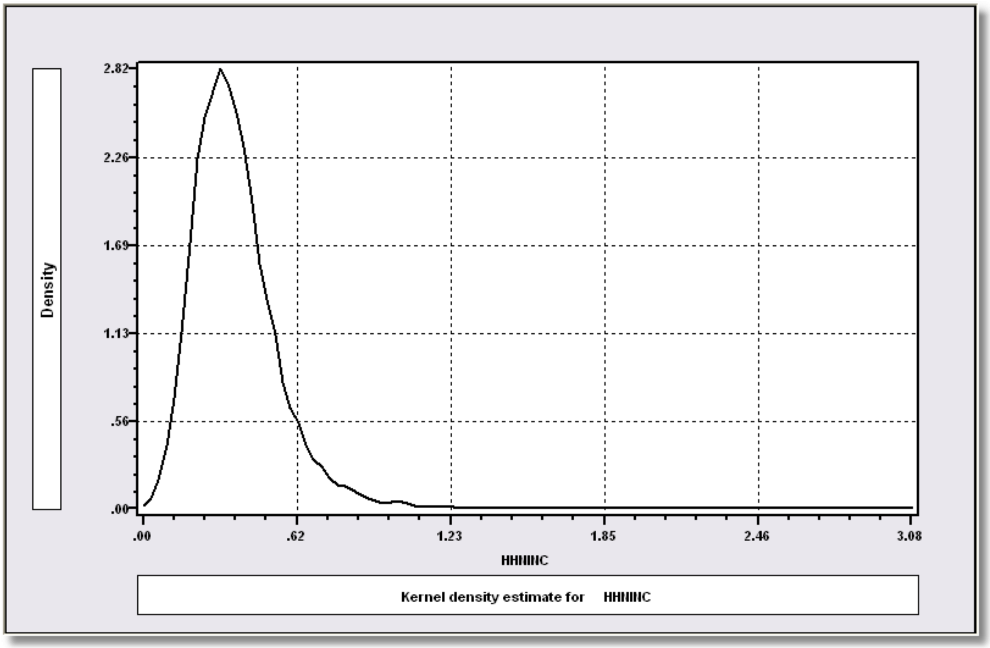


Figure E3.19 Command Builder for Kernel Density Estimator

For an example, we will compute the kernel density that is a smoothed counterpart to the histogram for income distribution in Figure E3.5. The command is

KERNEL ; Rhs = hhninc $

The histogram is repeated to show the similarity.



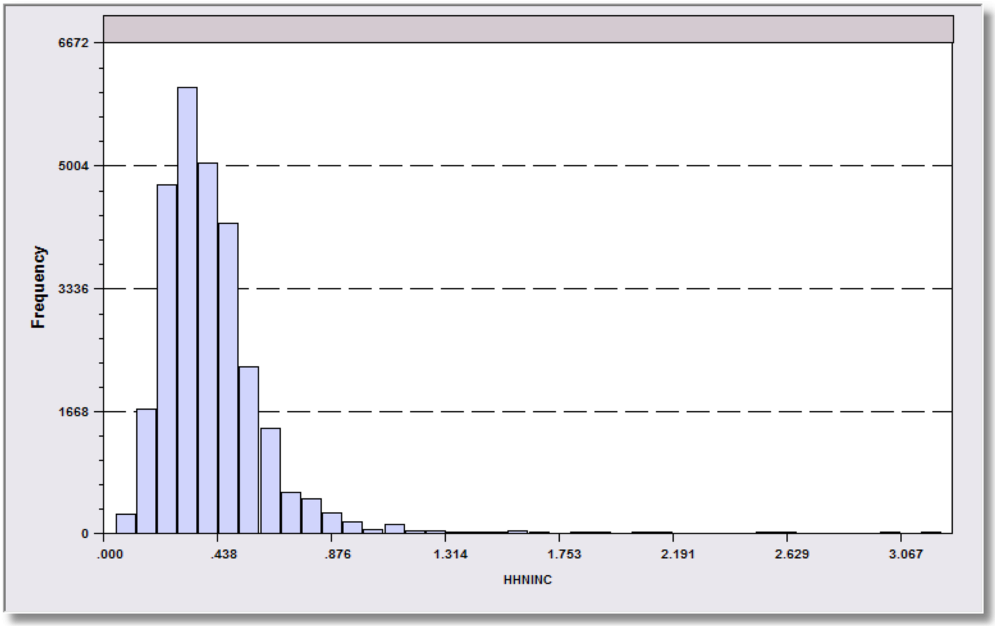


Figure E3.20 Kernel Density Estimator and Histogram for Incomes

The kernel density also produces some summary statistics, as shown below for the example in Figure E3.20.

```
+---------------------------------------+
| Kernel Density Estimator for HHNINC   |
| Observations       =          27326   |
| Points plotted     =            100   |
| Bandwidth          =        .020632   |
| Statistics for abscissa values----    |
| Mean               =        .352135   |
| Standard Deviation =        .176857   |
| Minimum            =        .001500   |
| Maximum            =       3.067100   |
| ----------------------------------    |
| Kernel Function    =       Logistic   |
| Cross val. M.S.E.  =        .000000   |
| Results matrix     =         KERNEL   |
+---------------------------------------+
```

The data used to plot the kernel estimator are also retained in a new matrix named *kernel*. Figure E3.21 shows the results for the preceding plot:

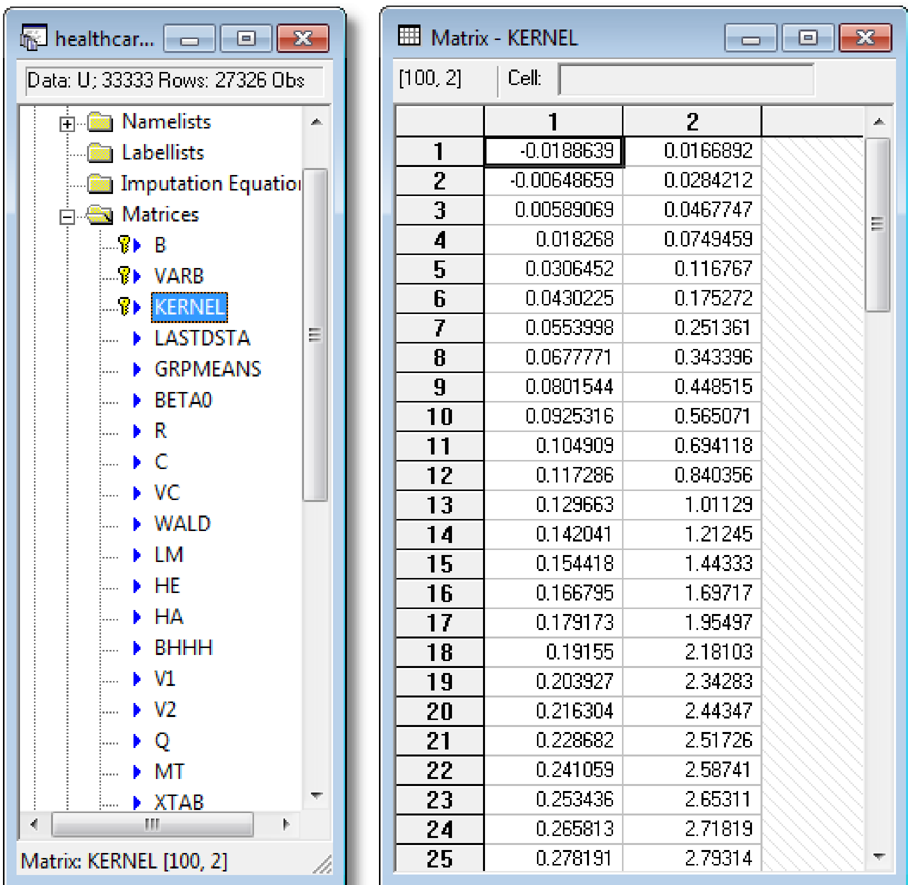


Figure E3.21 Matrix Results from KERNEL

E3.6.1 Options for Kernel Density Estimation

The weighting function for the kernel estimator is specified with

; Kernel = one of the eight types of kernels listed earlier: Epanechnikov, Normal, Logistic, Cosine, Uniform, Beta, Triangle, or Parzen.

The bandwidth may be specified with

; Smooth = the bandwidth parameter.

The default number of points specified is 100, with *zj* = a partition of the range of the variable. You may specify the number of points, up to 1000 with

; Pts = number of points to compute and plot

More than a few hundred points is not helpful, since the resolution of a modern display will not exceed a width of 2,000 points. The set of points z_j is then (for any number of points),

$$z_j = z_L + j*[(z_U - z_L)/M],\ j = 1,...,M\ z_L = \min(x)-h \text{ to } z_U = \max(x)+h.$$

Results of this procedure are a $M\times 2$ matrix named *kernel* in which the first column contains z_j and the second column contains the values of $f(z_j)$ and plot of the second column against the first – this is the estimated density function.

You may fix the limits on the vertical axis of the figure with

; Limits = high, low

This overrides the default limits computed internally. *Note reversal of the usual order.* The alternative specification,

; Limits = low, high

is used to restrict the sample values of the variable used to compute the kernel density estimator to those in the range from low to high. You can also fix the limits on the horizontal axis with

; Endpoints = low, high

Do note that this may conflict with the parameters being used to define the kernel estimator, however. If the range of your data being analyzed is 0-10, for example, and you specify **; Endpoints = 0,5**, the figure may be distorted. The data are not adjusted to conform to the endpoints.

Observation weights may be applied to the kernel estimator. Weights of any sort, including complex survey weights, may be applied, so that the revised estimator is

$$f(z_j) = \frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{h}K\left[\frac{\left(z_j - x_i\right)}{h}\right], \text{ such that } \Sigma_{i=1}^{n} w_i = 1,\ j = 1,...,M.$$

Note that the adjustment of the sum of the weights may be necessary if you specify that only a subsample of the current sample is to be used. As such, when you specify

KERNEL ; Rhs = the variable ; Wts = the weighting variable $

weights are automatically scaled, whether or not you have used use **,Noscale**. As such, you should not use **,Noscale** with this computation. (See Section R8.8.)

Finally, you may add a title to the figure with

; Title = up to 60 characters.

A somewhat neater version of Figure E3.20 which corresponds to Figure E3.8, is produced by

KERNEL ; If [hhninc <= 1.6]
; Rhs = hhninc
; Title = Income Distribution Truncated at 1.6
; Endpoints = 0,1.6 $

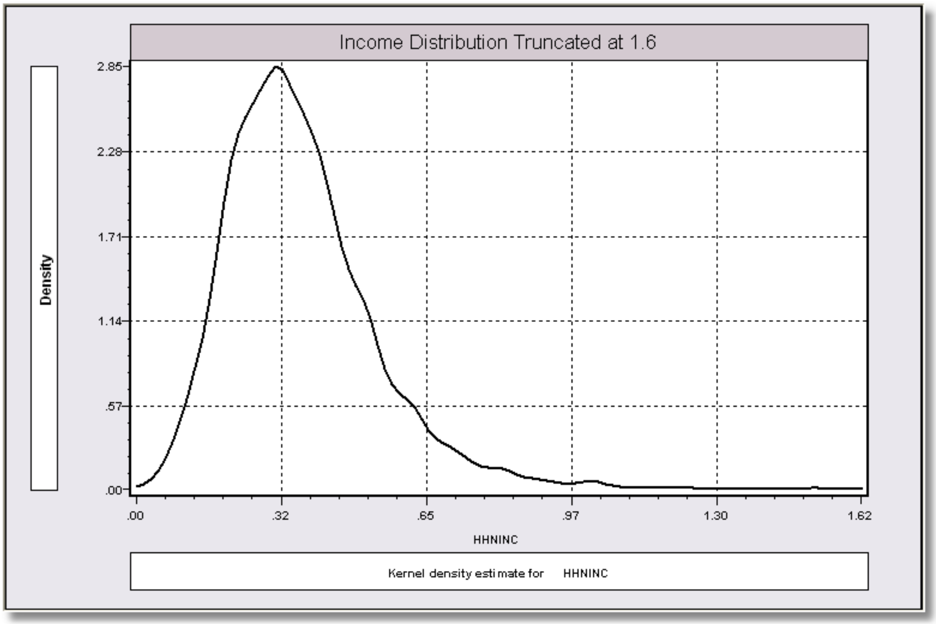


Figure E3.22 Kernel Density Estimator for Truncated Distribution

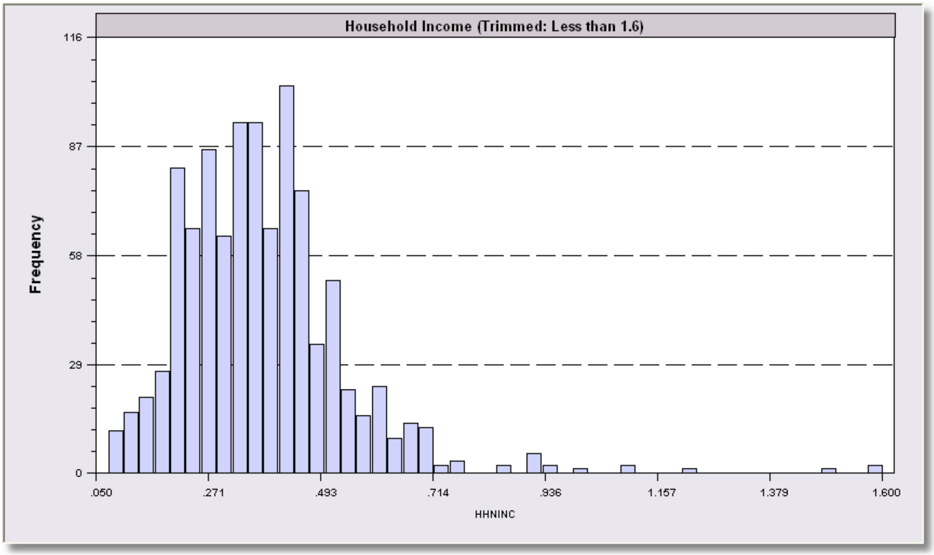


Figure E3.23 Kernel Density Estimator and Histogram for Incomes, with Trimming

E3.6.2 Multiple Kernel Estimators

Multiple kernel estimates can be placed in the same figure by including up to four variables in the Rhs list in the **KERNEL** command. In the example below, estimators of technical efficiency produced by two different stochastic frontier models are compared. (See Chapter E62.)

```
NAMELIST    ; x = one,x1,x2,x3,x4 $
FRONTIER    ; Lhs = yit ; Rhs = x ; Techeff = e_hlfnrm $
FRONTIER    ; Lhs = yit ; Rhs = x ; Techeff = e_expon
            ; Model = Exponential $
KERNEL      ; Rhs = e_hlfnrm,e_expon
            ; Title = Estimates of Technical Efficiency $
```

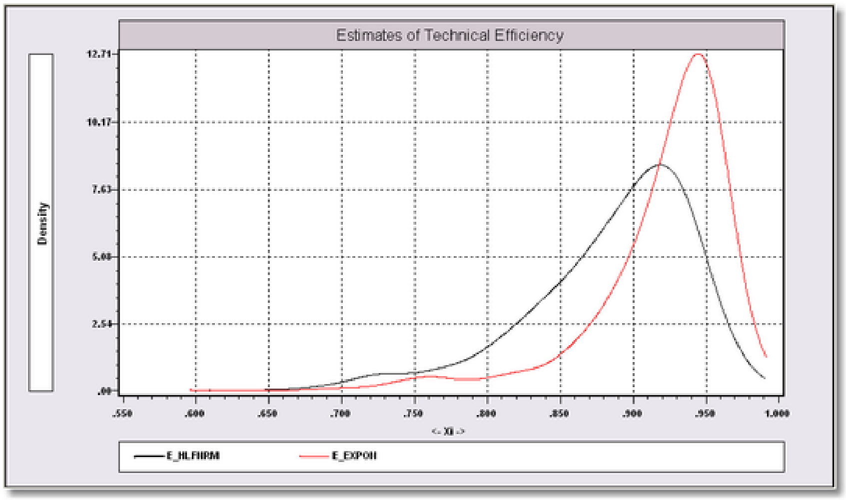


Figure E3.24 Multiple Kernel Estimators

E3.6.3 Sample Strata

To compare subgroups within a sample, use

```
KERNEL     ; Rhs = variable
           ; Group = variable $
```

The group variable partitions the sample in up to five subsamples. In the figure below, the variable female is coded 0 for males and 1 for females, so it partitions the sample into two groups. Labels for the groups may be provided as well with

```
           ; Labels = list of labels
```

The example below compares the responses of men and women to the health status question in the GSOEP data.

```
SAMPLE     ; 1-5000 $
KERNEL     ; Rhs = hsat
           ; Group = female
           ; Labels = men,women
           ; Title = Health Satisfaction: Male vs. Female $
```

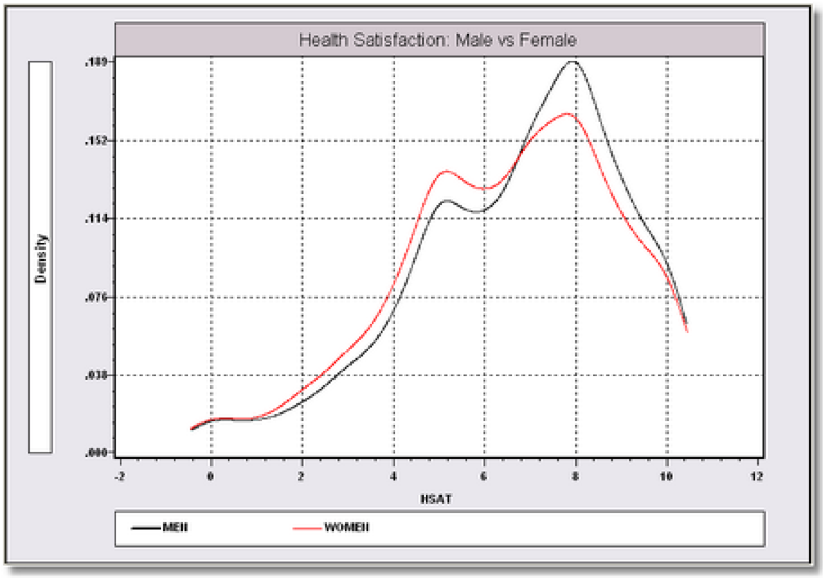


Figure E3.25 Kernel Estimators for Subsamples

E3.6.4 Comparison to Normal

The kernel estimator can be used to examine departures from or similarity to a normal distribution. To superimpose a normal distribution with the same mean and variance as the underlying variable in the kernel estimator, add

; Normal

to the **KERNEL** command. This is a common exercise in the examination of least squares prior to stochastic frontier modeling. The example below displays a kernel estimator and a normal density for the least squares residuals computed with the stochastic frontier models in Section E3.6.2.

REGRESS ; Lhs = yit ; Rhs = x ; Res = u $
KERNEL ; Rhs = u ; Normal
; Title = OLS Residuals with Evidence of Inefficiency $

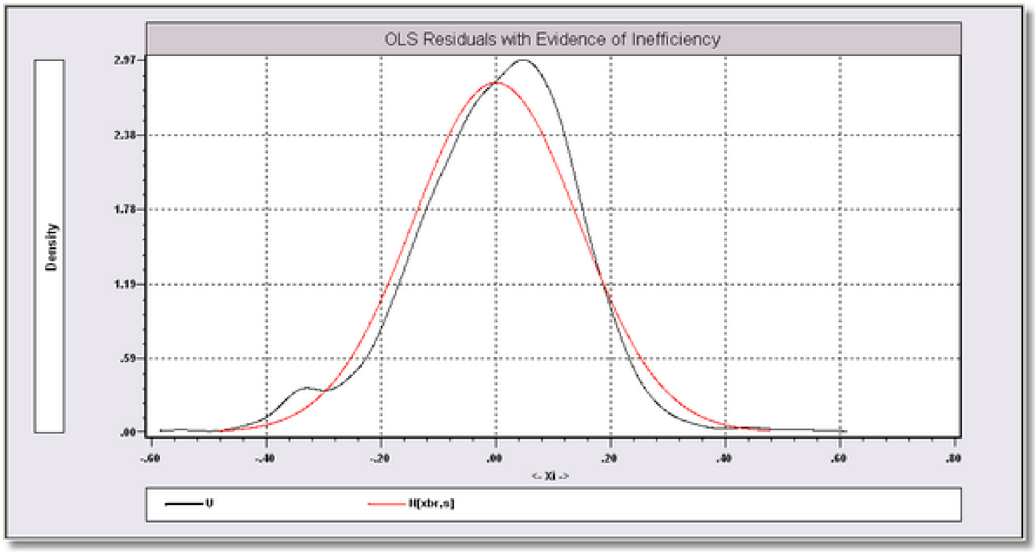


Figure E3.26 Kernel Density Estimator with Normal Density

E3.7 Testing for Normality

Two tests are generally used to test the resemblance of the distribution of a sample to an underlying normal distribution, the Bowman and Shenton (1975) test based on the third and fourth sample moments and the Kolmogorov-Smirnov test based on the sample CDF.

E3.7.1 Normality Test Based on Skewness and Kurtosis

The Bowman and Shenton chi squared statistic for testing against the null hypothesis of normality is

$$\chi^2[2] = N[(m_3/s^3)^2 /6+ (m_4/s^4 - 3)^2/24].$$

where m_3 is the average cubed deviation from the mean (the sample skewness) and m_4 is the average fourth power. In the two cases, the moment is divided by the third or fourth power of the sample standard deviation, respectively. There are several ways to compute this result:

For variables being analyzed with **DSTAT**, you can obtain this result by adding

; Normality test

to the **DSTAT** command. This will change the displayed output to include the statistic and the 'p value' which is the probability that a chi squared variable with two degrees of freedom would exceed this value. The 95% 'critical value' for chi squared[2] is 5.99, so based on this test, you would reject 'normality' at a 95% significance level if your statistic exceeds 5.99. For the least squares residuals used in the example in the previous section, we obtain

DSTAT ; Rhs = u ; Normality test $

```
--------+--------------------------------------------------------------------
Variable|    Mean       Std.Dev.     Minimum      Maximum        Cases Missing
--------+--------------------------------------------------------------------
       U| -.544414E-14   .14016     -.554019      .583301         1482       0
        | Skewness   -.30  Kurtosis     3.80    Chisq=   61.42    Prob = .0000
--------+--------------------------------------------------------------------
```

There are dedicated **CALC** functions for this computation:

CALC ; Sku(variable) $	Computes m_3.
CALC ; Krt(variable) $	Computes m_4.
CALC ; Sdv(variable) $	Computes s.
CALC ; Rb1(variable)$	Computes m_3/s^3.
CALC ; Bt2(variable) $	Computes $m_4/s^4 - 3$.

We can replicate the test result with

CALC ; List ; 1482*Rb1(u)^2/6 ; 1482*Bt2(u)^2/24 $

```
[CALC] *Result*=     21.6501337
[CALC] *Result*=     39.7657231
```

E3.7.2 Kolmogorov Smirnov Test of Normality

The Kolmogorov-Smirnov test is a nonparametric statistic used to test a distributional assumption. For the implementation here, we use the normal distribution as the null hypothesis. The statistic is computed as

$$D = \max_{1 \le i \le N}\left(F(x_i) - \frac{i-1}{N}, \frac{i}{N} - F(x_i) \right)$$

where F is the theoretical CDF being tested (normal). For the specified test,

CALC ; Kst(variable) $

reports the Kolmogorov-Smirnov test statistic. The null distribution is assumed to be the normal distribution. The mean and standard deviation of the normal distribution are estimated from the data. The derivation of the behavior of the test statistic, and the critical values, actually assume that the mean and variance of the distribution are known, not estimated from the data. So, the critical values given below should be viewed as approximate If you do know the mean and standard deviation of the distribution, provide them as the second and third parameters in the function, as in

CALC ; List ; Kst(variable, μ, σ) = Kolmogorov-Smirnov test against N[μ,σ^2].

Critical values of the distribution of the test statistic are as follows:

Sample Size	20	25	30	35	Over 35
95%	.294	.270	.240	.230	1.36/Sqr(*N*)
99%	.356	.320	.290	.270	1.63/Sqr(*N*)

For the least squares residuals used in the preceding example, we obtained

CALC ; List ; Kst(u) $

```
--------------------------------------
Kolmogorov-Smirnov test of F(U       )
vs. Normal[     .00000,      .14016^2]
******* K-S test statistic =  .0743086
******* 95% critical value =  .0353277
******* 99% critical value =  .0423412
Normality hyp. should     be rejected.
--------------------------------------
[CALC] *Result*=       .0743086
```

E4: Covariance and Correlation

E4.1 Introduction

This describes how to obtain covariances and various types of correlation coefficients. Scalar calculations of a single correlation for a pair of variables are given in Section E4.2. Covariance and correlation matrices are shown in Section E4.3. Correlations for discrete variables are described in Section E4.4. Section E4.5 shows how to compute and display cross tabulations

E4.2 Covariance and Correlation for Two Variables

CALC can be used to obtain a single covariance or correlation coefficient:

CALC ; [name =] Cov(variable 1, variable 2) $
or **CALC ; [name =] Cor(variable 1, variable 2) $**

E4.2.1 Kendall's Tau

Kendall's tau is a nonparametric measure of the concordance of the ranks of two variables. It is computed as

$$\tau = \frac{\sum_{i=1}^{N}\sum_{j=1}^{i-1} \operatorname{sgn}(x_i - x_j)\operatorname{sgn}(y_i - y_j)}{N(N-1)/2}, \ \operatorname{sgn}(z) = (\text{-}1,0,1) \text{ if } z\ (>0, =0, >0).$$

The exact distribution is unknown and would depend on the underlying population. Under the null hypothesis that τ equals zero, the distribution of τ is approximately normal with mean zero and variance $[(2N+5)/9]/[N(N+1)/2]$. **CALC** may be used with

CALC ; [name =] Ktr(variable 1, variable 2) $

E4.2.2 Rank Correlation

Spearman's correlation for a pair of sets of ranks is computed as

$$r_{RANKS} = \frac{6\sum_{i=1}^{N}(rank_{i,1} - rank_{i,2})^2}{N(N^2-1)} = \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}$$

Compute this with for two sets of ranks using

CALC ; [name =] Rkc(variable 1, variable 2) $

E4.3 Covariance and Correlation Matrices

Covariance and correlation matrices may be obtained as part of the output with **DSTAT** or separately using **MATRIX**.

E4.3.1 Matrix Output from DSTAT

After the table of results is given, you may elect to display a covariance or correlation matrix (or both) for the variables. The request is added to the command:

; Output = 1 to obtain the covariance matrix,
; Output = 2 to obtain the correlation matrix,
; Output = 3 for both covariance and correlation matrices.

The matrix displays,

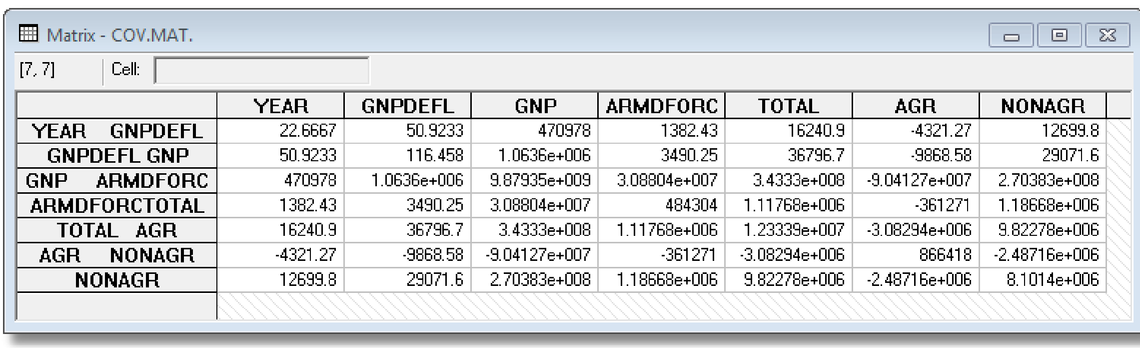
; Output = 1, **2**, or **3**,

will display either the full matrix in the output if it is small enough, or an embedded object if the matrix is too large (larger than 5×5 for a covariance matrix, 7×7 for a correlation matrix). The results for the data above are shown below. The table of statistics is the same, and is omitted. (Display of covariance matrices is difficult as the scale of the entries is the square of the data. Correlations can be displayed in much less cluttered format.) In this example, the covariance matrix appears as an embedded object.

DSTAT ; Rhs = year,gnpdefl,gnp,armdforc,total,agr,nonagr
; Output = 3 $

… other results from **DSTAT**

Matrix: COV.MAT.
[7,7]



Matrix - COV.MAT.
[7, 7] Cell:

	YEAR	GNPDEFL	GNP	ARMDFORC	TOTAL	AGR	NONAGR
YEAR GNPDEFL	22.6667	50.9233	470978	1382.43	16240.9	-4321.27	12699.8
GNPDEFL GNP	50.9233	116.458	1.0636e+006	3490.25	36796.7	-9868.58	29071.6
GNP ARMDFORC	470978	1.0636e+006	9.87935e+009	3.08804e+007	3.4333e+008	-9.04127e+007	2.70383e+008
ARMDFORCTOTAL	1382.43	3490.25	3.08804e+007	484304	1.11768e+006	-361271	1.18668e+006
TOTAL AGR	16240.9	36796.7	3.4333e+008	1.11768e+006	1.23339e+007	-3.08294e+006	9.82278e+006
AGR NONAGR	-4321.27	-9868.58	-9.04127e+007	-361271	-3.08294e+006	866418	-2.48716e+006
NONAGR	12699.8	29071.6	2.70383e+008	1.18668e+006	9.82278e+006	-2.48716e+006	8.1014e+006

Figure E4.1 Covariance Matrix as Embedded Object from DSTAT

```
--------+--------------------------------------------------------------
Cor.Mat.|     YEAR  GNPDEFL       GNP ARMDFORC     TOTAL       AGR   NONAGR
--------+--------------------------------------------------------------
    YEAR| 1.00000   .99115   .99527   .41725   .97133  -.97511   .93718
 GNPDEFL|  .99115  1.00000   .99159   .46474   .97090  -.98244   .94646
     GNP|  .99527   .99159  1.00000   .44644   .98355  -.97724   .95573
ARMDFORC|  .41725   .46474   .44644  1.00000   .45731  -.55771   .59909
   TOTAL|  .97133   .97090   .98355   .45731  1.00000  -.94308   .98266
     AGR| -.97511  -.98244  -.97724  -.55771  -.94308  1.00000  -.93877
  NONAGR|  .93718   .94646   .95573   .59909   .98266  -.93877  1.00000
```

The embedded object (Matrix:COV.MAT) shown above is the covariance matrix created by this command. This matrix, as displayed in the window is read only. However, the matrix can be exported to a spreadsheet program such as *Excel* just by clicking the empty square in the upper left corner, then using edit/copy in *LIMDEP* and edit/paste into *Excel*. The labeling and arrangement of the contents of the matrix will be preserved in a tab delimited format.

E4.3.2 Correlation and Covariance Matrices with MATRIX

Correlation matrices are produced by the matrix functions

MATRIX **[; List] ; [name =] Xcor (namelist for X) $**

which produces a correlation matrix for the variables in the namelist. (Optional specifications are shown in square brackets.) Cross correlations may be produced by using

MATRIX **[; List] ; [name =] Xcor(namelist for X, namelist for Y) $**

The correlation matrix shown in the first form results if X and Y are the same namelist. The observations may be weighted by using

MATRIX **[; List] ; [name =] Xcor(namelist for X, namelist for Y, weights) $**

where weights is a variable that contains the weights. The form **Xcor(X,X,w)** must be used to obtain a weighted correlation matrix. Covariance matrices are obtained by changing **Xcor** to **Xcov** in the preceding.

A matrix of Kendall's tau(x,y) for a set of variables is obtained with

MATRIX **[; List] ; [name =] Xtau (namelist for X) $**

Weights are not supported for the matrix of $\tau(x,y)$ correlation coefficients.

E4.4 Correlations for Discrete Variables

Correlations involving discrete variables are generally not computed using standard Pearson product moment correlation coefficients. Two for strictly discrete variables based on censoring an underlying normal variable are the tetrachoric correlation for two binary variables and the polychoric correlation for two ordered categorical variables or a binary variable and a categorical variable. The biserial correlation described in Section E4.4.3 is used for a binary variable and a continuous variable.

E4.4.1 Tetrachoric Correlation for Binary Variables

The tetrachoric correlation between binary variables d_1 and d_2 is described as the correlation that would be observed if the two variables were normally distributed around fixed means with variance one. Thus,

$$d_{\mathrm{j}} = 1(d_j^* > 0) \mid d_j^* \sim \mathrm{N}[0,1],\ \mathrm{j} = 1,2.$$

If we write this in full, we have the following bivariate model:

$$d_1^* = \alpha_1 + \varepsilon_1,\ d_1 = 1[d_1^* > 0]$$
$$d_2^* = \alpha_2 + \varepsilon_2,\ d_2 = 1[d_2^* > 0]$$
$$(\varepsilon_1,\varepsilon_2) \sim N_2[(0,0),(1,\rho,1)]$$

The correlation coefficient, ρ, is estimated by maximum likelihood. The structure can be seen to define a bivariate probit model in which the two regressor vectors are simply a constant term. The computation can be requested with

TCORRELATION ; Lhs = d1 ; Rh1 = d2 $

There are no other options for this model command. Note that the command is identical to

BIVARIATE PROBIT ; Lhs = d1,d2 ; Rh1 = one ; Rh2 = one $

and you can apply other optional features to this command, such as **; List** if you wish.

On computation, Olsson (1979) is among numerous sources that discuss maximum likelihood estimation of the tetrachoric correlation. The quite simple approach of treating this as the most simple type of bivariate probit model in this fashion seems to have gone unnoticed in the received literature.

In the example below, we have computed the tetrachoric correlation for the two behavioral variables in the health care data, *doctor* = 1[*docvis* > 0] and *hospital* = 1[*hospvis* > 0]. We have used the 1991 wave of the data set. The analysis begins with a descriptive cross tabulation. Crosstabs are described in Section E4.5.

CROSSTAB ; Lhs = doctor ; Rhs = hospital $

```
+-------------------------------------------------------------------+
|Cross Tabulation                                                   |
|Row variable is DOCTOR   (Out of range 0-49:       0)              |
|Number of Rows =  2      (DOCTOR   =  0 to  1)                     |
|Col variable is HOSPITAL (Out of range 0-49:       0)              |
|Number of Cols =  2      (HOSPITAL =  0 to  1)                     |
|Chi-squared independence tests:                                    |
|Chi-squared[   1] =   49.13158   Prob value =  .00000              |
|G-squared  [   1] =   54.56513   Prob value =  .00000              |
+-------------------------------------------------------------------+
|               HOSPITAL                                            |
+--------+-------------+------+                                     |
|  DOCTOR|      0      1| Total|                                    |
+--------+-------------+------+                                     |
|       0|   1340     68|  1408|                                    |
|       1|   2597    335|  2932|                                    |
+--------+-------------+------+                                     |
|   Total|   3937    403|  4340|                                    |
+-------------------------------------------------------------------+
```

TCORRELATION ; Lhs = doctor ; Rhs = hospital $

```
Normal exit:   6 iterations. Status=0, F=    4049.098
```

```
-----------------------------------------------------------------------------
FIML Estimation of Tetrachoric Correlation
Dependent variable                DOCTOR,HOSPITAL
Log likelihood function      -4049.09756
Estimation based on N =    4340, K =   3
--------+--------------------------------------------------------------------
  DOCTOR|                   Standard            Prob.      95% Confidence
HOSPITAL|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Estimated alpha for P[DOCTOR  =1] = F(alpha)
Constant|     .45536***      .01976    23.05  .0000      .41664    .49409
        |Estimated alpha for P[HOSPITAL=1] = F(alpha)
Constant|   -1.32336***      .02651   -49.92  .0000    -1.37532  -1.27141
        |Tetrachoric Correlation between DOCTOR   and HOSPITAL
RHO(1,2)|     .26548***      .03473     7.65  .0000      .19742    .33354
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E4.4.2 Polychoric Correlation for Two Ordered Qualitative Variables

The polychoric correlation is computed similarly to the tetrachoric correlation. The statistic has been recommended to measure the concordance of different judges or raters. Suppose, for example, there are a pair of judges each asked to rate restaurants on a scale from 0 to 4. The statistic is intended to measure the correlation of their ratings. The underlying model is precisely that underlying the bivariate ordered probit model shown in Section E35.9. For each individual (or utility function), we have

$$
\begin{aligned}
y_1^* &= \alpha_1 + \varepsilon_1, \quad y_1 = 0 \text{ if } y_1^* \le 0,\\
&\qquad\qquad y_1 = 1 \text{ if } 0 < y_1^* < \mu_1,\\
&\qquad\qquad y_1 = 2 \text{ if } \mu_1 < y_1^* < \mu_2,\\
&\qquad\qquad \ldots\\
&\qquad\qquad y_1 = J \text{ if } \mu_{J-1} < y_1^* < +\infty.\\
y_2^* &= \alpha_2 + \varepsilon_2, y_2 = 0 \text{ if } y_2^* \le 0,\\
&\qquad\qquad y_2 = 1 \text{ if } 0 < y_2^* < \lambda_1,\\
&\qquad\qquad y_2 = 2 \text{ if } \lambda_1 < y_2^* < \lambda_2,\\
&\qquad\qquad \ldots\\
&\qquad\qquad y_2 = M \text{ if } \mu_{M-1} < y_2^* < +\infty.\\
(\varepsilon_1,\varepsilon_2) &\sim N_2[(0,0),(1,1,\rho)].
\end{aligned}
$$

In this framework, then, ρ is the polychoric correlation. Either variable may be binary. If both are, then the tetrachoric correlation of the preceding section applies. The maximum likelihood estimate of the coefficient is obtained by treating the preceding as a bivariate ordered probit model in which both equations have only a constant term. The calculation is requested with

PCORRELATION ; Lhs = y1 ; Rhs = y2 $

In the example below, we have obtained the polychoric correlation between *docvis* (truncated at 5) and *hospvis* (truncated at 2). Both variables are counts. In this application, we are treating the counts as an indicator of underlying health in a particular dimension. We begin with a descriptive cross tabulation.

SAMPLE ; All $
REJECT ; year # 1991 | docvis > 5 | hospvis > 2 $
CROSSTAB ; Lhs = docvis ; Rhs = hospvis $

```
+-------------------------------------------------------------------+
|Cross Tabulation                                                   |
|Row variable is DOCVIS   (Out of range 0-49:      0)              |
|Number of Rows =  6      (DOCVIS   =  0 to  5)                    |
|Col variable is HOSPVIS  (Out of range 0-49:      0)              |
|Number of Cols =  3      (HOSPVIS  =  0 to  2)                    |
|Chi-squared independence tests:                                    |
|Chi-squared[  10] =   43.56260   Prob value =  .00000             |
|G-squared  [  10] =   39.00465   Prob value =  .00003             |
+-------------------------------------------------------------------+
|                HOSPVIS                                            |
+--------+--------------------+------+                              |
|  DOCVIS|      0      1      2| Total|                              |
+--------+--------------------+------+                              |
|       0|   1340     57      7|  1404|                              |
|       1|    721     49      8|   778|                              |
|       2|    596     42      4|   642|                              |
|       3|    481     44      3|   528|                              |
|       4|    189     19      5|   213|                              |
|       5|    150     23      3|   176|                              |
+--------+--------------------+------+                              |
|   Total|   3477    234     30|  3741|                              |
+-------------------------------------------------------------------+
```

PCORRELATION ; Lhs = docvis ; Rhs = hospvis $

```
Normal exit:  10 iterations. Status=0, F=    6942.729
-----------------------------------------------------------------------------
Polychoric Correlation for Ordered Variable
DOCVIS   = 0, 1, ...,  5
HOSPVIS  = 0, 1, ...,  2
--------+--------------------------------------------------------------------
  DOCVIS|                 Standard            Prob.       95% Confidence
 HOSPVIS|  Coefficient      Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Mean inverse probability for DOCVIS
Constant|     .31763***     .02087    15.22  .0000      .27672    .35855
        |Mean inverse probability for HOSPVIS
Constant|   -1.47189***     .03101   -47.46  .0000    -1.53267  -1.41111
        |Polychoric Correlation for DOCVIS   and HOSPVIS
RHO(1,2)|     .18523***     .03151     5.88  .0000      .12347    .24699
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E4.4.3 Biserial Correlation

The biserial correlation is sometimes used to assess the correlation between a continuous variable, *x*, and a binary variable, *d*. (See Glass and Hopkins (1995).) The computation is

$$r(x,d) = \frac{\bar{x}_{|d=1} - \bar{x}_{|d=0}}{s_x}\sqrt{\frac{n_1 n_0}{n^2}}$$

A standard error for the correlation is approximated by

$$s_{r(x,d)} = \sqrt{\frac{1-r^2(x,d)}{n_1+n_2}},$$

assuming that the sample is large enough that a normal approximation to the t distribution is satisfactory (i.e., greater than about 100). With a small sample, normally distributed, *x* and exogenous *d*, the t distribution is used and the degrees of freedom in the standard error are *n*-2. The biserial correlation, an estimate of the standard error and an estimated confidence interval are obtained with

CALC [; List] ; [name =] Bsr(x,d) $

In the example below, the coefficient is computed for income and the binary indicator for whether the individual has purchased public insurance.

CALC ; List ; Bsr(income,public) $

```
---------------------------------------------
Biserial correlation of INCOME   and PUBLIC
Estimated correlation coefficient = -.19415
Estimated standard error for bsr  =  .00593
Estimated 95%conf. interval=(-.2058,-.1825)
---------------------------------------------
[CALC] *Result*=      -.1941491
```

E4.5 Cross Tabulations

The command for crosstabs based on two variables is

CROSSTAB ; Lhs = rows variable ; Rhs = columns variable $

Use **CROSSTAB** to analyze a pair of discrete variables that are coded 0,1,... up to 49 (i.e., up to 2,500 possible outcomes). The table may be anywhere from 2×2 to 50×50. (Row and column sizes need not be the same.) Observations which do not take these values are tabulated as ‘out of range.’

This command assumes that your data are coded as integers, 0,1,... If you wish to analyze continuous variables, you must use the **RECODE** command (Section R4.7) to recode the continuous ranges to these values.

The categories are automatically labeled ‘NAME=0,’ ‘NAME=1,’..., etc. for the two variables. To provide your own labels and to specify the number of categories for the variables, add

; Labels = list of labels for Lhs / list of labels for Rhs

to the command. Labels may contain up to eight characters. Separate labels in the lists with commas. Cross tabulations may be computed with unequally weighted observations. The specification is

; Wts = variable

as usual. This scales the weights to sum to the sample size. If the weights are replications that should not be scaled, use

; Wts = variable,Noscale

The command builder dialog boxes that you can use to construct the command for CROSSTAB are found by selecting Model:Data Description/Crosstab. The Main and Options pages are shown below.

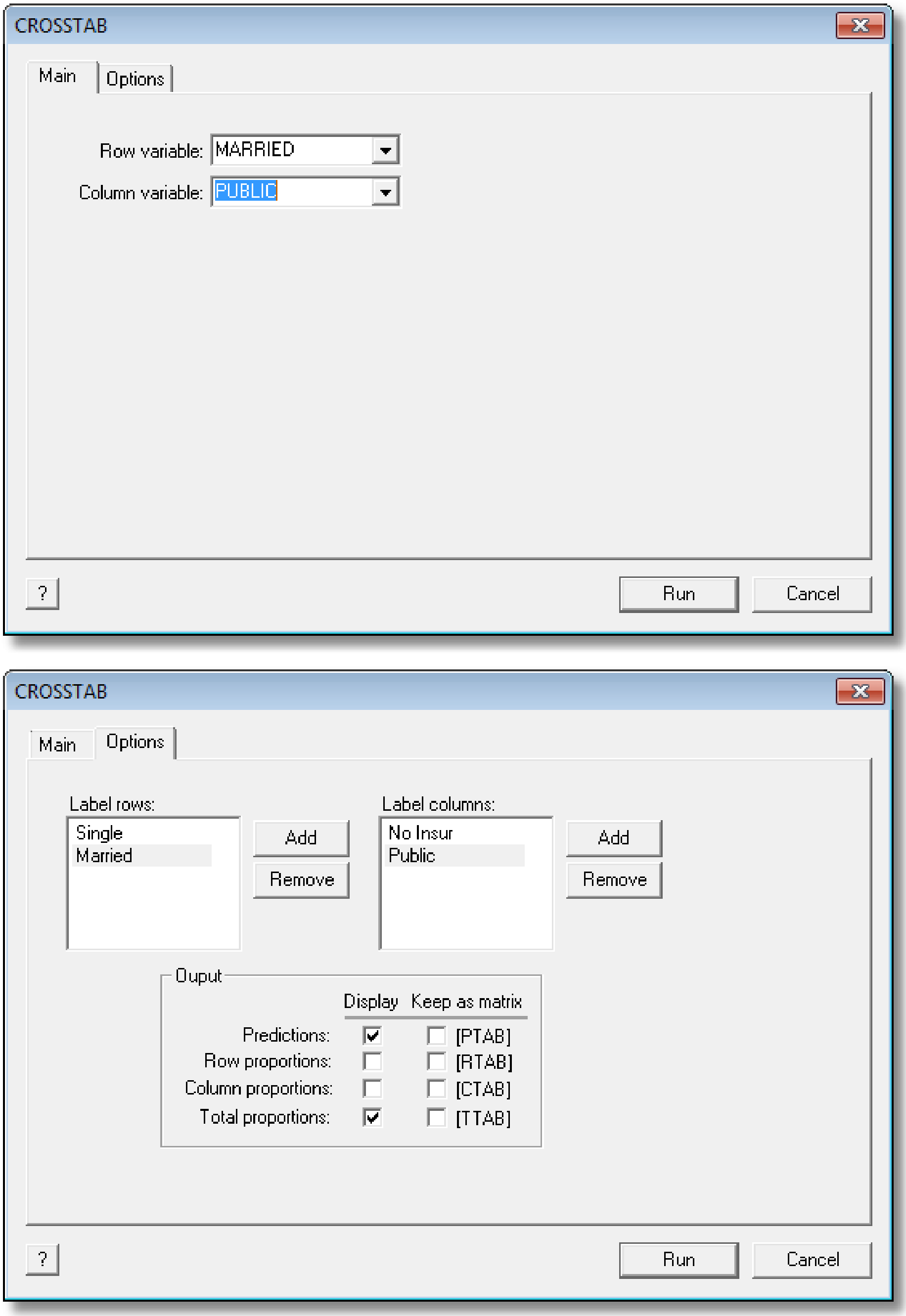


Figure E4.2 Command Builder for CROSSTAB

E4.5.1 Output

The **CROSSTAB** command will produce five types of tables. The default is a simple cross frequency table. Other options are

p = predictions, using the independence model (see below),
r = table entries are row proportions,
c = table entries are column proportions,
t = table entries are total sample proportions.

Use

; Output = any or all of p,r,c,t to request the display.
; Store = any or all of p,r,c,t to request keeping results as matrices.

Matrix *xtab*, the basic frequency table, is kept automatically. Other matrices that will be kept are

ptab if **; Store = ...,p...**
rtab if **; Store = ...,r...**
ctab if **; Store = ...,c...**
ttab if **; Store = ...,t...**

```
+------------------------------------------------------------------+
|Cross Tabulation                                                  |
|Row variable is MARRIED  (Out of range 0-49:      0)              |
|Number of Rows =  2      (MARRIED  =  0 to  1)                    |
|Col variable is PUBLIC   (Out of range 0-49:      0)              |
|Number of Cols =  2      (PUBLIC   =  0 to  1)                    |
|Chi-squared independence tests:                                   |
|Chi-squared[   1] =     .00108   Prob value =  .97379             |
|G-squared  [   1] =     .00108   Prob value =  .97379             |
+------------------------------------------------------------------+
|                 PUBLIC                                           |
+--------+-------------+------+                                    |
| MARRIED| NOINSU PUBLIC| Total|                                   |
+--------+-------------+------+                                    |
|SINGLE  |    141    968|  1109|                                   |
|MARRIED |    430   2942|  3372|                                   |
+--------+-------------+------+                                    |
|   Total|    571   3910|  4481|                                   |
+------------------------------------------------------------------+

--------+----------------------------------------
Predctns|       NOINSUR        PUBLIC         Total
--------+----------------------------------------
  SINGLE|       141.316       967.684       1109.00
 MARRIED|       429.684       2942.32       3372.00
  SINGLE|       141.316       967.684       1109.00
   Total|       571.000       3910.00       4481.00

--------+----------------------------------------
TtlPrcnt|       NOINSUR        PUBLIC         Total
--------+----------------------------------------
  SINGLE|      .0314662       .216023       .247489
 MARRIED|      .0959607       .656550       .752511
  SINGLE|      .0314662       .216023       .247489
   Total|       .127427       .872573       1.00000
```

E4.5.2 Testing the Independence Assumption

Frequency counts are often used for analyzing the hypothesis of independence of two variables. Suppose that the frequency table contains cell counts, n_{ij}, row sums, $n_{i.}$, and column sums, $n_{.j}$, and that there are total of n observations. The predicted cell frequencies are

$$Fij = n\times(n_{i.}/n)\times(n_{.j}/n) = n_{i.}\times n_{.j}/n.$$

Two statistics are computed for testing the independence hypothesis:

$$\text{Chi squared} = \Sigma_i\Sigma_j\,(F_{ij} - n_{ij})^2 / n_{ij}$$
$$G \text{ squared} = \Sigma_i\Sigma_j\, n_{ij}\log(F_{ij} / n_{ij}).$$

Both statistics are reported as zero if there are any cells with zero frequencies, since neither can be computed. Both of these are distributed in large samples as chi squared with degrees of freedom

$$K = (\text{number of rows - 1}) \times (\text{columns - 1}).$$

In addition to the matrices given earlier, the following scalars are kept by this procedure:

gsqrd	will contain G squared.
csqrd	will contain chi squared.
degfrdm	will contain the degrees of freedom, K.

LIMDEP also reports the probability that the chi squared variable would be at least this large (the p value) for the two statistics.

E4.5.3 Analyzing Frequency Data

If your data are already tabulated in the form of a frequency table, you can compute the independence tests, and predicted frequencies as follows:

```
MATRIX      ; nij  =  the table of frequencies $
CROSSTAB    ; Lhs  =  nij $
```

The command is different in that the **; Lhs** specifies a matrix, not a variable, and there is no **; Rhs**. This produces the same results as if the data were individual. No note is made in the results that the data were already tabulated. For example,

```
MATRIX      ; c = [44,52,11,14/12,99,88,21/22,42,86,19] $
CROSSTAB    ; Lhs = c $
```

produces the following:

```
+-------------------------------------------------------------------+
|Cross Tabulation                                                   |
|Row variable is C          (Out of range 0-49:      0)             |
|Number of Rows =  3        (C         =  0 to  2)                  |
|Col variable is C          (Out of range 0-49:      0)             |
|Number of Cols =  4        (C         =  0 to  3)                  |
|Chi-squared independence tests:                                    |
|Chi-squared[   6] =   96.82121   Prob value =  .00000              |
|G-squared  [   6] =  101.97340   Prob value =  .00000              |
+-------------------------------------------------------------------+
|                      C                                            |
+--------+----------------------------+------+                      |
|       C|      0      1      2      3| Total|                      |
+--------+----------------------------+------+                      |
|       0|     44     52     11     14|   121|                      |
|       1|     12     99     88     21|   220|                      |
|       2|     22     42     86     19|   169|                      |
+--------+----------------------------+------+                      |
|   Total|     78    193    185     54|   510|                      |
+-------------------------------------------------------------------+
```

E4.5.4 An Application

The following example based on data about political ideology is given by Agresti (1984). The actual frequencies are

	Political Ideology			
Party Affiliation	Liberal	Moderate	Conservative	Total
Democrat	143	156	100	399
Independent	119	210	141	470
Republican	15	72	127	214
Total	277	438	368	1083

The commands are:

```
MATRIX      ; nij = [143,156,100 / 119,210,141 / 15,72,127] $
CROSSTAB    ; Lhs = nij
            ; Labels = democrat, indpndnt, repubcln  /
                       liberal, moderate, consrvtv
            ; Store = p,r,c,t $
MATRIX      ; List ; xtab ; ptab ; rtab ; ctab ; ttab $
```

```
+----------------------------------------------------------------------+
|Cross Tabulation                                                      |
|Row variable is NIJ      (Out of range 0-49:      0)                 |
|Number of Rows =  3      (NIJ      =  0 to  2)                        |
|Col variable is NIJ      (Out of range 0-49:      0)                 |
|Number of Cols =  3      (NIJ      =  0 to  2)                        |
|Chi-squared independence tests:                                       |
|Chi-squared[   4] =  102.04903   Prob value =  .00000                 |
|G-squared  [   4] =  105.66216   Prob value =  .00000                 |
+----------------------------------------------------------------------+
|                     NIJ                                              |
+--------+--------------------+------+                                 |
|     NIJ| LIBERA MODERA CONSRV| Total|                                |
+--------+--------------------+------+                                 |
|DEMOCRAT|    143    156    100|   399|                                |
|INDPNDNT|    119    210    141|   470|                                |
|REPUBCLN|     15     72    127|   214|                                |
+--------+--------------------+------+                                 |
|   Total|    277    438    368|  1083|                                |
+----------------------------------------------------------------------+

    XTAB|             1              2              3              4
--------+-------------------------------------------------------------
       1|       143.000        156.000        100.000        399.000
       2|       119.000        210.000        141.000        470.000
       3|       15.0000        72.0000        127.000        214.000
       4|       277.000        438.000        368.000        1083.00
    PTAB|             1              2              3              4
--------+-------------------------------------------------------------
       1|       102.053        161.368        135.579        399.000
       2|       120.212        190.083        159.705        470.000
       3|       54.7350        86.5485        72.7165        214.000
       4|       277.000        438.000        368.000        1083.00
    RTAB|             1              2              3              4
--------+-------------------------------------------------------------
       1|       .358396        .390977        .250627        1.00000
       2|       .253191        .446809        .300000        1.00000
       3|      .0700935        .336449        .593458        1.00000
       4|       .255771        .404432        .339797        1.00000
    CTAB|             1              2              3              4
--------+-------------------------------------------------------------
       1|       .516245        .356164        .271739        .368421
       2|       .429603        .479452        .383152        .433980
       3|      .0541516        .164384        .345109        .197599
       4|       1.00000        1.00000        1.00000        1.00000
    TTAB|             1              2              3              4
--------+-------------------------------------------------------------
       1|       .132041        .144044       .0923361        .368421
       2|       .109880        .193906        .130194        .433980
       3|      .0138504       .0664820        .117267        .197599
       4|       .255771        .404432        .339797        1.00000
```

E5: Descriptive Statistics for Time Series Data

E5.1 Introduction

This chapter will detail some of *LIMDEP*'s time series capabilities. Although *LIMDEP* is primarily oriented to cross section and panel data analysis, some common applications in time series can be handled as well.

E5.2 Box-Jenkins Time Series Identification

To produce a plot of the autocorrelations and partial autocorrelations for a time series variable, use the command

IDENTIFY ; Rhs = variable ; Pds = number of lags $

The number of lags is limited to one quarter of the sample size, and may not exceed 25. It will be reset to the limit value if necessary. The plot is accompanied by a tabulation of the Box-Pierce statistic at each lag for testing the hypothesis that the series is not autocorrelated. (Examples are shown below.)

This procedure creates a scalar named *nlag* which contains the value you give in **; Pds = nlag** and an *nlag*×2 matrix named *acf_pacf* which contains the results.

E5.2.1 Command Builder

The command builder for this computation is accessed by selecting Model:Time Series/Identify/Spectral. The dialog box for **IDENTIFY** requests only the variable (**; Rhs**), the number of lags (**; Pds**), and whether the Burg method is to be used for the partial autocorrelations (**; Alg = Burg**). The Burg specification is the only option for this command. See Figure E5.1.

E5.2.2 Computations

Calculations for this procedure are as follows: Data are $z_1, z_2, ..., z_T$. For lag length K+1, we assemble the following K variables:

$$d_0 = z_{K+1}, \dots z_T$$

$$d_1 = z_K, \dots , z_{T-1}$$

$$d_2 = z_{K-1}, \dots, z_{T-2}$$

$$\dots$$

$$d_K = z_1, \dots, z_{T-K}$$

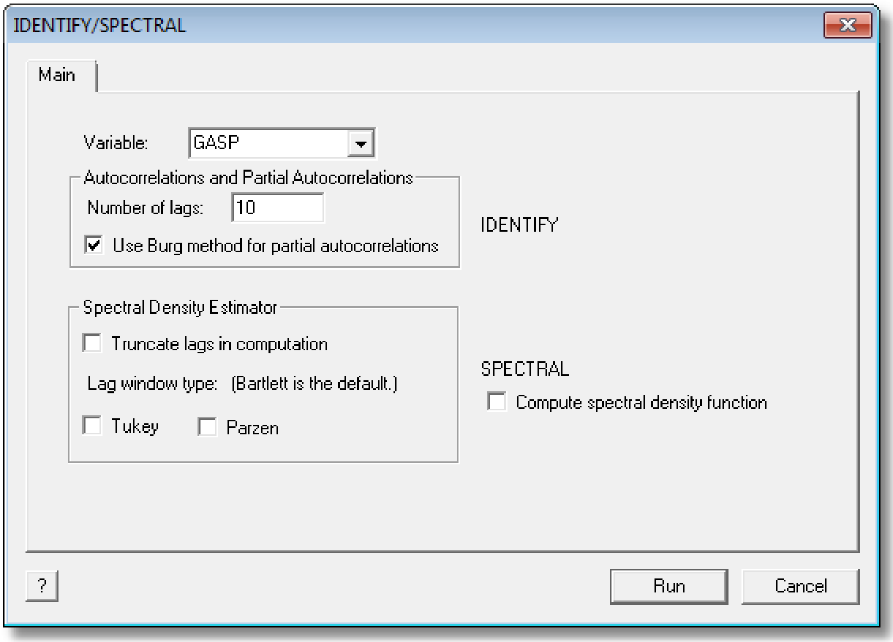


Figure E5.1 Command Builder for Time Series Description

Then, autocorrelations and other statistics are based on simple moments of these K+1 variables, each of which has exactly T-K observations. All moments are based on centered data. Under stationarity, the means should be asymptotically equivalent.

1. Autocovariances: c_i = sample covariance of $[d_i, d_0]$, $i = 0,...,K$,
2. Autocorrelations: r_i = sample correlation of $[d_i, d_0]$,
3. Partial correlations: $\mathbf{C}$ = $(K+1)\times(K+1)$ covariance matrix of ds,

   $\mathbf{c}$ = $K\times 1$ vector $[c_1,...,c_K]$,

   $\mathbf{C}_{(i)}$ = leading $i\times i$ principal submatrix of $\mathbf{C}$, $i = 1,...,K$,

   $\mathbf{c}_{(i)}$ = first i elements of $\mathbf{c}$. Then,

   $r_i{*}$ = last element of $[\mathbf{C}_{(i)}]^{-1}\mathbf{c}_{(i)}$..

For example, the first partial autocorrelation is c_1/c_0 while the second is the second element in the vector

$$\mathbf{r} = \begin{bmatrix} c_0 & c_1 \\ c_1 & c_0 \end{bmatrix}^{-1} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Finally, the Box-Pierce statistic at lag i is $BP_i = (T\text{-}K)\,\Sigma_i\, r_i.^2$. Two summary statistics are presented with the results:

$$\text{Box-Pierce} = (T-K)\sum_{s=1}^{J} r_s^2 ,$$

$$\text{Box-Ljung} = \frac{T}{T+2}\sum_{s=1}^{J} \frac{r_s^2}{T-s} .$$

E5.2.3 The Burg Variant of the PACF

McCullough (1999) and others have argued that using the Yule-Walker equations, as we do in the preceding section, to compute the partial autocorrelation functions produces results that can be numerically unstable and prone to numerical errors. Among the problems of the Yule-Walker approach is its reliance on inversion of the moment matrix. This will be problematic in and of itself in a highly autocorrelated data set. The Burg estimator is computed by a simple recursion, and is relatively more stable and less affected by numerical problems. A few programs have begun to implement the method, but it remains relatively new. The Burg method uses a fairly esoteric procedure documented further in McCullough. For the vector

$$\mathbf{z} = [z_1, z_2, \ldots , z_{T\text{-}1}, z_T]$$

define the circular shift operation

$$c(\mathbf{z}) = [z_T, z_1, z_2, \ldots , z_{T\text{-}1}]$$

and the subvector extraction operation

$$e_{j,k(\mathbf{z})} = [z_j, z_{j+1}, z_2, \ldots , z_{k\text{-}1}, z_k] = \text{elements } j \text{ through } k \text{ of } \mathbf{z}.$$

The dot product of two vectors, $\mathbf{x}$ and $\mathbf{z}$, which is usually denoted $\mathbf{x}'\mathbf{z}$ is computed by the function

$$d(\mathbf{x}, \mathbf{z}) = \mathbf{x}'\mathbf{z}$$

The squared norm of a vector $\mathbf{z}$ is, therefore, $d(\mathbf{z}, \mathbf{z}) = \|\mathbf{z}\|^2$. Finally, define the augmented $T+p$ vector padded with p zeros

$$\overrightarrow{\mathbf{z}}(0) = [z_1, z_2, \ldots , z_{T\text{-}1}, z_T, 0, \ldots, 0]$$

and $$\overleftarrow{\mathbf{z}}(0) = c[\overrightarrow{\mathbf{z}}(0)] = [0, z_1, z_2, \ldots , z_{T\text{-}1}, z_T, 0, \ldots, 0]$$

where the latter is obtained by shifting one of the zeros from the right end to the left so that p-1 zeros remain at the right. With all this in place, the Burg estimator of the partial autocorrelation coefficients is

$$r_i = \frac{2d\left(e_{i+1,T}\left(\overrightarrow{\mathbf{z}}(i-1)\right), e_{i+1,T}\left(\overleftarrow{\mathbf{z}}(i-1)\right)\right)}{\left\|e_{i+1,T}\left(\overrightarrow{\mathbf{z}}(i-1)\right)\right\|^2 + \left\|e_{i+1,T}\left(\overleftarrow{\mathbf{z}}(i-1)\right)\right\|^2}$$

where $$\overrightarrow{\mathbf{z}}(i) = \overrightarrow{\mathbf{z}}(i\text{-}1) - r_i \overleftarrow{\mathbf{z}}(i\text{-}1)$$

and $$\overleftarrow{\mathbf{z}}(i) = c\,[\overleftarrow{\mathbf{z}}(i\text{-}1) - r_i \overrightarrow{\mathbf{z}}(i\text{-}1)].$$

In spite of its seeming complexity, the Burg method is a relatively straightforward recursion which begins with r_1 = the first autocorrelation. The Burg method is requested by adding

; Alg = Burg

to the **IDENTIFY** command.

TECHNICAL NOTE: Here is the algorithm used internally for computing the Burg estimator.

1. Set $N = T + J$, T = number of observations, J = number of partial autocorrelations.

2. Initialize $\mathbf{z}_L = [z_1, z_2, ..., 0, 0, ..., 0]$ = N elements, T of z_t and J zeros.
   Initialize $\mathbf{z}_R$ = circular shift of $\mathbf{z}_L = [0, z_1, z_2, ..., 0, 0, ..., 0]$, 0, $\mathbf{z}$, J-1 zeros.

3. For $i = 1, ..., J$ in recursion, i starts at 1

   a. Move $T - i$ elements starting at i+1 from $\mathbf{z}_R$ into vector $\mathbf{d}_R$,
   b. Move $T - i$ elements starting at i+1 from $\mathbf{z}_L$ into vector $\mathbf{d}_L$,
   c. $r_i = 2\, \mathbf{d}_R'\mathbf{d}_L / (\mathbf{d}_R'\mathbf{d}_R + \mathbf{d}_L'\mathbf{d}_L)$ (dot products involve T-i values),
   d. Make $\mathbf{d}_R = \mathbf{z}_R - r_i\, \mathbf{z}_L$,
   e. Make $\mathbf{d}_L = \mathbf{z}_L - r_i\, \mathbf{z}_R$,
   f. Set first element of $\mathbf{z}_L$ equal to Nth element of $\mathbf{d}_L$,
   g. Move first N-1 elements from $\mathbf{d}_L$ into $\mathbf{z}_L$ starting at second position in $\mathbf{z}_L$,
   h. Move N elements from $\mathbf{d}_R$ into $\mathbf{z}_R$,
   i. Increment i and return to a.

E5.2.4 Application

The data listed in Table E5.1 apply to the U.S. gasoline market from 1953 to 2004. (This data set was assembled from several sources by Professor Chris Bell, Department of Economics, University of North Carolina, Asheville.) We will construct examples using this data set. The data are provided in files named gas.lpj, gas.dat, and gas.csv. We begin by computing the autocorrelations and partial autocorrelations for the gasoline price variable. Relying on the canon of Box and Jenkins, we conclude from this figure that these data are clearly characterized by an AR(1) process.

Based on the Yule-Walker Equations:

IDENTIFY ; Rhs = gasp ; Pds = 10 $

```
Time series identification for GASP
Box-Pierce Statistic =   229.6895      Box-Ljung Statistic  =   260.4700
Degrees of freedom   =         10      Degrees of freedom   =         10
Significance level   =      .0000      Significance level   =      .0000
* => |coefficient| > 2/sqrt(N) or > 95% significant.
PACF is computed using Yule-Walker equations.
----+-------------------------------+-------+--------------------------------+
Lag |  Autocorrelation Function     |Box/Prc|    Partial Autocorrelations    |
----+-------------------------------+-------+--------------------------------+
  1 | .907*|           |**********  | 42.82*| .907*|            |**********  |
  2 | .828*|           |*********   | 78.48*|-.211 |         ** |            |
  3 | .781*|           |*********   |110.18*| .198 |            |**          |
  4 | .721*|           |********    |137.20*|-.122 |          * |            |
  5 | .643*|           |*******     |158.71*|-.139 |         ** |            |
  6 | .597*|           |*******     |177.26*| .032 |            |*           |
  7 | .567*|           |******      |193.99*| .093 |            |*           |
  8 | .522*|           |******      |208.14*|-.027 |          * |            |
  9 | .476*|           |*****       |219.91*| .006 |            |*           |
 10 | .434*|           |*****       |229.69*|-.068 |          * |            |
----+-------------------------------+-------+--------------------------------+
```

Based on the Burg Method:

IDENTIFY ; Rhs = gasp ; Pds = 10 ; Alg = Burg $

(*Note that the first autocorrelation is adjusted to equal the first partial autocorrelation, not the reverse.*)

```
Time series identification for GASP
Box-Pierce Statistic =   229.6895      Box-Ljung Statistic  =   260.4700
Degrees of freedom   =         10      Degrees of freedom   =         10
Significance level   =      .0000      Significance level   =      .0000
* => |coefficient| > 2/sqrt(N) or > 95% significant.
PACF and Rho(1) computed using method of Burg
----+-------------------------------+-------+--------------------------------+
Lag |  Autocorrelation Function     |Box/Prc|    Partial Autocorrelations    |
----+-------------------------------+-------+--------------------------------+
  1 | .972*|           |*********** | 49.16*| .972*|            |***********|
  2 | .828*|           |*********   | 84.82*|-.313*|        *** |            |
  3 | .781*|           |*********   |116.53*| .316*|            |***         |
  4 | .721*|           |********    |143.54*|-.296*|        *** |            |
  5 | .643*|           |*******     |165.05*|-.100 |          * |            |
  6 | .597*|           |*******     |183.60*| .075 |            |*           |
  7 | .567*|           |******      |200.33*| .099 |            |*           |
  8 | .522*|           |******      |214.48*|-.057 |          * |            |
  9 | .476*|           |*****       |226.25*| .076 |            |*           |
 10 | .434*|           |*****       |236.03*|-.162 |         ** |            |
----+-------------------------------+-------+--------------------------------+
```

Year	GasQ	GasP	GasCPIU	PCInc	PNC	PUC	PPT	PN	PD	PS	POP
1953	25.415	16.668	21.2	8802	47.2	26.7	16.8	37.7	29.7	19.4	159565
1954	26.223	17.029	21.8	8757	46.5	22.7	18.0	36.8	29.7	20.0	162391
1955	28.505	17.210	22.1	9177	44.8	21.5	18.5	36.1	29.5	20.4	165275
1956	30.229	17.729	22.8	9450	46.1	20.7	19.2	36.1	29.9	20.9	168221
1957	31.393	18.497	23.8	9508	48.5	23.2	19.9	37.2	30.9	21.8	171274
1958	32.884	18.316	23.5	9433	50.0	24.0	20.9	37.8	31.7	22.6	174141
1959	34.573	18.576	23.7	9685	52.2	26.8	21.5	38.4	31.5	23.3	177130
1960	35.757	19.112	24.4	9735	51.5	25.0	22.2	38.1	32.0	24.1	180760
1961	36.126	18.924	24.1	9901	51.5	26.0	23.2	38.1	32.2	24.5	183742
1962	37.658	19.043	24.3	10227	51.3	28.4	24.0	38.5	32.5	25.0	186590
1963	38.815	18.997	24.2	10455	51.0	28.7	24.3	38.6	32.9	25.5	189300
1964	40.940	18.873	24.1	11061	50.9	30.0	24.7	39.0	33.2	26.0	191927
1965	42.874	19.587	25.1	11594	49.7	29.8	25.2	38.8	33.8	26.6	194347
1966	45.549	20.038	25.6	12065	48.8	29.0	26.1	38.9	35.1	27.6	196599
1967	47.029	20.700	26.4	12457	49.3	29.9	27.4	39.4	35.7	28.8	198752
1968	50.304	21.005	26.8	12892	50.7	30.7	28.7	40.7	37.1	30.3	200745
1969	53.686	21.696	27.7	13163	51.5	30.9	30.9	42.2	38.9	32.4	202736
1970	57.009	21.890	27.9	13563	53.0	31.2	35.2	44.1	40.8	35.0	205089
1971	59.770	22.050	28.1	14001	55.2	33.0	37.8	46.0	42.1	37.0	207692
1972	62.206	22.336	28.4	14512	54.7	33.1	39.3	46.9	43.5	38.4	209924
1973	65.440	24.473	31.2	15345	54.8	35.2	39.7	48.1	47.5	40.1	211939
1974	62.217	33.059	42.2	15094	57.9	36.7	40.6	51.5	54.0	43.8	213898
1975	64.070	35.278	45.1	15291	62.9	43.8	43.5	57.4	58.3	48.0	215981
1976	66.633	36.777	47.0	15738	66.9	50.3	47.8	60.9	60.5	52.0	218086
1977	68.675	38.907	49.7	16128	70.4	54.7	50.0	64.4	64.0	56.0	220289
1978	70.258	40.597	51.8	16704	75.8	55.8	51.5	68.6	68.6	60.8	222629
1979	69.315	54.406	70.2	16931	81.8	60.2	54.9	75.4	77.2	67.5	225106
1980	65.358	75.509	97.5	16940	88.4	62.3	69.0	83.0	87.6	77.9	227726
1981	66.349	84.018	108.5	17217	93.7	76.9	85.6	89.6	95.2	88.1	230008
1982	67.176	79.768	102.8	17418	97.4	88.8	94.9	95.1	97.8	96.0	232218
1983	68.676	77.160	99.4	17828	99.9	98.7	99.5	99.8	99.7	99.4	234333
1984	70.833	76.005	97.8	19011	102.8	112.5	105.7	105.1	102.5	104.6	236394
1985	72.225	76.619	98.6	19476	106.1	113.7	110.5	106.8	104.8	109.9	238506
1986	75.734	60.175	77.0	19906	110.6	108.8	117.0	106.6	103.5	115.4	240683
1987	77.762	62.488	80.1	20072	114.6	113.1	121.1	108.2	107.5	120.2	242843
1988	79.810	63.017	80.8	20740	116.9	118.0	123.3	110.4	111.8	125.7	245061
1989	81.515	68.837	88.5	21120	119.2	120.4	129.5	112.2	118.2	131.9	247387
1990	80.750	78.385	101.0	21281	121.0	117.6	142.6	113.4	126.0	139.2	250181
1991	79.862	77.338	99.2	21109	125.3	118.1	148.9	116.0	130.3	146.3	253530
1992	83.093	77.040	99.0	21548	128.4	123.2	151.4	118.6	132.8	152.0	256922
1993	85.204	76.257	97.7	21493	131.5	133.9	167.0	121.3	135.1	157.9	260282
1994	86.338	76.614	98.2	21812	136.0	141.7	172.0	124.8	136.8	163.1	263455
1995	87.935	77.826	99.8	22153	139.0	156.5	175.9	128.0	139.3	168.7	266588
1996	89.888	82.596	105.9	22546	141.4	157.0	181.9	129.4	143.5	174.1	269714
1997	92.666	82.579	105.8	23065	141.7	151.1	186.7	128.7	146.4	179.4	272958
1998	96.941	71.874	91.6	24131	140.7	150.6	190.3	127.6	146.9	184.2	276154
1999	100.351	78.207	100.1	24564	139.6	152.0	197.7	126.0	151.2	188.8	279328
2000	100.000	100.000	128.6	25472	139.6	155.8	209.6	125.4	158.2	195.3	282429
2001	101.481	96.289	124.0	25698	138.9	158.7	210.6	124.6	160.6	203.4	285366
2002	102.871	90.405	116.0	26229	137.3	152.0	207.4	121.4	161.1	209.8	288217
2003	103.587	105.154	135.1	26570	134.7	142.9	209.3	117.5	165.3	216.5	291073
2004	103.245	123.901	159.7	27208	133.9	133.3	209.1	114.8	172.2	222.8	293951

GasQ = quantity index of gasoline consumption
GasP = price index for gasoline
GasCPIU = consumer price index
PCInc = real per capita disposable income
PNC = price index for new cars
PUC = price index for used cars
PPT = price index for public transportation
PN = aggregate price index for consumer nondurables
PD = aggregate price index for consumer durables
PS = aggregate price index for consumer services
POP = population in thousands

Table E5.1 Data on the U.S. Gasoline Market

Figure E5.2 shows the new matrix created by the second **IDENTIFY** command.

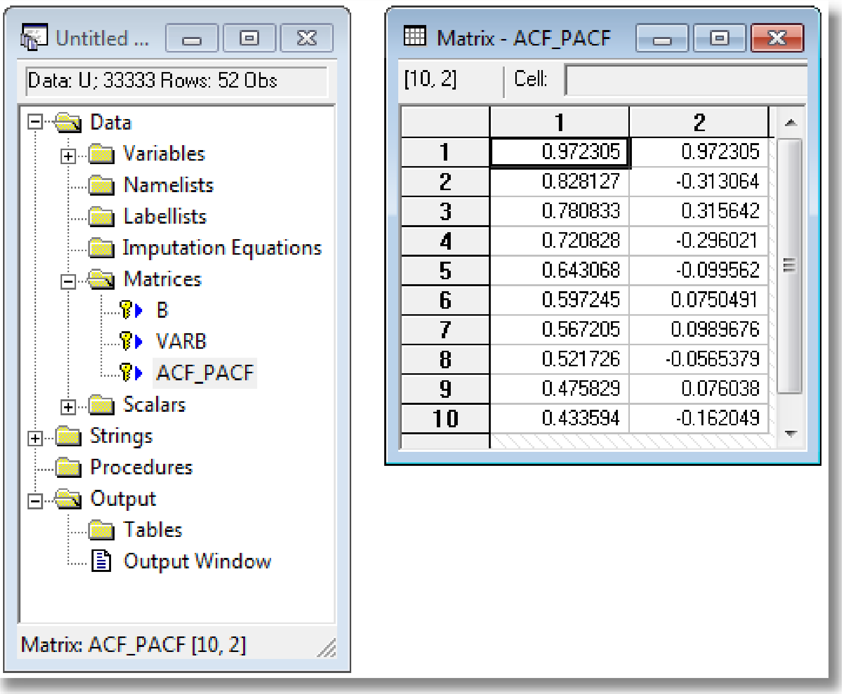


Figure E5.2 Matrix Results from IDENTIFY Command

E5.3 Spectral Density Estimation

The basic command for requesting a spectral density estimator is

SPECTRAL ; Rhs = variable $

The spectral density function or periodogram is used to decompose the variance of a time series. The underlying assumption is that the time series x_t can be written as a weighted sum of underlying series, z_t each varying at a different frequency. The variation in x_t is then decomposed into the contributions made at the different frequencies by each of these underlying series. (Hamilton (1994) is a good reference on time series analysis. Chatfield (1996) is an alternative, somewhat less technical reference to consider.)

We compute the spectrum as follows: (Again, see one of the aforementioned sources for background): For frequencies ranging from 0 to π (the function is symmetric from π to 2π), the spectral density function is computed as the following function of the autocorrelations, r_j:

$$\omega_i = \left(\frac{i-1}{T-1}\right)\pi\,,\ i = 1,\ldots,T,$$

$$f(\omega_i) = \frac{1}{2\pi}\sum_{j=1}^{L} w_j\left[2r_j\cos(j\omega_i)\right],\ L = T\text{-}1 \text{ or } 2\sqrt{T} \text{ if the function is truncated,}$$

where $$r_j = \frac{\sum_{t=j+1}^{T}(x_t - \bar{x})(x_{t-j} - \bar{x})}{\sum_{t=j+1}^{T}(x_t - \bar{x})^2}.$$

The weighting function, w_j is the 'lag window.' Lag windows are used to stabilize the function and to reduce the influence of unstable autocorrelations at longer lags. In addition to the default window (none), $w_j = 1$, you may use any of

Bartlett's window: $$w_j = 1 - \frac{j}{T-1} = B_j$$

Tukey's window: $$w_j = \frac{1}{2}\left[1 + \cos\left(\frac{j\pi}{T-1}\right)\right]$$

Parzen's window: $$w_j = 1 - 6B_j(1 - B_j)^2 \text{ if j} \leq T/2$$

$$= 2B_j^2 \text{ if } j > T/2$$

A second method of stabilizing the estimated function is to truncate the lags in the summation. Note, in the definition, L is the lag length, T-1, or $2\sqrt{T}$ if you elect to truncate the sum.

The general command

SPECTRAL ; Rhs = the variable
; Window = Bartlett, Tukey, or Parzen (optional)
; Truncate (also optional) $

requests these computations. Output from the procedure consists of the plot of the periodogram and a matrix named *freq_sdf* which has two columns. The first is the frequencies, and the second is the function values.

The command builder for **SPECTRAL** uses the same dialog box as **IDENTIFY**, as shown in Figure E5.3

NOTE: The default command in this command builder is **IDENTIFY**. To produce the **SPECTRAL** command, you must check the box to the right of the grouping for the spectral density estimator. (This is marked by the arrow in the figure.)

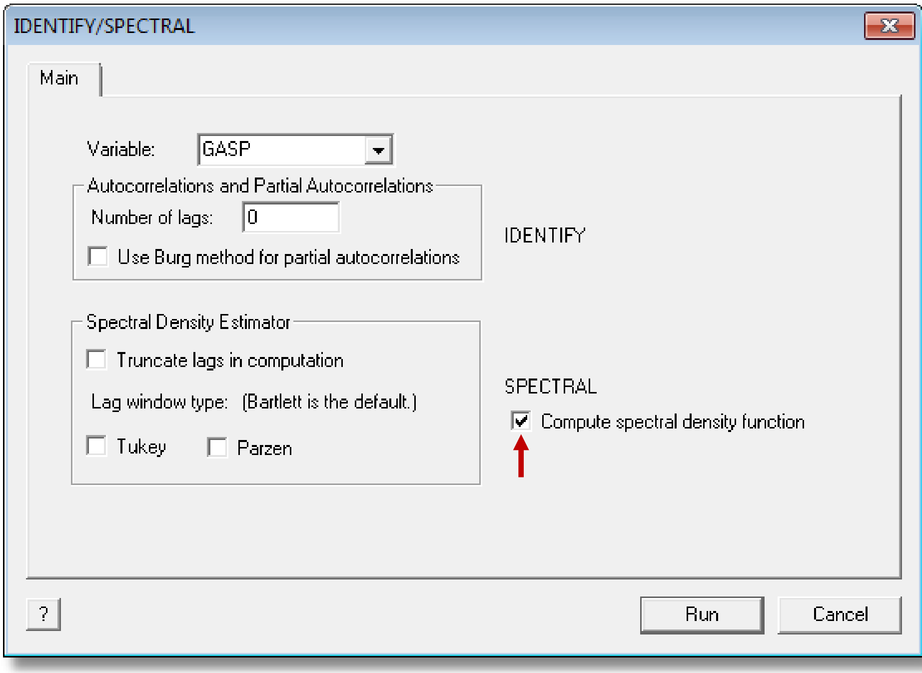


Figure E5.3 Command Builder for Spectral Density Estimation

For example, Figure E5.4 shows the spectral density estimator for the log of per capita income in the gasoline data.

CREATE ; logpcinc = Log(pcincome) $
SPECTRAL ; Rhs = logpcinc $

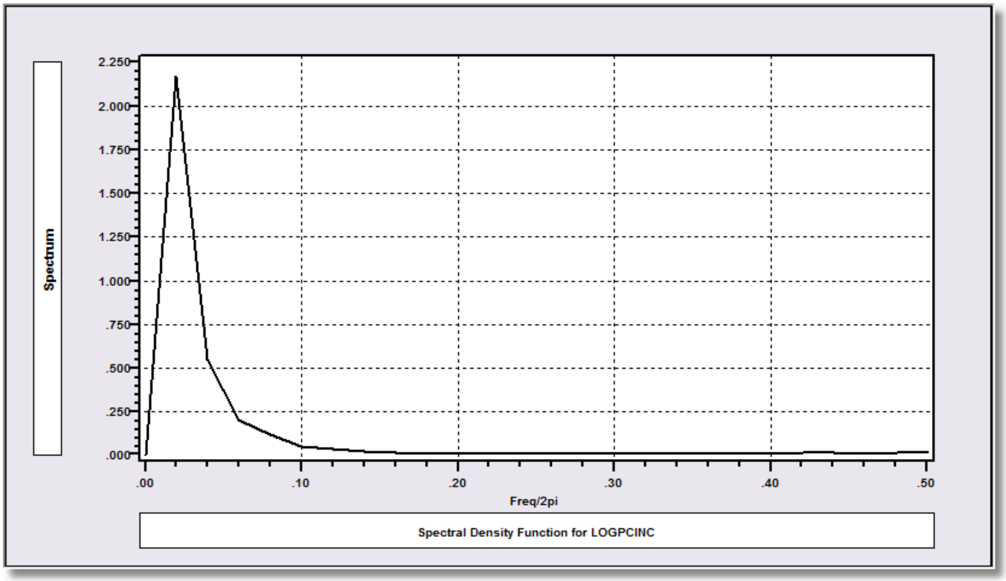


Figure E5.4 Spectral Density Estimator for Log Per Capita Income

E5.4 Phillips-Perron Test for a Unit Root

Various devices have been presented for testing for unit roots in time series data. Most familiar is the Dickey-Fuller (1979) test, which is carried out simply by referring familiar regression statistics to the appropriate table. (See the next section.) Another in wide use is the Phillips-Perron (1988) test, which is carried out by subjecting the residuals from the regression

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$

to a prescribed test procedure. The regression may be computed with additional regressors, a time trend, or without the constant term. The cases in the Phillips-Perron derivation are:

Case 1: No constant term
Case 2: Constant term
Case 4: Constant term and time trend.

(Case 3 is not considered here. See Hamilton (1994) for details.) The procedure is based on two statistics,

$$Z_\rho = T(\hat{\rho}-1) - \frac{1}{2}\left(\frac{T^2 v^2}{s^2}\right)(a - c_0)$$

$$Z_\tau = \sqrt{\frac{c_0}{a}}\left(\frac{\hat{\rho}-1}{v}\right) - \frac{1}{2}(a - c_0)\frac{Tv}{\sqrt{as^2}}$$

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-K}$$

v^2 = estimated asymptotic variance of $\hat{\rho}$

$$c_j = \frac{1}{T}\sum_{s=j+1}^{T} e_t e_{t-s}\ ,\ \text{j} = 0,\ldots,L = j\text{th autocorrelation of residuals}$$

c_0 = $[(T\text{-}K)/T]s^2$.)

$$a = c_0 + 2\sum_{j=1}^{L}\left(1 - \frac{j}{L+1}\right)c_j$$

The test statistics are referred to the Dickey-Fuller tables. *LIMDEP* uses linear interpolation in a few critical values from the tables: For each statistic, the internal values are for significance levels of .01, .05, and .10, and sample sizes 25, 20, 100, 250, 500 and ∞. The T=25 value is used if the sample is under 25. The value for ∞ is used if T is greater than 500. The hypothesis of a unit root is rejected if the test statistics are less than the critical values given.

The Phillips-Perron test is requested by setting up the AR(1) regression, and adding

; PPT ; Pds = L

where L is the desired number of Newey-West lags for the computation. (Note the definition of a above.)

NOTE: The Newey-West autocorrelation consistent covariance matrix for the OLS estimator is not available when you request the Phillips-Perron test.

You can request this estimator in the command builder by setting up the regression as a linear model, then choosing the Phillips-Perron test instead of the Newey-West estimator. (See the Options page of the **REGRESSION** command builder, which can be found by selecting Model:Linear Models/Regression.)

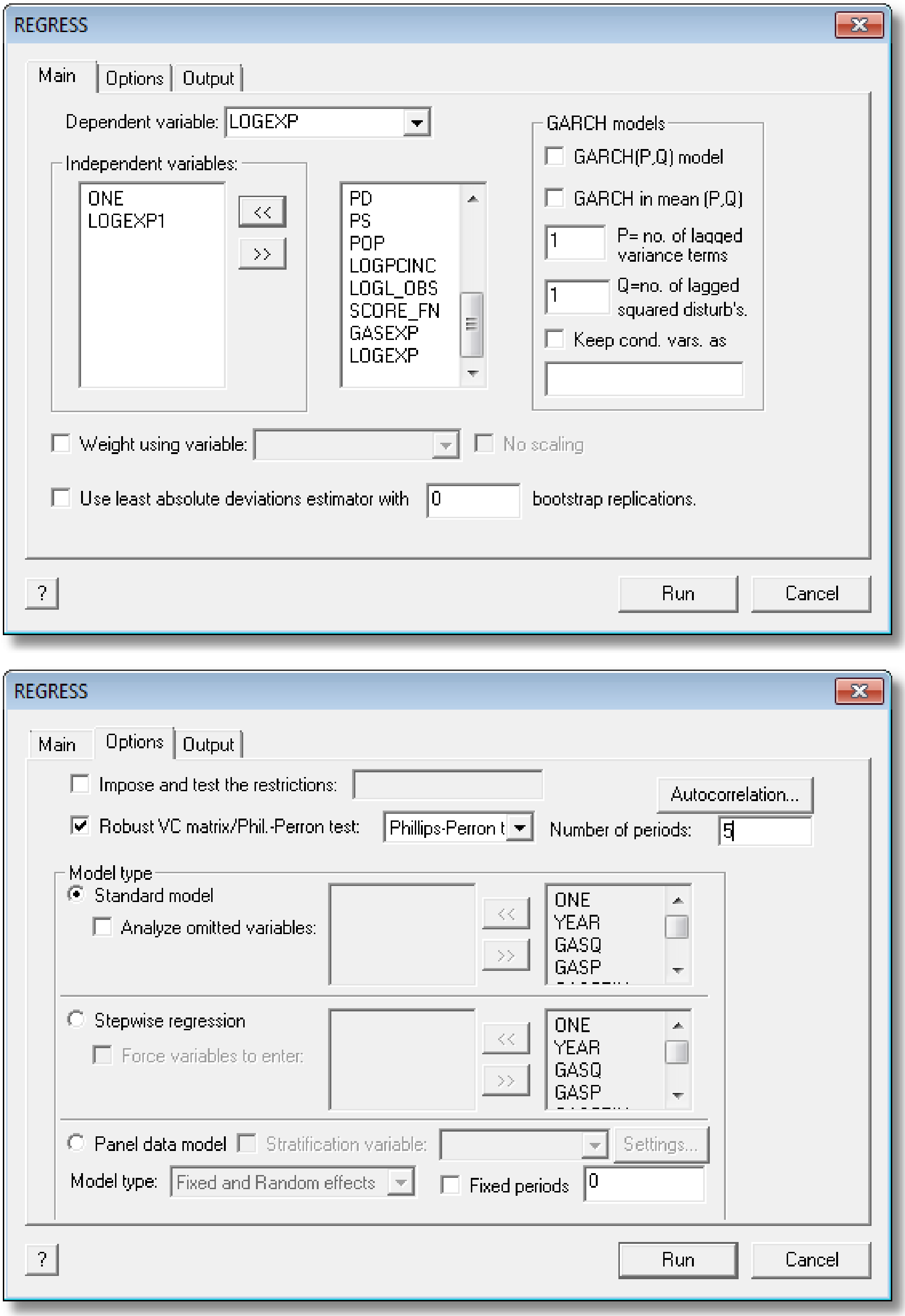


Figure E5.5 Command Builder for Phillips-Perron Test for a Unit Root

For the sample used in the previous example, we have applied the test to the log of per capita gasoline expenditure. The command sequence is

```
SAMPLE     ; All $
CREATE     ; logexp = Log(gasp*gasq) $
CREATE     ; logexp1 = logexp[-1] $
SAMPLE     ; 2-52 $
REGRESS    ; Lhs = logexp
           ; Rhs = one,logexp1
           ; PPT ; Pds = 10 $
```

The results are shown below. Note that the statistic and the critical values are presented in a separate table, with no further action taken. This procedure generates no additional retained values beyond those generated by the regression as usual. The statistics are small and negative, suggesting that the hypothesis of a unit root should not be rejected.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGEXP   Mean                    =       7.87249
             Standard deviation      =       1.03681
             No. of observations     =            51  Degrees of freedom
Regression   Sum of Squares          =       53.4018          1
Residual     Sum of Squares          =       .347364         49
Total        Sum of Squares          =       53.7491         50
             Standard error of e     =        .08420
Fit          R-squared               =        .99354  R-bar squared =    .99341
Model test   F[  1,     49]          =    7532.98233  Prob F > F*   =    .00000
Diagnostic   Log likelihood          =      54.85893  Akaike I.C.   = -4.91078
             Restricted (b=0)        =     -73.70467
             Chi squared [  1]       =     257.12719  Prob C2 > C2* =    .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGEXP|  Coefficient      Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|     .13434        .08993     1.49  .1416     -.04192    .31061
 LOGEXP1|     .99135***     .01142    86.79  .0000      .96896   1.01374
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+-----------------------------------------------------------------------+
| Phillips - Perron tests for unit root                                 |
| Sample size  =    51, Number of regressors =    2                     |
| Sample statistics:        Z(tau) =     -.7386 |Z(rho) =     -.4939    |
|                           ------[Z(tau)]------+------[Z(rho)]------   |
| 3 cases (models)            99%     95%     90% | 99%    95%    90%   |
| 1. y(t)=ry(t-1)+u(t)       -2.62   -1.95   -1.61|-12.90  -7.70  -5.50 |
| 2. y(t)=a+ry(t-1)+u(t)     -3.58   -2.93   -2.60|-18.90 -13.30 -10.70 |
| 4. y(t)=a+ry(t-1)+dt+u(t)  -4.25   -3.80   -3.18|-25.70 -19.80 -16.80 |
+-----------------------------------------------------------------------+
```

E5.5 Augmented Dickey-Fuller Tests

The Adf function in **CALC** automates the Dickey-Fuller test for unit roots in time series data. The syntax is

CALC ; Adf (variable, type, lags for augmentation) $

where	*variable*	is the single time series variable to be analyzed,
	type = 1, 2 or 3	for unit root, drift, trend, lags $\geq$ 0,
	lags for augmentation	is the number of additional lagged values to include.

Users are referred to any of the standard texts, e.g., Greene (2011, Chapter 21) for details.

The Phillips-Perron test for a unit root in per capita gasoline expenditure in the preceding section is recalculated here, using the Adf computation instead.

CALC ; Adf (logexp,1,3) $

```
+-----------------------------------------------+
| Augmented Dickey Fuller Test for LOGEXP       |
| Form: Random walk                             |
| Number of lagged differences in model is   3 |
| DF(tau) =    2.44635, DF(gamma) =     .40584 |
| Critical values for    47 observations:      |
| DF(tau)                                       |
| 01 is  -2.62, .025 is  -2.25, .05 is  -1.95  |
| DF(gamma)                                     |
| 01 is -12.80, .025 is  -9.90, .05 is  -7.70  |
+-----------------------------------------------+
```

As before, the statistics are between zero and the critical values, indicating that the hypothesis of a unit root is not rejected.

E6: Scatter Diagrams and Plotting

E6.1 Introduction

This chapter will describe commands for producing high resolution graphs. This feature can be used for simple scatter diagrams, time series plots, and for plotting functions such as log likelihoods. You can print the graphics on standard printers and create plotter files for export to word processing programs such as Microsoft *Word*. Generalities about plotting and the following five commands are described in this chapter:

PLOT	is the standard scatter plotting function,
SPLOT	is for producing several scatter plots at the same time,
MPLOT	is for plotting the elements of matrices,
FPLOT	is for plotting functions of one variable,
CPLOT	is used for creating contour plots

E6.2 Printing and Exporting Figures

The following procedures apply to the various plotting commands described in this chapter as well as to the other uses of graphics in *LIMDEP*, which include: **KERNEL**, **HISTOGRAM**, **SPECTRAL**, **SIMULATE ; Plot**, **PARTIALS ; Plot**, **SURVIVAL**, **DSTAT ; Boxplots**, **DSTAT ; Quantiles ; Plot**, **REGRESS ; Cusums** and **EXECUTE ; Bootstrap**.

LIMDEP uses the standard Windows interface between input and output devices. When a plot appears in a window, you can use File:Print to send a copy to your printer (see Section E6.2.1). You can also save the graph as a Windows metafile (.wmf format) by using File:Save or File:Save As (see Section E6.2.2). Finally, you can simply use Edit:Copy and Edit:Paste to transport the graphic figure into another software program (see Section E6.2.3). For purposes of illustrating these functions, we will use Figure E6.1 which was generated by a **PLOT** command. The figure shows *LIMDEP*'s base format for graphics. Every graph generated is placed in its own scalable window, as shown in Figure E6.2, apart from the project, editing and output windows already open. This window will remain open until you close it. When you are finished reviewing the figure, you should close the window to avoid proliferating windows. You will be prompted to save the graph if you have not already done so.

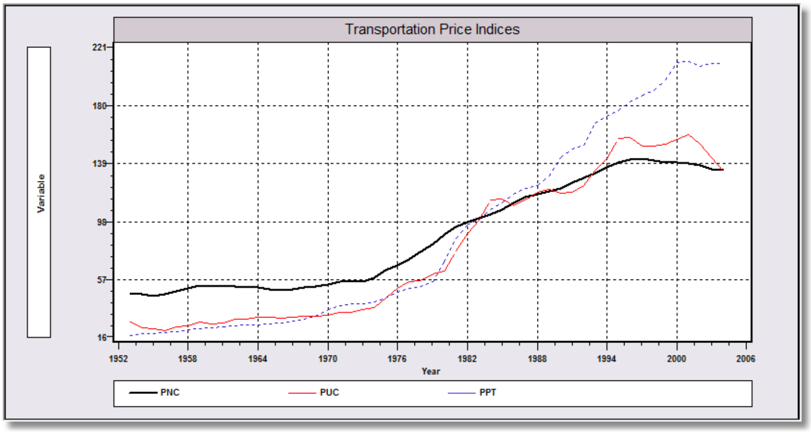


Figure E6.1 Time Series Plot Using the PLOT Command

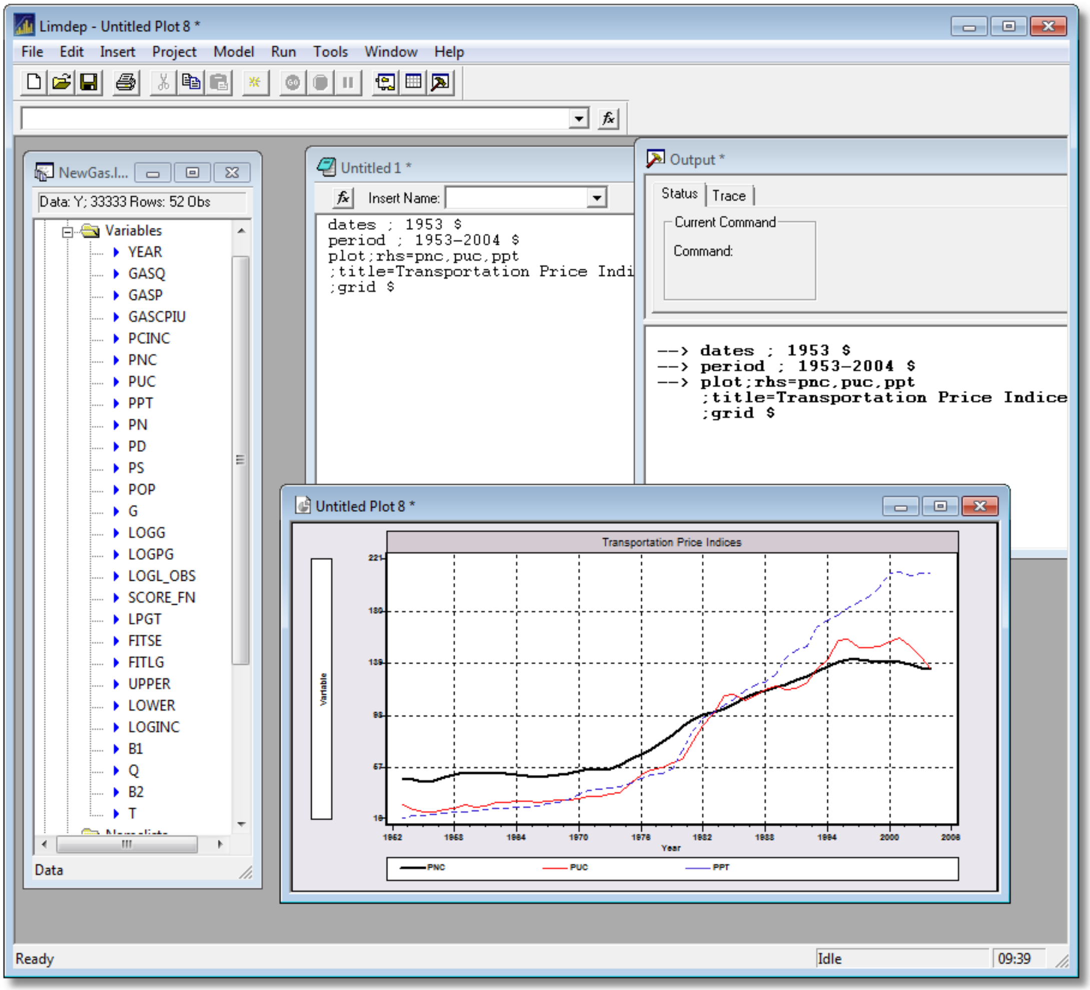


Figure E6.2 *LIMDEP* Desktop with Graphics Window

E6.2.1 Printing

LIMDEP prints graphs in landscape mode. (You can change this to portrait mode by using Page Setup in the file menu.) To print a graph, first make the plot window the active one by clicking the large banner at the top of the window (marked 'Untitled Plot 8*' in Figure E6.2 above). Then, to print without adjustment, select File:Print from the main menu. If you wish to change the size and orientation of a graph, select File:Page Setup first. Windows will handle the printer interface, as in other operations.

TIP: In *LIMDEP*, click the right mouse button inside the graphics window to open a menu with options to Copy, Save As, Page Setup, Print Preview and Print.

E6.2.2 Saving a Graph as a Graphics File

You may save any figure from a graphics window to disk in the Windows metafile (.wmf) format. Select File:Save or File:Save As. This file type is transportable to many other programs, including Microsoft *Word* and *Excel*. For example, in *Word*, click the Insert menu and then select Picture, then From File to import your .wmf file into your *Word* document. The Windows .wmf format includes codes that allow you to scale the figure to whatever size you desire.

E6.2.3 Pasting a Graph into a Document or Spreadsheet

You can also put a copy of your graph in *Word* or *Excel* without first saving it as a .wmf file. Select Edit:Copy in *LIMDEP*, and then select Edit:Paste in your other software to paste the graph into your document or spreadsheet. An example is shown in Figure E6.3. We have used Edit:Copy in *LIMDEP* followed by Edit:Paste in *Excel*, first to transport the data from *LIMDEP*'s data editor, then to transport the scatter plot. Once in *Excel*, the data and accompanying graphic can be viewed together and the graphic can be formatted. In our example, the graphic was resized and a line box was added. The formats used for graphics by *LIMDEP* are standards that are used generally in other commercial packages.

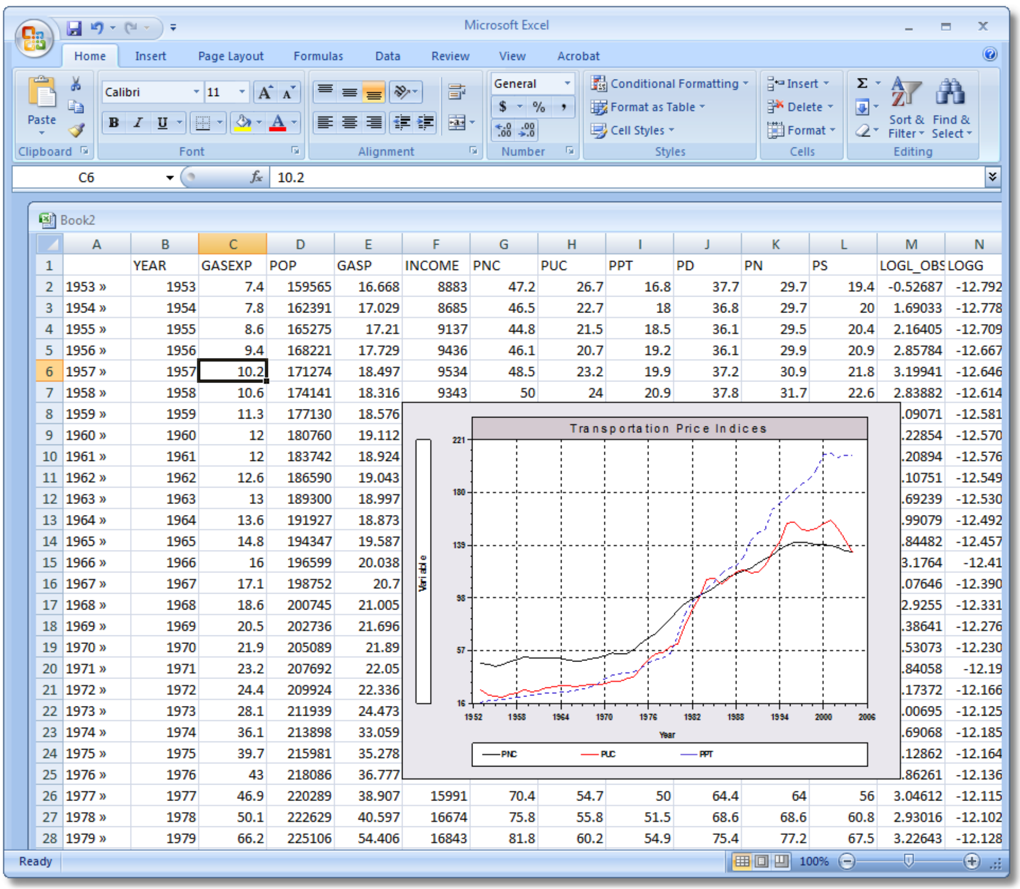


Figure E6.3 *Excel* Spreadsheet with *LIMDEP* Data and Graph Imported

E6.3 The PLOT Command

The command for producing a basic scatter (*XY*) plot of one or more variables against another variable is

PLOT ; Lhs = variable on horizontal axis
; Rhs = variables (up to five) on vertical axis $

Note the reversal of LIMDEP's usual convention. This command puts the Lhs variable on the horizontal axis, whereas a regression might be expected to do the reverse.

You may add a title to the figure by including

; Title = the title to be used

The title is placed at the top of the figure. The vertical axis of the plot is usually labeled with some variable name. You can override this with

; Yaxis = the label to be used, up to eight characters

This will often be useful when you plot a function or more than one variable. A longer descriptive label for the vertical axis can be provided with

; Vaxis = the test string to be used, up to 60 characters

The command builder for **PLOT** may be found by selecting Model:Data Description/Plot Variables. See Figures E6.4 and E6.5.

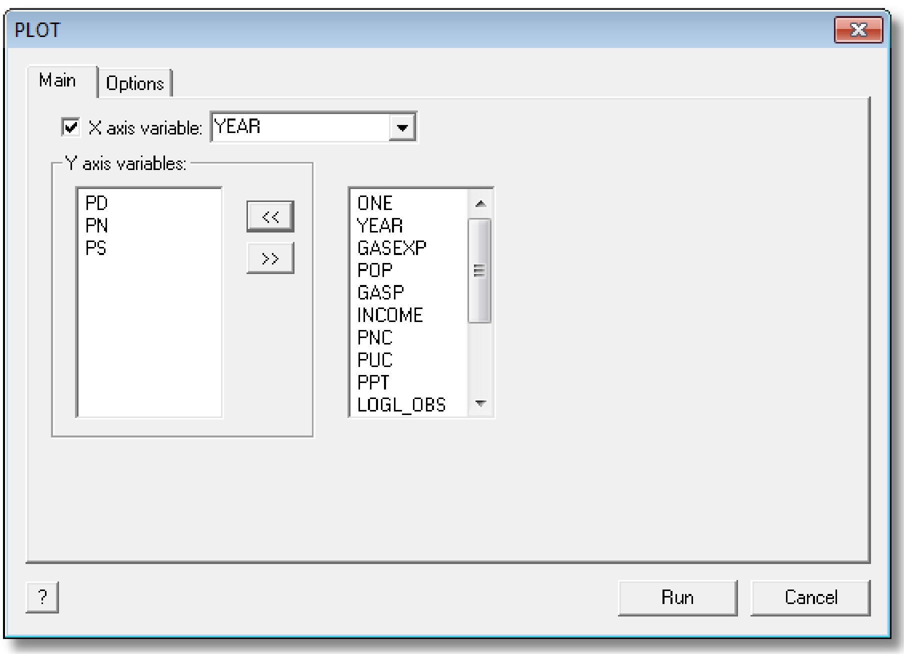


Figure E6.4 Main Page of Command Builder for PLOT

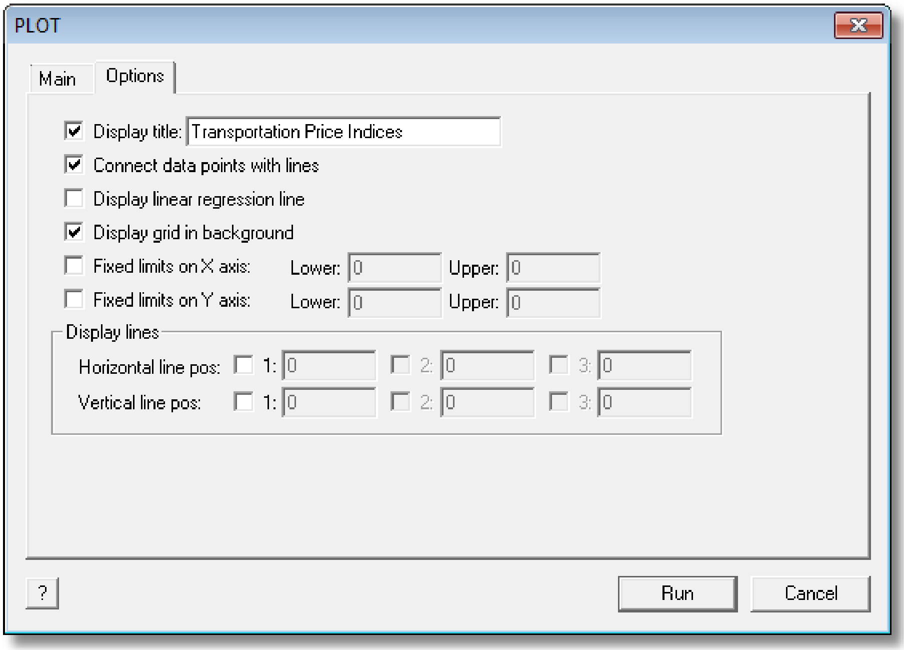


Figure E6.5 Options Page of Command Builder for PLOT

NOTE: The figures displayed here are based on yearly data for the U.S. gasoline market: The data are listed in Section E5.2.4 in Table E5.1.

E6.3.1 Scatter Plot of One Variable Against Another

To produce a scatter plot of one variable (*y*) against another (*x*) variable, the **PLOT** command is given with only a single Rhs variable. The command would be

PLOT ; Lhs = x ; Rhs = y $

Figure E6.6 was produced with the commands listed.

```
CREATE   ; g = gasq/(100*pop/282429) $
PLOT     ; Lhs = g
         ; Rhs = gasp
         ; Title = Simple Plot of Gas against Price $
```

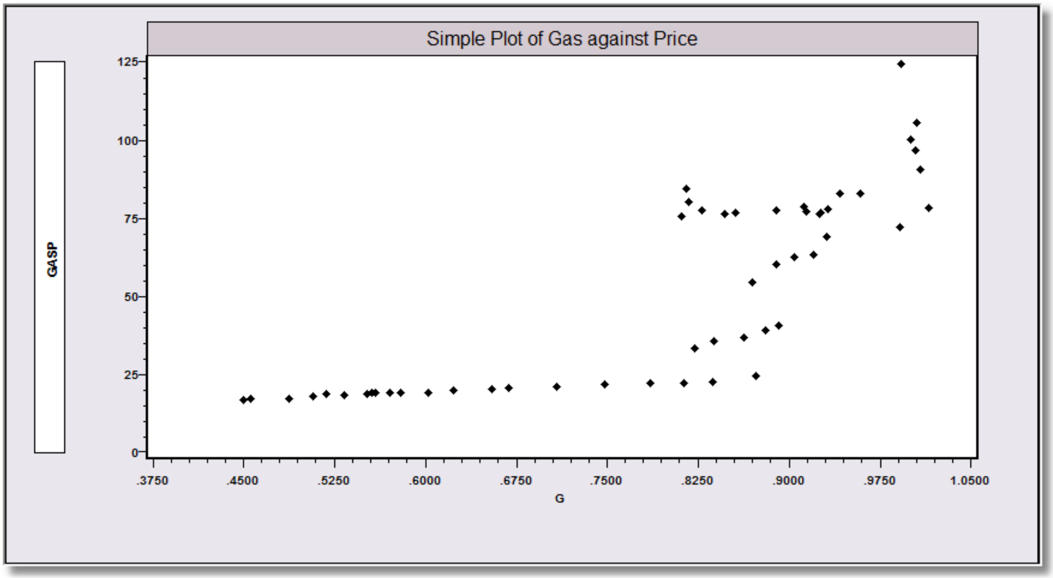


Figure E6.6 Simple Scatter Plot

E6.3.2 Plotting a Simple Linear Regression

To add a regression line to a figure, add

; Regression

to the **PLOT** command. By adding **; Regression** to the preceding command, we obtain the plot in Figure E6.7. You can also obtain this by selecting Display linear regression line in the Options page of the command builder. (In previous versions of *LIMDEP*, the regression equation would replace the title in the figure. In this version, the title appears as a header and the regression equation is placed in the legend at the right of the figure.)

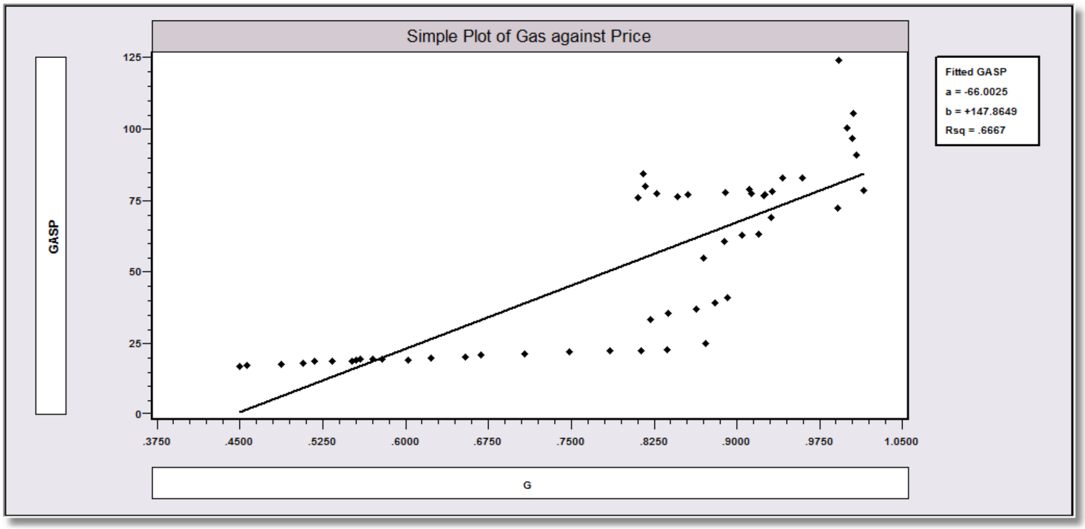


Figure E6.7 Scatter Plot with Linear Regression

E6.3.3 Time Series Plots

Time series plots, that is, plots of variables against the date can be obtained by using **DATES** and **PERIOD** to set up the dating, then omitting the **; Lhs** part of the **PLOT** command. When you omit the **; Lhs** part of the command, it is assumed that this is a time series plot, and the adjacent points are automatically connected. The figure is also automatically labeled with the dates. Figure E6.8 is a time series plot of the three macroeconomic price series for the data above. Note the use of the **; Grid** specification to improve the readability of the figure. (See Section E6.3.6 for details on this specification.)

NOTE: If your data are not dated using **DATES** and **PERIOD**, i.e., they are undated, then if you omit the **; Lhs = variable** in the **PLOT** command, the observations are plotted against the observation number, beginning with Observ.# = 1.

The commands used for Figure E6.8 are

```
DATES       ; 1953 $
PERIOD      ; 1953-2004$
PLOT        ; Rhs = pnc,puc,ppt
            ; Title = Transportation Price Indices
            ; Grid
            ; Vaxis = Prices of New and Cars and Public Transport $
```

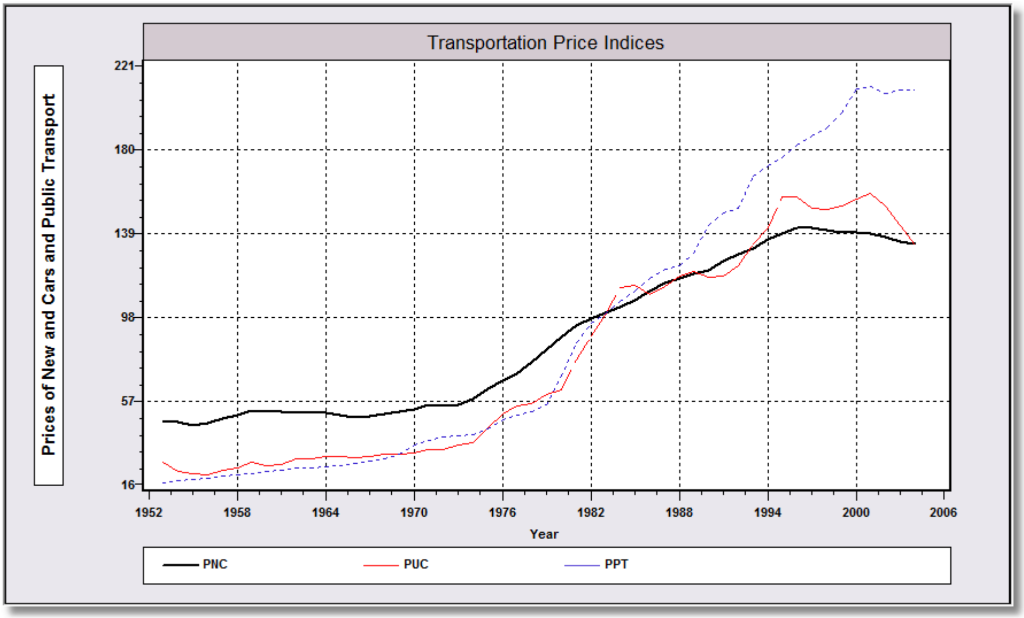


Figure E6.8 Time Series Plot for Several Variables

E6.3.4 Plotting Several Variables Against One Variable

To plot several variables against a single one, just include more than one Rhs variable in the command. The command is

PLOT ; Lhs = variable on horizontal axis
; Rhs = up to five variables to be plotted
; … other options, such as ; Grid and ; Fill $

(Note that the command builder dialog box allows you to specify multiple variables.) A different line style is used for each variable if you use **; Fill**. (The time series plot above is an example in which the Lhs variable is the automatically supplied date.) A different type of point is constructed for each if you are using a cross section. Generally, **PLOT** with **; Fill** (see the next section) creates a figure with one or more line plots, joining segments at the points, but suppressing any symbols for the points. The symbols (dots, stars, etc.) may be retained with **; Symbols**. The **; Regression** command is ignored if more than one variable is being plotted. An example is shown in Figure E6.9:

PLOT ; Rhs = pn,pd,ps
; Lhs = year
; Title = Scatter Plot of Price Series
; Yaxis = Prices
; Grid
; Fill
; Symbols $

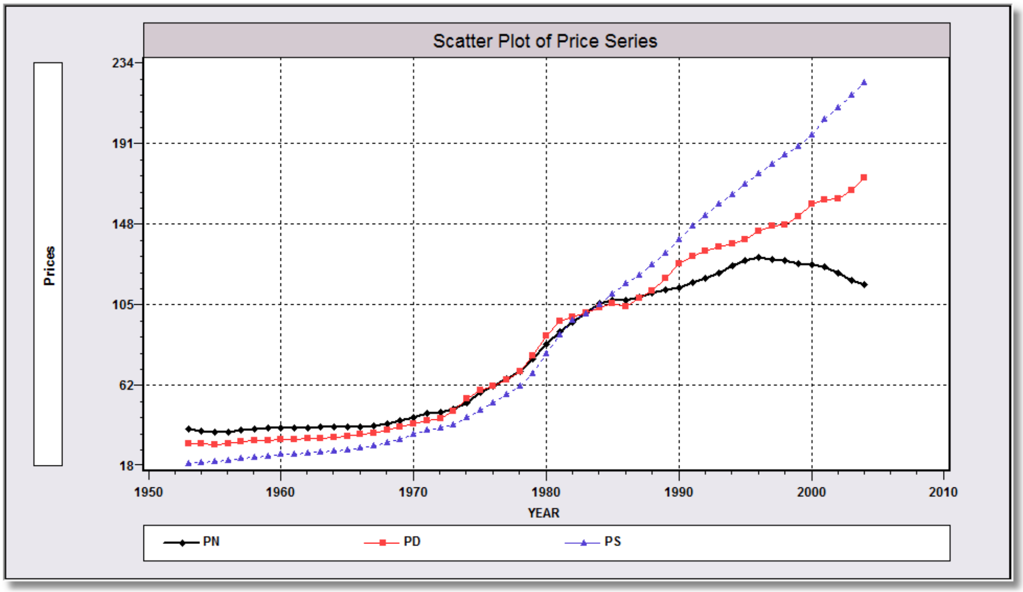


Figure E6.9 Multiple Plots in the Same Figure

NOTE: This is not a time series plot, in spite of the fact that *year* is the variable on the horizontal axis. Although at this point, *LIMDEP* does know that these are time series data, it does not know that '*year*' is a date variable; *year* is just another variable in the data set. If you omit the **; Lhs = variable** specification in the command, *LIMDEP* will label the *x* axis 'YEAR,' but this is not with respect to a variable in your data set; it is the date labeling that you gave in your **DATES** command. To see this at work, note that even if you did not have a variable named *year* in your data set, you could obtain a time series style plot with yearly observations, and labeled as such.

It might be useful in a figure to differentiate between certain variables by creating a line plot for some while plotting only the symbols for others. Plotting fitted and actual values in a regression, as in Figure E6.7, would be a common application. If the function being plotted is not a linear regression, or is some other function of a variable, you can create a scatter plot for some variables and a line plot for others by using

```
PLOT        ; Lhs = variable for horizontal axis
            ; Rh1 = variables to be shown with symbols only
            ; Rh2 = variables to be shown with a line plot
            ; … any other options $
```

In the example shown in Figure E6.10, we have computed a semilog regression, then plotted the predicted and actual values for the retransformed data.

```
REGRESS     ; Lhs = Log(hhninc) ; Rhs = one,educ ; Keep = loginc_f $
CREATE      ; inc_f = Exp(loginc_f) $
PLOT        ; Lhs = educ
            ; Rh1 = hhninc
            ; Rh2 = inc_f
            ; Grid
            ; Title = Predicted and Actual Income vs. Education
            ; Yaxis = Income $
```

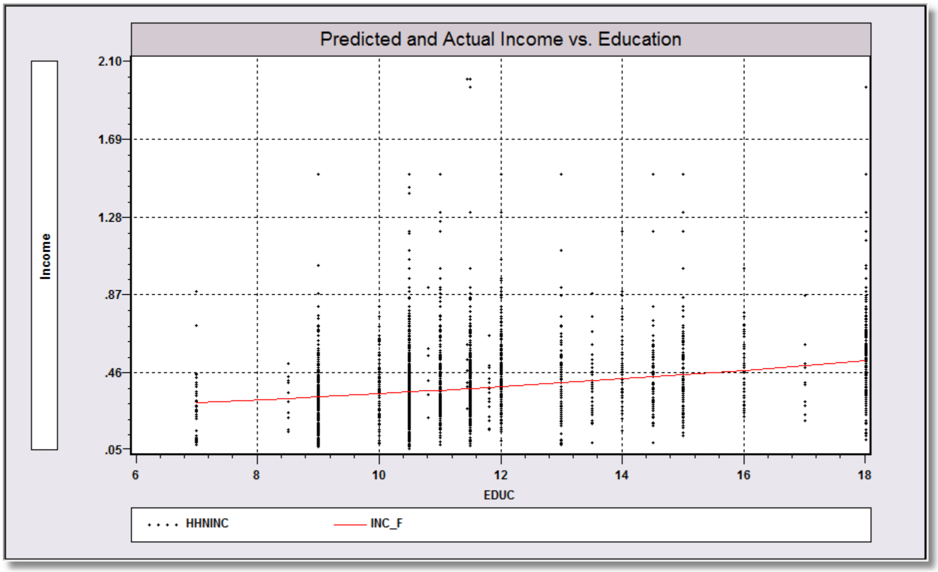


Figure E6.10 Simultaneous Scatter and Line Plots

E6.3.5 Combining Plots

You can combine up to five plots in the same field by using the command

```
PLOT        ; Lhs = variablex1, variablex2, …
            ; Rhs = variabley1, variabley2, …
            ; … any other options $
```

In this specification, each *y* variable is plotted against the corresponding *x* variable, all in the same plotting field. The field is dimensioned to be large enough to accommodate all of the graphs in the single figure. (Note that if your variables do not have similar scales, this can create a very unattractive figure.) Figure E6.11 plots one of the aggregate price indexes against another and one of the micro- indices against another in the same figure.

```
PLOT        ; Rhs = pn,ppt
            ; Lhs = pd,puc
            ; Fill
            ; Title = Price Indexes
            ; Vaxis = PN vs. PD and PPT vs. PUC $
```

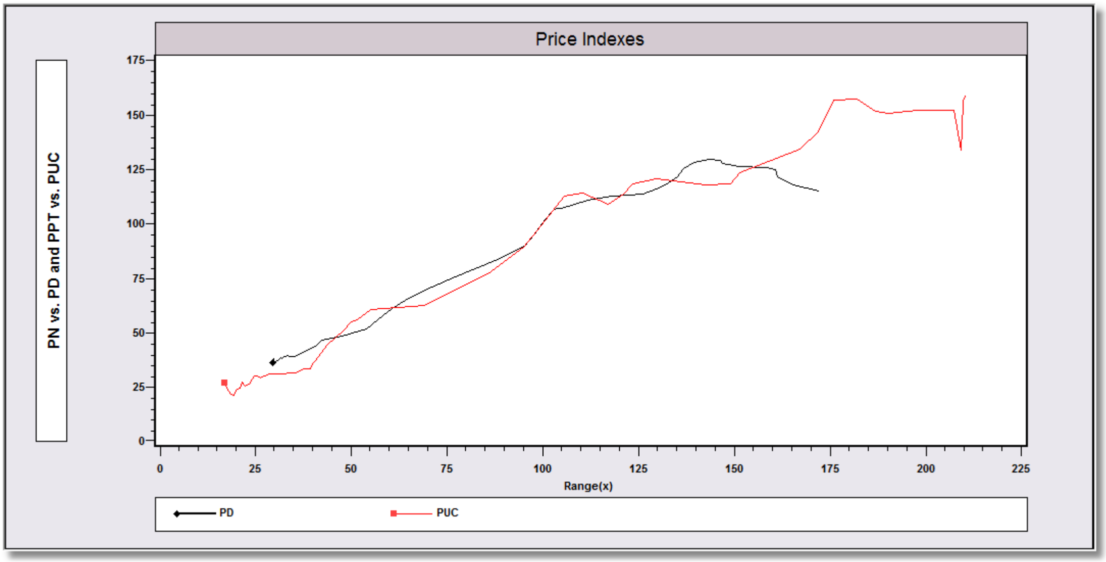


Figure E6.11 Simultaneous Plots

E6.3.6 Options for Scaling and Labeling the Figure

Scaling

The limits for the vertical and horizontal axes are chosen automatically so that every point appears in the figure. Boundaries are set by the ranges of the variables. You can override these settings as follows:

To set the limits for the horizontal axis use **; Endpoints = lower value, upper value**
To set the limits for the vertical axis use **; Limits = lower value, upper value**

HINT: If you plot variables of very different magnitudes in the same figure, or if your series has outliers in it, the scaling convention that seeks to include every point in the graph may severely distort your figure.

NOTE: If the endpoints or limits that you specify push any points out of the figure – *x* or *y* values are outside the limits – then the specifications are ignored, and the original default values are used.

For example, the top panel in Figure E6.12 is the same as Figure E6.6, produced by the command below without the specification of the endpoints and limits. The lower panel shows the effect of expanding the limits

```
PLOT        ; Rhs = gasp
            ; Lhs = g
            ; Title = Gasoline Consumption vs. Price
            ; Yaxis = Gas_Cons
            ; Grid
            ; Limits = 0,125       ?  Set the vertical axis limits
            ; Endpoints = 0,1.2  $  Set the horizontal axis limits
```

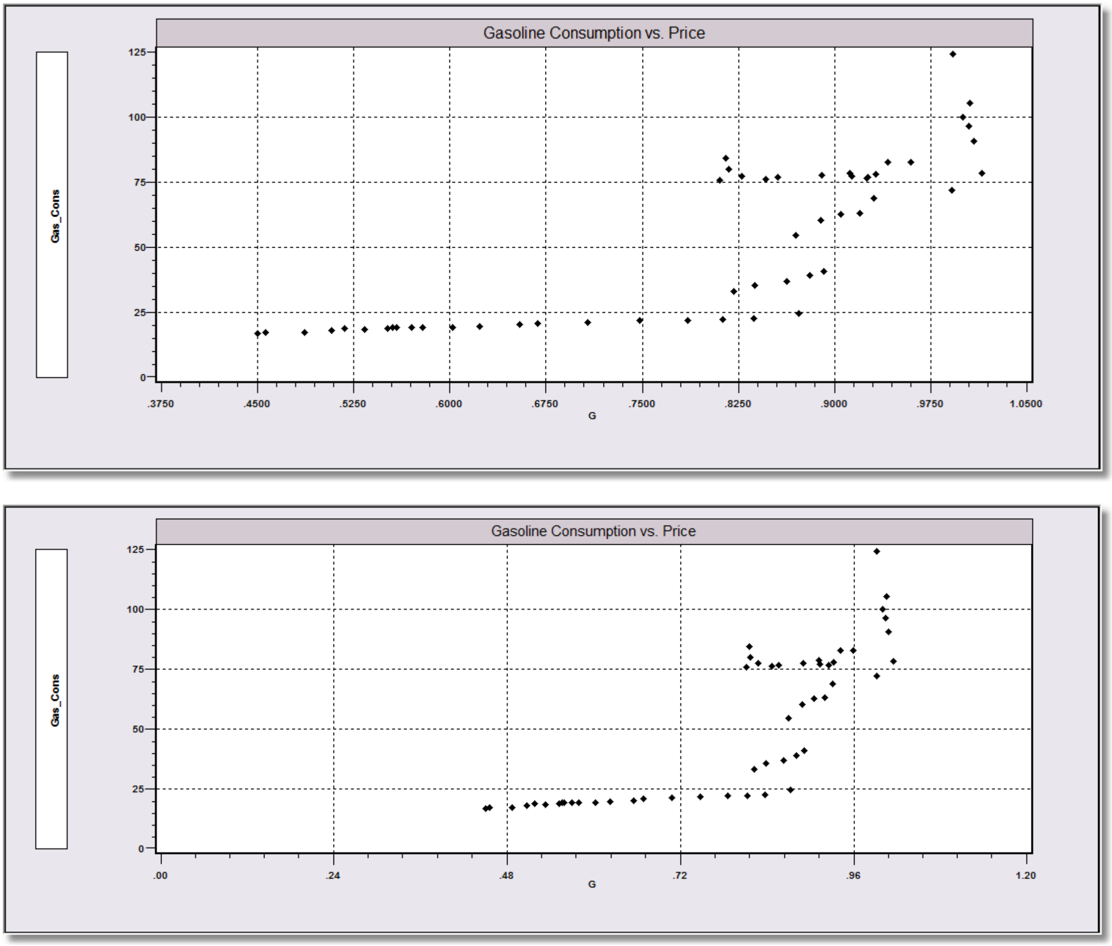


Figure E6.12 Scatter Plots with Rescaled Axes

The following describe some devices for changing the appearance of the figure, and creating particular types of graphs. Some of these have been used in the examples above. More extensive applications appear below.

Grids and Lines in the Plotting Field

It is sometimes helpful when plotting to put a grid in the figure. This makes it easier to relate the points in the graph to the distances on the axes. You may request a grid to be placed in the figure with

; Grid

This divides the screen into a grid of rectangles using dotted bars. The option was used in several of the preceding examples. You may also put horizontal and/or vertical lines in the figure at specific numerical benchmarks. The syntax is

; Spikes = up to five value(s) to put vertical lines at particular values
; Bars = up to five value(s) to put horizontal lines at particular values

The vertical or horizontal line is drawn from axis to axis, the full width or height of the box. Figure E6.13 uses these devices to focus on the means in a regression.

Connecting Points in the Plotting Field

If you are plotting a function or a time series, it may also be useful to connect adjacent points. To do so, add

; Fill

to the command. One way you might use this device would be to draw a function by creating a set of equally spaced values, then plotting the function of these values, connecting the points to create the continuous function. This device was used above in Figures E6.9 and E6.11.

For an example, the following fairly involved program plots the predicted values from a regression with the upper and lower confidence limits for the forecasts. Lines are placed in the figure at the point of means. The underlying computations are based on the bivariate regression

$$\log g_t = b_1 + b_2 \log pg_t + e_t$$

where g_t is the gasoline consumption variable and pg_t is the corresponding observation on the price variable, and b_1 and b_2 are the least squares constant and regression slope. The prediction and associated confidence limits are

$$\log \hat{g}_t = b_1 + b_2 \log pg_t$$

$$s_{ft} = \text{ forecast standard error (estimated)} = s\sqrt{1+\frac{1}{n}+\frac{\left(\log pg_t - \overline{\log pg}\right)^2}{\sum_{t=1}^{T}\left(\log pg_t - \overline{\log pg}\right)^2}}$$

confidence limits = $\log \hat{g}_t \pm t^* \times s_{ft}$

where *s* is the estimated standard error of the regression and *t** is the appropriate critical value from the table of the t distribution. The following does all of these computations and plots the forecast limits for 100 points which span the observed range of *pg*. The end results are shown in Figure E6.13. (The figure lacks the textbook butterfly shape because for these data, the forecast standard error is dominated by the term 1+1/*n*.)

Compute the regression using the sample data and collect statistics

```
SAMPLE      ; 1-52 $
CREATE      ; logg = Log(g)
            ; logpg = Log(gasp) $
REGRESS     ; Lhs = logg
            ; Rhs = one,logpg $
CALC        ; minlpg = Min(logpg)
            ; maxlpg = Max(logpg)
            ; delta =  (maxlpg - minlpg) / 100
            ; meanlpg = Xbr(logpg)
            ; vlpg = (n-1)*Var(logpg)
            ; meanlg = Xbr(logg)
            ; critcalt = Ttb(.95,degfrdm) $
```

Base the remaining computations on 100 generated points.

```
SAMPLE      ; 1-100 $
```

First, create 100 equally spaced points in the range of *pg*.

```
CREATE       ; lpgt = minlpg + delta * Trn(0,1)
```

Obtain forecast standard error and predicted values.

```
            ; fitse = s * Sqr(1 + 1/nreg + (lpgt - meanlpg)^2/vlpg)
            ; fitlg = b(1) + b(2) * lpgt
```

Compute the confidence limits.

```
            ; upper = fitlg + critcalt * fitse
            ; lower  = fitlg  - critcalt * fitse $
```

Now, plot all three series, marking the means of the two variables.

```
PLOT        ; Lhs = lpgt
            ; Rhs = fitlg,upper,lower
            ; Fill
            ; Bars = meanlg
            ; Spikes = meanlpg
            ; Yaxis = Fitted_LG
            ; Title = Forecast Interval for Fitted LogG $
```

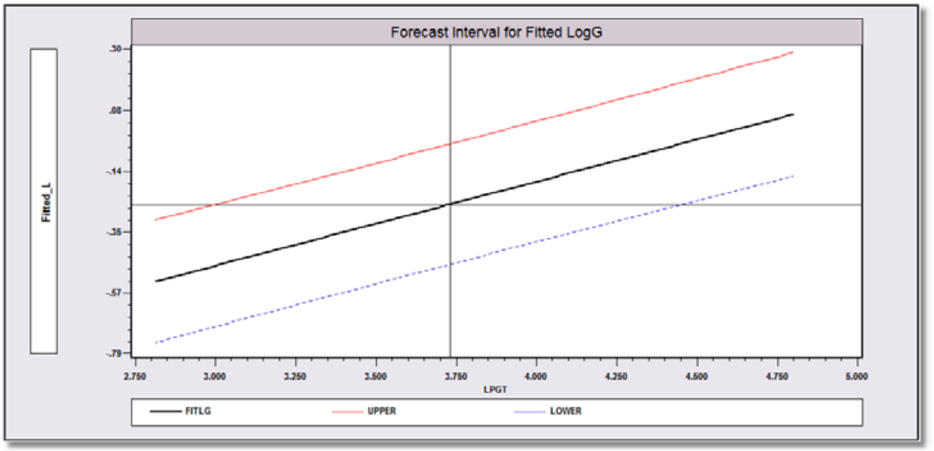


Figure E6.13 Plot of a Forecast Interval

The foregoing could be automated in a procedure. It might lack generality, however, as it is limited to simple bivariate regressions. This result can be replicated with the automatic procedure provided by the **SIMULATE** command. Figure E6.14 produces an interval estimate for the same regression. The commands are

```
SAMPLE      ; 1-52 $
REGRESS     ; Lhs = logg
            ; Rhs = one,logpg $
SIMULATE    ; Scenario: & logpg = minlpg(delta)maxlpg
            ; Plot(ci) $
```

The familiar butterfly effect can be seen in Figure E6.14. Note, the difference between E6.13 and E6.14 is that in E6.13, we have produced a 'forecast interval,' using the textbook formula that includes the variation of the unobserved disturbance. This is the source of the leading 1 under the square root. In Figure E6.14, we have simulated the dependent variable by computing a confidence interval for the fitted values in the regression, not a forecast interval. The interval in Figure E6.14 is correspondingly narrower because the leading 1 in the computation of the standard error is not included.

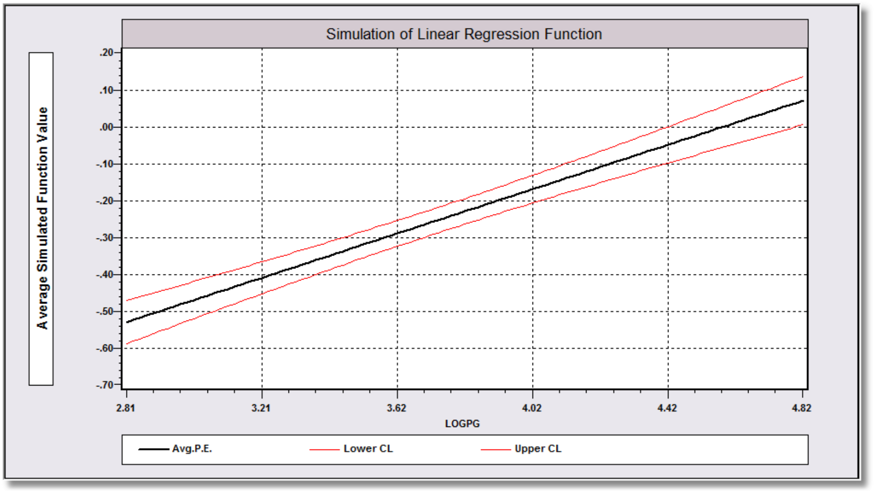


Figure E6.14 Prediction Interval Using SIMULATE

E6.3.7 Fenceposts Plot

A plot that displays vertical distances from the horizontal axis to the point in the field (it will resemble a row of fenceposts) is obtained by adding

; Post

to the **PLOT** command. This form of the figure may take a couple experiments to obtain the desired result, as it is necessary to adjust the location of the horizontal axis. The symbol (dot) that would normally appear in the figure absent the post can be included at the top (or bottom) of the post by adding

; Symbol

to the command.

The following example illustrates. The data used in the plot are a panel of data on 48 states for 17 years used in Munell (1990). The data were downloaded from the website for Badi Baltagi's text: http://www.wiley.com/legacy/wileychi/baltagi/supp/PRODUC.prn (*Econometric Analysis of Panel Data* (2005)). Figure E6.15 displays the within group (state) residual variances based on a loglinear regression of log of public capital on a constant, the log of gross state product and the log of total employment.

```
SAMPLE      ; 1-816 $
REGRESS     ; Lhs = Log(p_cap) ; Rhs = one,Log(gsp),Log(emp)
            ; Res = e $
CREATE      ; Esq = e^2 $
CREATE      ; vstate = Group Mean(esq, Str = state) $
REJECT      ; YR < 1986 $ (Use only last year of the data)
PLOT        ; Lhs = state ; Rhs = vstate
            ; Post
            ; Endpoints = 0,50 ; Limits = 0,.2 ; Grid
            ; Title = Residual Variance by State $
```

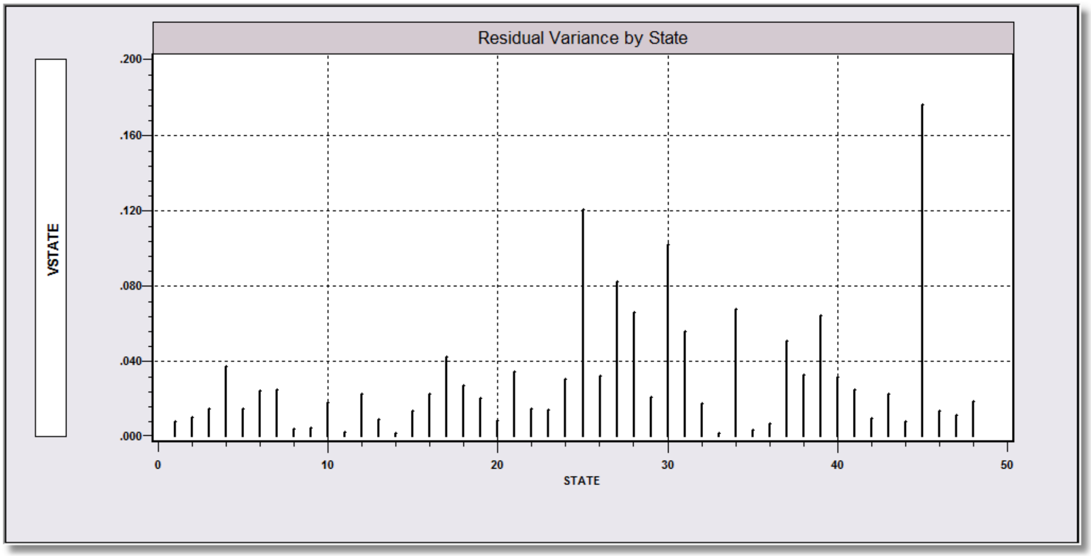


Figure E6.15 Fencepost Plot of Residual Variances by State

E6.3.8 Centipede Plot

When you send an Lhs variable and two Rhs variables to **PLOT**, you can produce a plot which puts the two values above the corresponding Lhs value and draws a vertical line between them with a dot in the center. This is a particular type of plot that we label a centipede plot, because of its appearance which will be clear shortly. The figure is requested with

```
PLOT        ; Lhs = one variable
            ; Rhs = two variables
            ; Centipede
            ... (all other options for plot are the same) $
```

In the following example, we compute a regression for each of the 48 states in the panel data set used in the preceding example. We then construct the 95% confidence limits for a confidence interval for the 48 estimated coefficients on the third variable. The centipede plot shows how the confidence limits vary from state to state. The horizontal bar at zero reveals which of the estimates are significantly different from zero – that is, those for which the confidence interval does not contain zero. To make this convenient, we use a procedure to do the computations in a loop.

```
SAMPLE      ; 1-816 $
MATRIX      ; bi = Init(48,1,0) ; si = Init(48,1,0)$
PROCEDURE
INCLUDE     ; New ; State = i $
REGRESS     ; Lhs = Log(p_cap)
            ; Rhs = one,Log(gsp),Log(emp)
            ; Quietly $
MATRIX      ; bi(i) = b(3) $
CALC        ; sdi = Varb(3,3) ; sdi = Sqr(Sdi) $
MATRIX      ; si(i) = sdi $
ENDPROC $
EXECUTE     ; i = 1, 48 $
CALC        ; tstar = Ttb(.975,14)$
MATRIX      ; upper = bi + tstar*si ; lower = bi - tstar*si $
SAMPLE      ; 1-816 $
REJECT      ; Yr < 1986 $
CREATE      ; b_upper = upper(state) ; b_lower = lower(state) $
PLOT        ; Lhs = state
            ; Rhs = b_lower,b_upper
            ; Centipede
            ; Endpoints = 0,48
            ; Bars = 0
            ; Vaxis = Confidence Limits
            ; Title = Confidence Limits for Employment Elasticity by State $
```

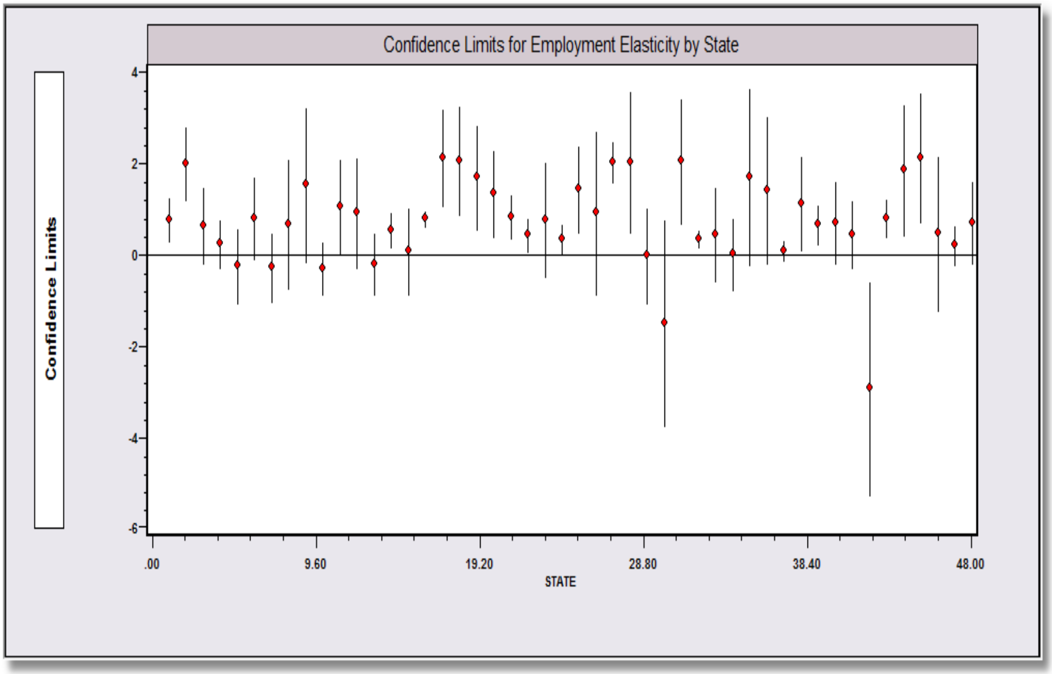


Figure E6.16 Centipede Plot of Confidence Intervals for b(log*employment*)

E6.3.9 A Program for Plotting Confidence Regions

In the example below, which is a general program, a confidence ellipse is drawn for two coefficient estimates. The one dimensional confidence intervals are also drawn. The computation is as follows: The confidence ellipse is the set of points (β_1,β_2) for which

$$F = ½[(b_1 - \beta_1)^2 s_{22}/D + (b_2 - \beta_2)^2 s_{11}/D - 2(b_1 - \beta_1)(b_2 - \beta_2)s_{12}/D] = F^*$$

where F^* is the critical value from the appropriate F table; b_1 and b_2 are the parameter estimates; s_{11}, s_{12}, and s_{22} are estimated asymptotic (co)variances of b_1 and b_2; and $D = s_{11}s_{12}(1-r_{12}^2)$ is the determinant of the 2×2 covariance matrix. The ellipse is defined over values of b_1 and b_2 for which the equality is met. The procedure will do this computation for any model.

The program is used here to produce a confidence region for the price and income coefficients in an equation for the gasoline market examined earlier. The commands used to produce the figure are

```
SAMPLE      ; 1-52 $
CREATE      ; g = gasq/(100*pop/282429) ; logg = Log(g) $
CREATE      ; loginc = Log(pcinc) ; logpg = Log(gasp) $
REGRESS     ; Lhs = logg ; Rhs = one,logpg,loginc $
EXECUTE     ; Proc = confregn(2,3) $
```

Figure E6.17 below shows the results of the computation. The procedure is listed below.

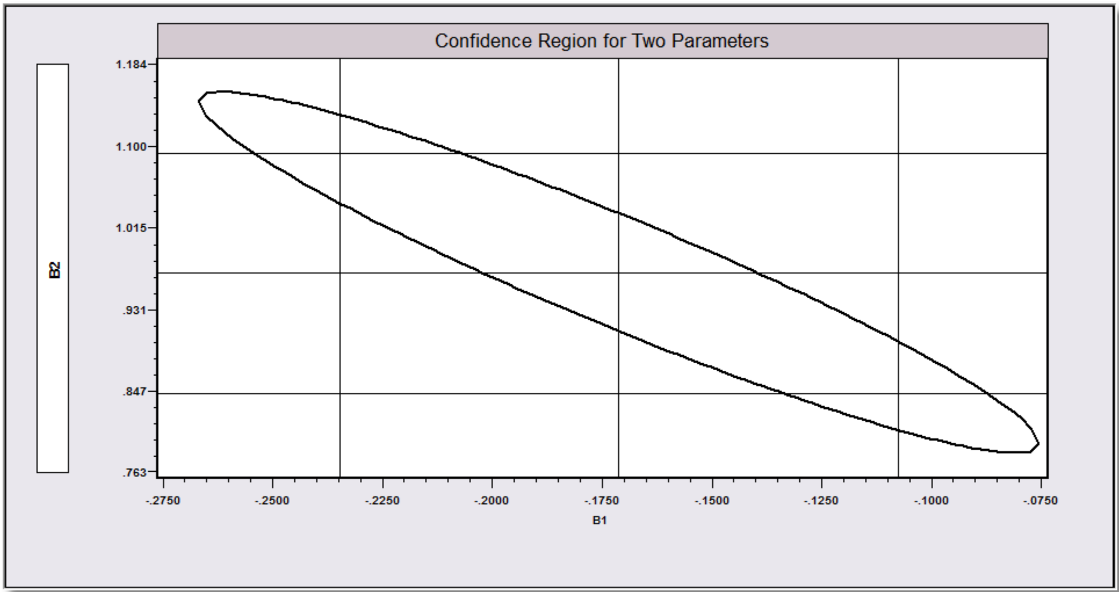


Figure E6.17 Confidence Region for Two Parameters

Compute confidence ellipse and confidence regions for two coefficients. The program does not require the model to be a regression model. It may be used with any model estimated by the program. Theoretically, if the model is not a classical regression model, a large sample appeal to the Wald statistic and asymptotic normality of the estimates is needed. Use the following three steps:

Step 1. Store the procedure.
Step 2. Estimate the model.
Step 3. Execute the following procedure.

No modification is necessary. The plot will contain the confidence ellipse and individual confidence intervals for the selected parameters. The first coefficient is labeled $b1$ and the second is labeled $b2$. The procedure will do the plot for any pair of coefficients.

```
PROC = confregn ( coef1 , coef2 ) $
```

Gather the coefficients, variances, standard errors, t and F values. Do the computations for the intervals. **CALC** obtains the values for the range of $b1$ for the plot.

```
CALC      ; sb1b1 = Varb(coef1,coef1) ; sb1 = Sqr(sb1b1) ; c1 = b(coef1)
          ; sb2b2 = Varb(coef2,coef2) ; sb2 = Sqr(sb2b2) ; c2 = b(coef2)
          ; sb1b2 = Varb(coef1,coef2) ; q12 = sb1b2/sb1b1
          ; u  = Sqr(sb2b2 - sb1b2*q12)
          ; fc = Ftb(.95,2,(n-kreg))
          ; tc = Ttb(.95,(n-kreg))
          ; max = Sqr(2*fc*sb1b1)
          ; min  = -max
          ; db1  = (max - min)/100 $
SAMPLE    ; 1-201 $
```

Plot 201 points in an ellipse. The first 100 are the lower part, the second 100 are the upper part. Point 201 equals point 1, so the ellipse is closed. Compute *b*1-*beta*1 then *b*2-*beta*2 as a function of *b*1-*beta*1, then *b*1 and *b*2.

```
CREATE      ; If(_obsno <= 100) b1 = min+Trn(1,1)       * db1
            ; If(_obsno >   100) b1 = max-(Trn(1,1)-100)* db1
```

The first line contains a protection against taking the square root of a negative number.

```
            ; q = u*Sqr(0 ! (2*fc - b1*b1/sb1b1))
            ; If(_obsno <= 100) b2 = c2 + b1 * q12 + q
            ; If(_obsno >   100) b2 = c2 + b1 * q12  -  q
            ; b1 =  b1 + c1
            ; If(_obsno = 201) b1 = b1[-200]
            ; If(_obsno = 201) b2 = b2[-200] $
```

These are the one dimensional upper and lower confidence bounds for the two coefficients. Top, bot, etc. make the box a little bigger.

```
CALC        ; ucb2 = c2 + tc*sb2 ; lcb2 = c2 - tc*sb2
            ; ucb1 = c1 + tc*sb1 ; lcb1 = c1 - tc*sb1
            ; top = 1.025 * Max(b2) ; bot = .975 * Min(b2)
            ; lft = .975 * Min (b1) ; rt = 1.025 * Max(b1) $
```

Finally, plot the ellipse with the confidence limits and a bar and spike to show the original coefficients, themselves.

```
PLOT        ; Lhs = b1 ; Rhs = b2
            ; Bars = ucb2,lcb2,c2 ; Spikes = ucb1,lcb1,c1
            ; Limits = bot,top ; Endpoints = lft,rt ; Fill ; Nosort
            ; Title = Confidence Region for Two Parameters $
ENDPROC
```

The preceding program is written specifically for a linear regression model with normally distributed disturbances – it is based on the F statistic for the joint test of the significance of the two coefficients. It can be made more general by setting it up for the Wald (chi squared) statistic that would rely on the asymptotic distribution. The following lists the corresponding program, without the surrounding annotation. It is executed the same as in the earlier program.

```
PROC  = confregn ( coef1 , coef2 ) $
CALC       ; sb1b1 = Varb(coef1,coef1) ; sb1 = Sqr(sb1b1) ; c1 = b(coef1)
           ; sb2b2 = Varb(coef2,coef2) ; sb2 = Sqr(sb2b2) ; c2 = b(coef2)
           ; sb1b2 = Varb(coef1,coef2) ; q12 = sb1b2/sb1b1
           ; u = Sqr(sb2b2 - sb1b2*q12)
           ; max = Sqr(5.99*sb1b1) ; min = -max ; db1 = (max - min)/100 $
SAMPLE     ; 1-201 $
CREATE     ; If(_obsno <= 100) b1 = min + Trn(1,1)        * db1
           ; If(_obsno >   100) b1 = max - (Trn(1,1)-100) * db1
           ; q = u*Sqr( 0 ! (5.99 - b1*b1/sb1b1))
           ; If(_obsno <= 100) b2= c2 + b1 * q12 + q
           ; If(_obsno >   100) b2= c2 + b1 * q12 -  q
           ; b1 = b1 + c1
           ; If(_obsno = 201) b1 = b1[-200]
           ; If(_obsno = 201) b2 = b2[-200] $
CALC       ; ucb2 = c2 + 1.96*sb2 ; lcb2 = c2 - 1.96*sb2
           ; ucb1 = c1 + 1.96*sb1 ; lcb1 = c1 - 1.96*sb1
           ; top  = Max(b2) + .1*Abs(Max(b2))
           ; bot  = Min (b2)  - .1*Abs(Min(b2))
           ; lft   = Min (b1)  - .1*Abs(Min(b1))
           ; rt    = Max(b1) + .1*Abs(Max(b1)) $
PLOT       ; Lhs = b1 ; Rhs = b2
           ; Bars = ucb2,lcb2,c2 ; Spikes = ucb1,lcb1,c1
           ; Limits = bot,top ; Endpoints = lft,rt ; Fill ; Nosort
           ; Title = Confidence Region for Two Parameters$
ENDPROC
```

E6.3.10 Sorting the Data Before Plotting

The **PLOT** command contains a switch **; Nosort** which is used to request the plotting program not to sort the values of the variable on the horizontal axis (carrying all those for the vertical axis, of course) before plotting. The sort is normally done when you use **; Fill** to connect points so that the plot looks like a function and not a jumble of lines. When you plot a time series, this sort is not necessary since dates are already sorted by construction. But, it might be necessary if you are plotting a cross section of values. The following experiment demonstrates: The data on *x* are randomly drawn from the standard normal distribution, but *y* is a simple, deterministic quadratic function of *x*. The results are shown in Figure E6.18. The plot on the left shows the expected result. The plot on the right shows what happens if the data are not sorted.

```
SAMPLE     ; 1-100 $
CREATE     ; x = Rnn(0,1) ; y = .3 * (x-1)^2 $
PLOT       ; Lhs = x ; Rhs = y ; Fill $  Then, we add ; NoSort
```

Data sorted before plotting as usual **Data not sorted before plotting**

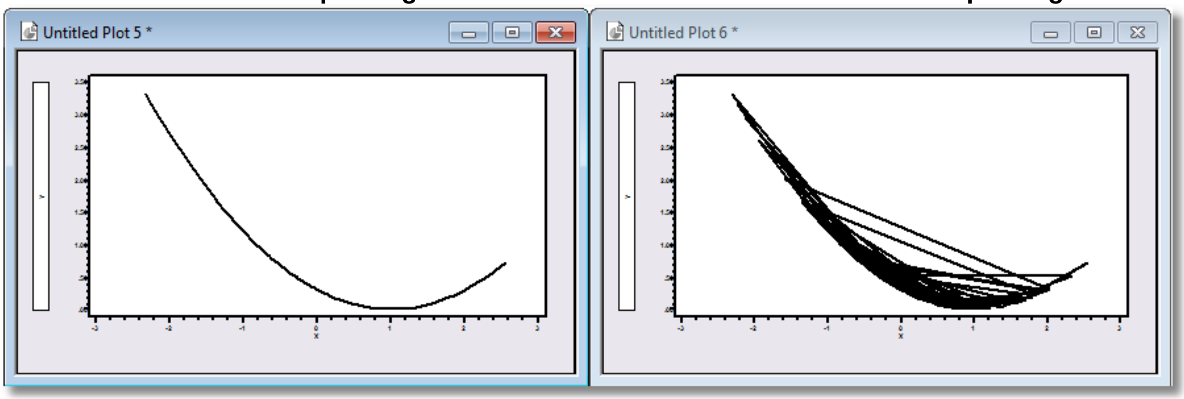


Figure E6.18 The Effect of Sorting the Lhs Variable in PLOT ; Fill ... $

This issue is relevant in our confidence region plot because of the way it is constructed. The data on b_1 are

$$b_1(min), b_1(min)+db1, b_1(min)+2db1 \ldots b_1(max), b_1(max)-db1, b_1(max)-2db1, \ldots, b_1(min)$$

so the series ascends from $b_1(min)$ to $b_1(max)$ in equal steps, then descends from $b_1(max)$ to $b_1(min)$ in equal steps. This allows us to draw a figure that (if it were done slowly enough for you to watch it) would move the pen across the field from left to right (the lower half of the ellipse), then move back across the field from right to left (the upper half of the ellipse). This is necessary because, in fact, the ellipse is not a function; the values the horizontal axis are not each associated with a single *y*. If the confidence ellipse is plotted with the sort, Figure E6.19 results; intriguing, perhaps, but not what we had in mind.

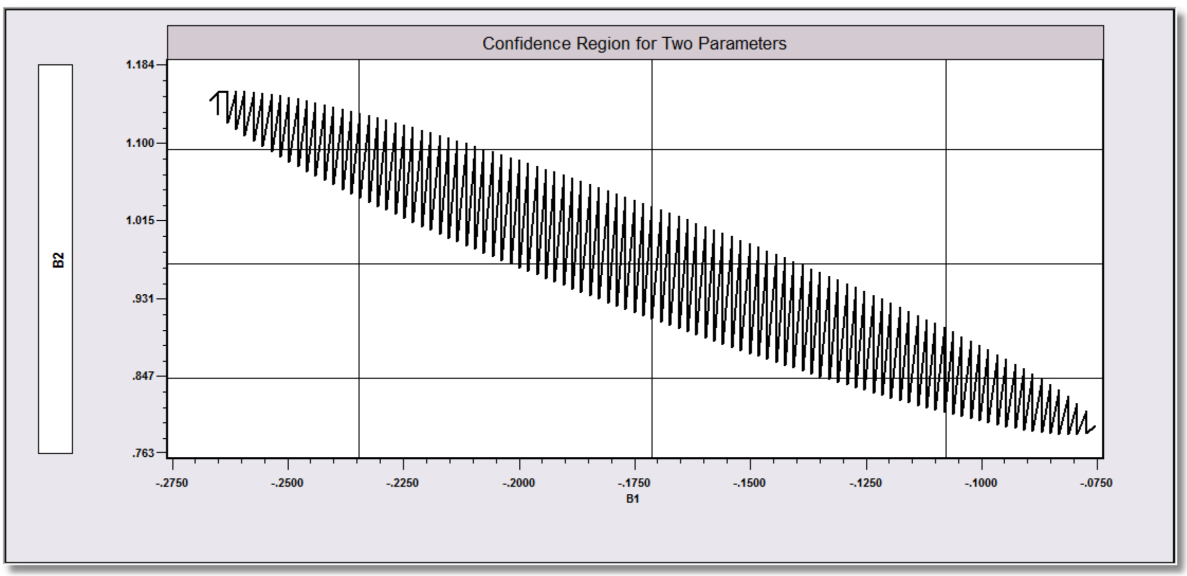


Figure E6.19 Function Plot with Inappropriate Sorting

E6.3.11 Plotting a Function

Frequently, a simple way to plot a function is to plot the function values at a set of equally spaced points and connect the dots in the figure. The following produces Figure B.3 in Greene (2011, page 1023), where the density of the t distribution is plotted for several values of the degrees of freedom parameter: The following plots the density of *t* for 2, 10, 40, and (essentially) infinite degrees of freedom.

```
SAMPLE      ; 1-101 $                Plot and connect 100 segments
CREATE      ; t = Trn(-4,.08) $      Values -4 to +4 in steps of .08
```

This procedure obtains the value of the density over a grid of values contained in variable *t*, and puts them in a variable passed as *fn*.

```
PROC = tdensity(fn,t,d) $
CREATE      ; fn = Gma((d+1)/2)/Gma(d/2) / Sqr(d*pi) * (1+t*t/d)^(-(d+1)/2)$
ENDPROC $
```

Compute for 2, 10, 40, infinity (last is N(0,1)). Then plot the four densities in the same figure.

```
EXECUTE     ; Proc = tdensity(t2,t,2) $
EXECUTE     ; Proc = tdensity(t10,t,10) $
EXECUTE     ; Proc = tdensity(t40,t,40) $
CREATE      ; tinf = N01(t) $
PLOT        ; Lhs = t ; Rhs = t2,t10,t40,tinf ; Fill intervals ; Yaxis = Density
            ; Title = t Densities with Different Degrees of Freedom $
```

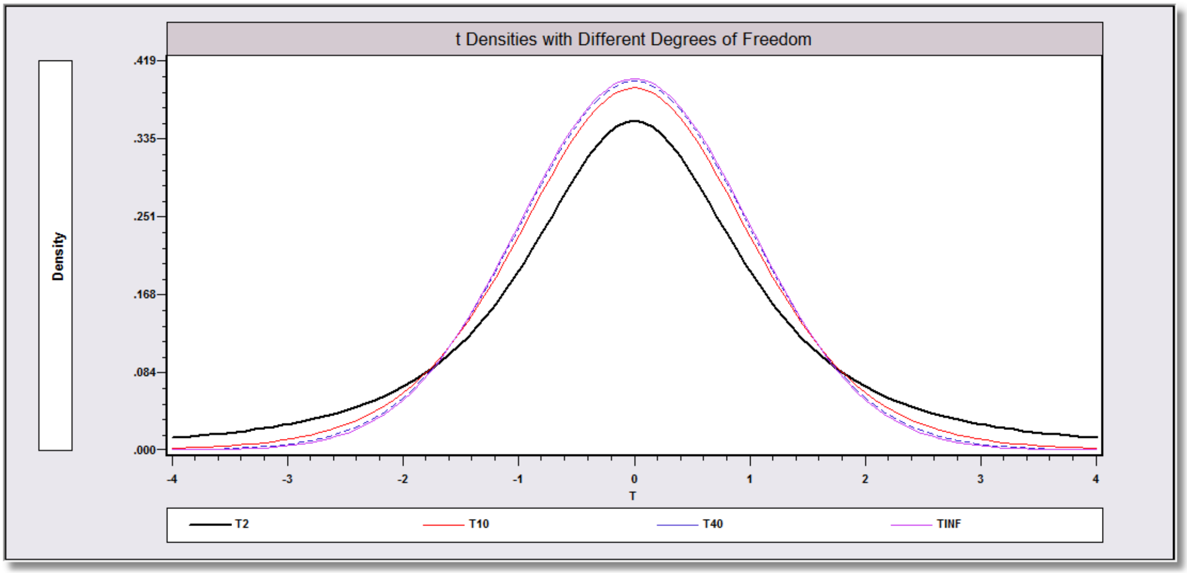


Figure E6.20 Plot of t Density with Varying Degrees of Freedom

An area under the function being plotted can be filled by using

; Area = Left, Right

in the command. This has the effect of filling the figure from the left margin (negative infinity) to *Left* and from *Right* to the right margin (positive infinity). The following familiar figure shows the upper and lower 2.5% critical regions for a t distribution with 40 degrees of freedom.

```
SAMPLE      ; 1-1000 $
CREATE      ; t = Trn(-4,.008) $       Values -4 to +4 in steps of .08
EXECUTE     ; Proc = tdensity(t40,t,40)$
CALC        ; t025 = Ttb(.025,40) ; t975 = -t025$
PLOT        ; Lhs = t ; Rhs = t40
            ; Yaxis = density
            ; Title = Upper and Lower 95% Critical Values for t[40]
            ; Area = t025,t975 $
```

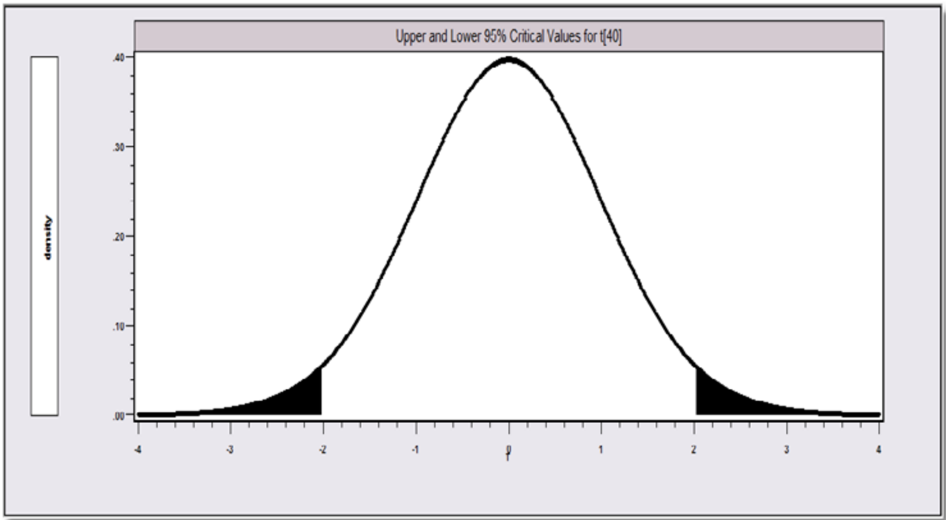


Figure E6.21 Critical Regions for the t[40] Distribution

You can hide the left tail area by making *Left* low enough to be out of the picture, such as -5 in the figure above and the right area by making *Right* large enough to be out of the picture (e.g., +5 above). In order to obtain a complete fill, the number of points plotted should be at least 500. If the area has gaps, they will close when you resize the graph.

A function plot for a discrete variable with a small number of values will appear like the top panel of Figure E6.22 which plots a Poisson probability distribution. Connections between the dots are meaningless (there is no function value for xp = 2.5). Nonetheless, it is customary to accentuate the plot by including the connections. For this sort of figure, use

; Fill ; Symbols

The commands are:

```
SAMPLE      ; 1-10 $
CREATE      ; xp = Trn(0,1)$
CREATE      ; prob_xp = Exp(-2)*2^xp/Gma(xp+1) $
PLOT        ; Lhs = xp ; Rhs = prob_xp
            ; Fill ; Symbols
            ; Title = Poisson Probabilities with Lambda = 2
            ; Grid
            ; Vaxis = Poisson Probability $
```

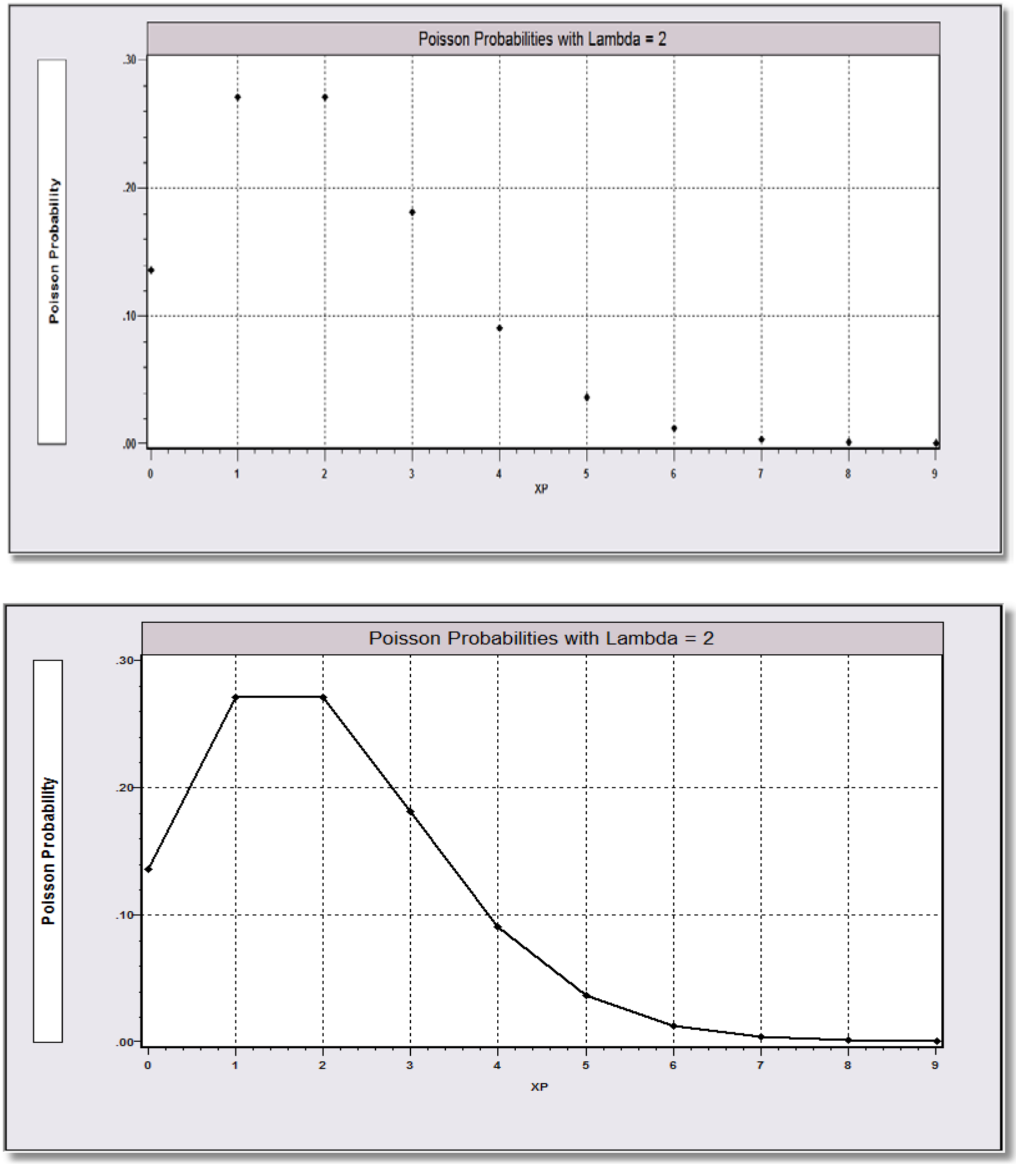


Figure E6.22 Poisson Probabilities

E6.3.12 Stratified Scatter Plots

A scatter plot can be stratified by a stratification variable, so that the values associated with each stratum can be given a different symbol in the figure. You may plot up to five strata in the same figure. The command format is

PLOT ; Lhs = variable on horizontal axis
; Rhs = variable on vertical axis
; Str = stratification variable which takes values 1, 2, ... up to 5
[any other options – all available except ; Fill and ; Regression] $

Figure E6.23 shows a plot of *age* vs. *income* stratified by marital status using the health care system data that we used in Section E3.3 (histograms). In order to reduce the density of the plot, only a small subsample of observations is used

SAMPLE ; 1-500 $
CREATE ; married = married + 1 $
PLOT ; Lhs = age
; Rhs = hhninc
; Str = married
; Title = Income vs. Age for Married and Nonmarried $

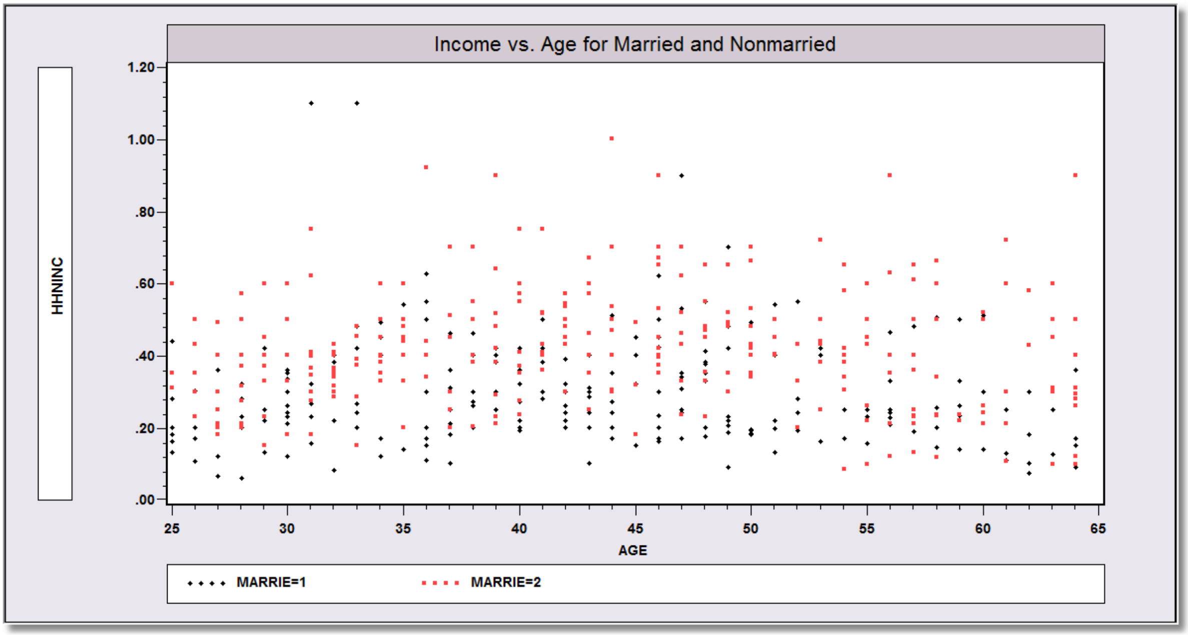


Figure E6.23 Scatter Plot with Stratification

E6.4 Multiple Scatter Plots – The SPLOT Command

The command for computing and plotting several scatter diagrams simultaneously is

SPLOT ; Rhs = list of from three to five variables $

This command requests a simultaneous plot of every variable in the list against every other variable. This can produce up to 4×5 = 20 plots at the same time. (Variables are not plotted against themselves.) The command builder is the same as the first page for **PLOT** and can be accessed by selecting Model:Data Description/Multiple Scatter Plots. There are no options or other specifications for this command. An example in which we obtain simultaneous scatter plots for five of the price indices in the gasoline market data appears in Figure E6.24.

SPLOT ; Rhs = pn,pd,ps,pnc,puc
; Title = Aggregate and Transport Price Indexes $

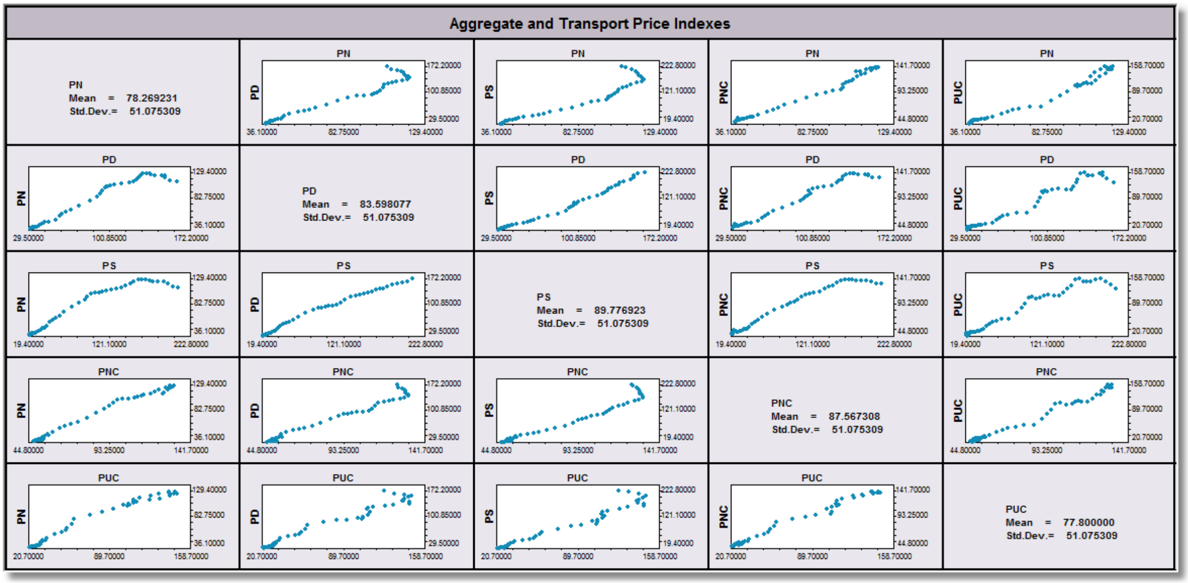


Figure E6.24 Multiple Scatter Plots with SPLOT

E6.5 Plotting Matrices – The MPLOT Command

The command for plotting the values in one matrix against those in another is

MPLOT ; Lhs = matrix 1 ; Rhs = matrix 2 $

The plot is in the same fashion as a pair of variables. All other options are the same as in the **PLOT** command. The graph is produced by treating the corresponding elements of the two matrices as if they were observations on a pair of variables. Only one Rhs matrix may be given. All other options, such as **; Limits**, **; Fill**, etc. are the same as for **PLOT**. We consider two examples below. The command builder for **MPLOT** is the same as for **PLOT.** It can be accessed by selecting Model:Data Description/Plot Matrix.

E6.5.1 Plotting Autocorrelation and Partial Autocorrelation Functions

The **IDENTIFY** command described in Chapter E5 produces character graphic plots of the autocorrelation and partial autocorrelation functions for a variable. It also saves these results as a matrix, *acf_pacf*, so if you want to produce sharper figures, you can use **MPLOT**. The program segment below shows how to obtain such a plot for the ACF and PACF for any time series. (The character plot will actually be more informative since it also lists the test statistics.)

Get the ACF and PACF.

```
IDENTIFY    ; Rhs = the variable ; Pds = the value you choose $
```

(We used *PD* in the gasoline data for our example.) Find out how many lags were used.

```
CALC        ; lags = Row(acf_pacf) $
```

Extract the two columns of the results matrix.

```
MATRIX      ; acf = Part(acf_pacf,1,lags,1,1) ; pacf = Part(acf_pacf,1,lags,2,2) $
```

Use a trick to create a matrix containing 1,2,...

```
SAMPLE      ; 1 - lags $
CREATE      ; t = Trn(1,1) $
MATRIX      ; k = t $
```

Now, plot the figures.

```
MPLOT       ; Lhs  = k ; Rhs = acf,pacf
            ; Limits = -1,1 ; Bars = 0 ; Symbols
            ; Title = ACF and PACF for Price Index for Durable Goods $
```

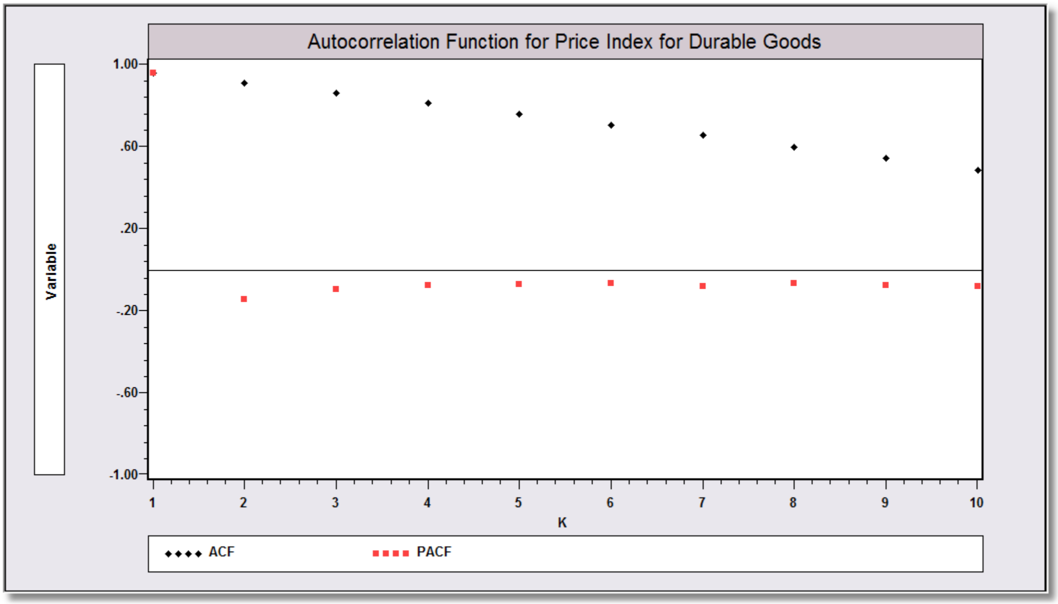


Figure E6.25 ACF and PACF for Durable Goods Price Index

E6.5.2 Examining an Estimation Criterion (Log Likelihood) Function

The next example involves a graphical analysis of a nonlinear least squares problem. The example is taken from Greene (2011). A generalized production function is written

$$\log y + \theta y = \beta_1 + \beta_2 \log k + \beta_3 \log l + \varepsilon.$$

The log likelihood function for this model, concentrated over $\sigma^2 = \mathrm{Var}[\varepsilon]$, is

$$\log L = \Sigma_i [\log(1 + \theta y_i) - \log y_i] - (n/2)[1 + \log(2\pi) + \log(\mathbf{e'e}/n)]$$

where $\mathbf{e'e}$ is the sum of squared residuals in the least squares regression of $\log y+\theta y$ on the Rhs variables. Conditioned on θ, the MLE is OLS, and we scan over θ to find the MLE. The data are set up as follows:

IMPORT $

```
state,valueadd,capital,labor,estabs
Alabama          126.148        3.804      31.551     68
California      3201.486      185.446     452.844   1372
Connecticut      690.670       39.712     124.074    154
Florida           56.296        6.547      19.181    292
Georgia          304.531       11.530      45.534     71
Illinois         723.028       58.987      88.391    275
Indiana          992.169      112.884     148.530    260
Iowa              35.796        2.698       8.017     75
Kansas           494.515       10.360      86.189     76
Kentucky         124.948        5.213      12.000     31
Louisiana         73.328        3.763      15.900    115
Maine             29.467        1.967       6.470     81
Maryland         415.262       17.546      69.342    129
Massachusetts    241.530       15.347      39.416    172
Michigan        4079.554      435.105     490.384    568
Missouri         652.085       32.840      84.831    125
New_Jersey       667.113       33.292      83.033    247
New_York         940.430       72.974     190.094    461
Ohio            1611.899      157.978     259.916    363
Pennsylvania     617.579       34.324      98.152    233
Texas            527.413       22.736     109.728    308
Virginia         174.394        7.173      31.301     85
Washington       636.948       30.807      87.963    179
West_Virginia     22.700        1.543       4.063     15
Wisconsin        349.711       22.001      52.818    142
```

CREATE ; y = valueadd/estabs
; logy = Log(y)
; logcaptl = Log(capital/estabs)
; loglabor = Log(labor/estabs) $

The commands for producing trace of the log likelihood in Figure E6.26 are as follows:

```
CALC          ; i = 0$
MATRIX        ; ti = Init(26,1,0) ; li = ti $
PROCEDURE
CREATE        ; d = logy + t*y ; jacobian = Log(1+t*y) - logy $
REGRESS       ; Quietly ; Lhs = d ; Rhs = one, logcaptl, loglabor $
CALC          ; loglik = Sum(jacobian) - n/2*(1+Log(2*pi)+Log(sumsqdev/n)) $
MATRIX        ; {i = i+1} ; ti(i) = t ; li(i) = loglik $
ENDPROCEDURE
EXEC          ; Silent ; t = 0,1,.04 $
MPLOT         ; Lhs = ti ; Rhs = li ; Fill ; Grid
              ; Endpoints = 0,1
              ; Vaxis = Log Likelihood
              ; Title  = Log Likelihood for a Generalized Production Model $
```

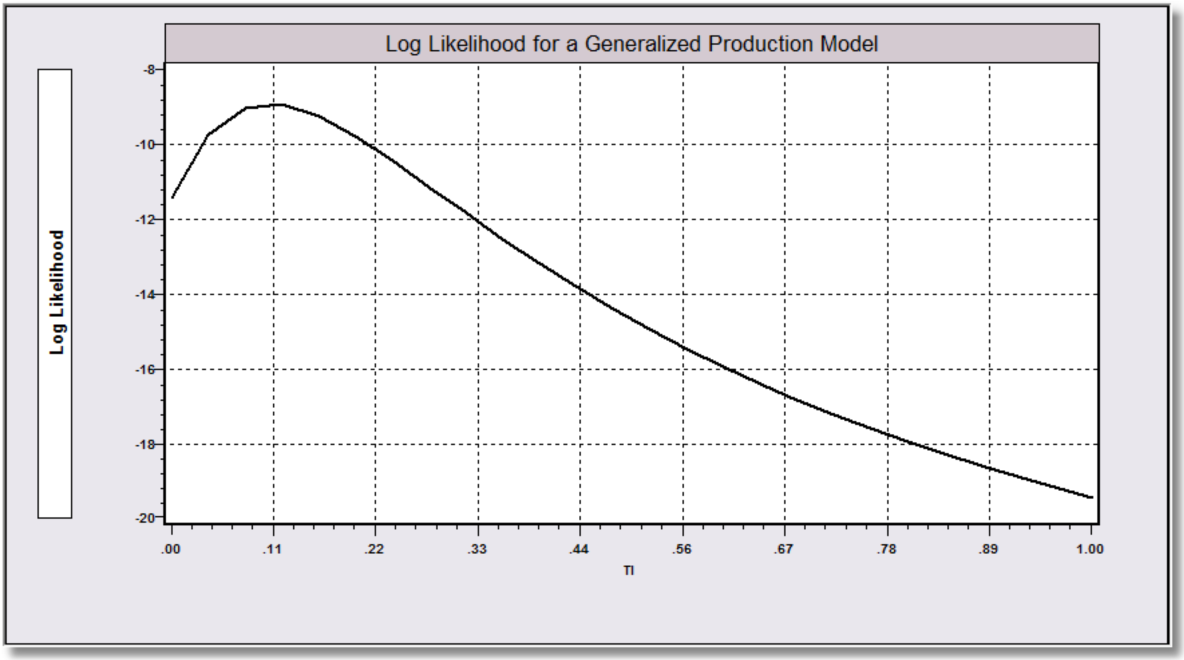


Figure E6.26 Matrix Plot of a Log Likelihood Function

E6.6 Plotting Functions – The FPLOT Command

The command for plotting a function of one variable is

FPLOT ; Fcn = function definition
; ... several other mandatory specifications $

The full, general form of the command is

SAMPLE ; 1 $
FPLOT ; Fcn = the function definition of f(x| any other variables)
; Labels = x ... any other variables
; Plot (x)
; Start = an interior point in the range of x
; Pts = number of points to plot
; Limits = lower,upper limit of range of x over which to plot
; Endpoints = lower,upper limits for horizontal axis (optional) $

There are no other options for this plotting function. The command has this structure because it is also used with the **MINIMIZE** command, described in Chapter E66. The function definition may use any of the features described in Section E14.3 and Chapter E66 for functions for the **MINIMIZE** command. The purpose of the **; Start = value** part of the command is for you to provide a point at which *LIMDEP* can test the function definition that you have given to see if it is computable. If it is, processing continues. The function may involve any number of parameters specified by the **; Labels** specification. The **; Plot(x)** must specify one of the variables in the **; Labels** list.

The **SAMPLE ; 1 $** command is used to prevent needless computing. This command can be used to plot a function which is a sum of terms where the sum is taken over the current sample. But, for a simple function such as the one examined in the example below, this summing operation would just compute and add the same function *n* times.

You may specify the desired limits on vertical axis with third and fourth values in the **; Limits** specification. The first and second give the range of variation on the horizontal axis for the variable being plotted. They do not control the limits on the actual graph plotted. Plot limits for the horizontal axis (only to control the display) are specified with **; Endpoints = lower,upper**. These must widen the interval specified by **; Limits = lower,upper**. They may not narrow it at either end. Thus, for example, the command containing **; Limits = 0,1 ; Endpoints = -1,2** evaluates the function for the variable ranging from zero to one, then constructs a graph in which the horizontal axis contains a range from minus one to two. There will be blank space in the graph from minus one to zero on the left of the plotted function, and from one to two on the right.

The command builder dialog box for the **FPLOT** command is found in the Model: Numerical Analysis/Plot Function option. An illustration appears with the example below.

The following is the example in Section E4.3 of Greene (2011, page 1108), maximizing over ρ and β the function

$$F(\rho,\beta) = \rho\log\beta - \log\Gamma(\rho) - 3\beta + \rho - 1.$$

At the maximum, we must have $\partial F/\partial\beta = \rho/\beta - 3 = 0$ which implies that at the solution, $\beta = \rho/3$. Considering only these points, then, the concentrated function to be maximized is

$$F^*(\rho) = \rho\log(\rho/3) - \log\Gamma(\rho) - 1.$$

We do this graphically with

```
SAMPLE    ; 1 $
FPLOT     ; Fcn = r*Log(r/3) - Lgm(r) - 1
          ; Labels = r
          ; Plot(r) ; Start = 5
          ; Pts = 100 ; Limits = 1,16 $
```

The command builder that will produce this command is shown in Figure E6.27.

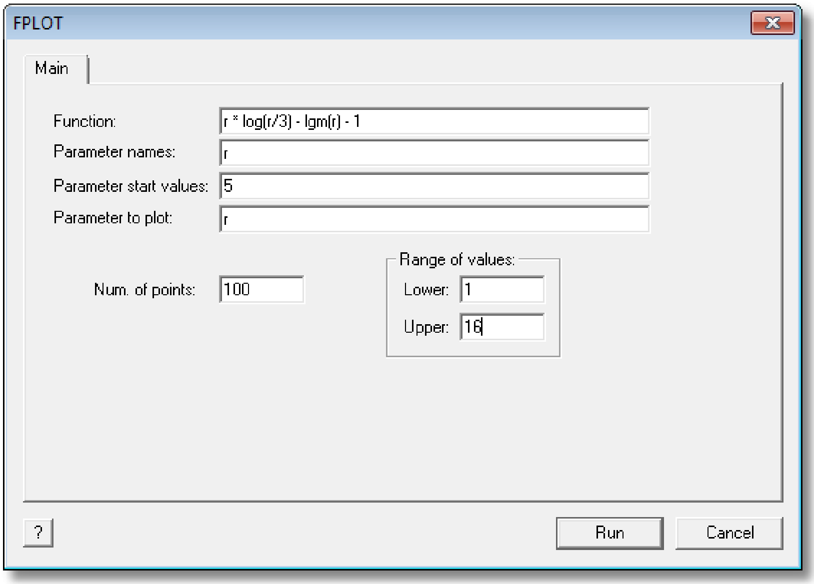


Figure E6.27 Command Builder for FPLOT

The results in Figure E6.28 reveal that the maximum is near 5.2. (The correct value is 5.23.)

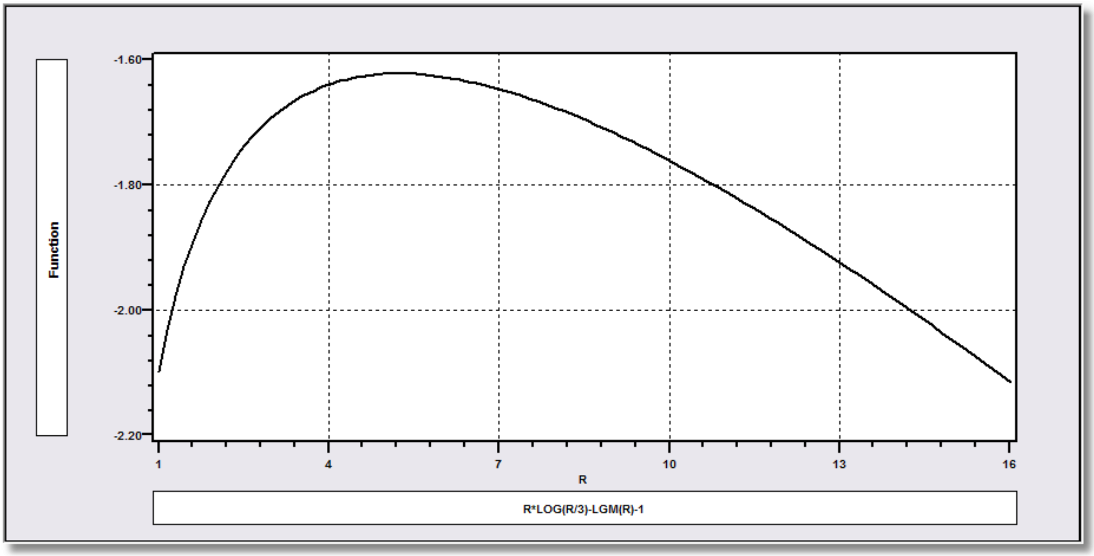


Figure E6.28 Function Plot of a Concentrated Log Likelihood

E6.7 Contour Plots – The CPLOT Command

CPLOT may be used to plot contours of a function in two dimensions. The feature may be used to plot contours of a function

$$f(x1,x2 \mid x3,\ldots) = c$$

for values of c. We have used it in the example below to plot the contours of a log likelihood function. For a probit model, the command format is

```
CPLOT       ; Fcn = any function of at least two arguments
            ; Labels = the labels in the function
            ; Plot(x,y) = two of the variables identified by the labels
            ; Pts = number of evaluation points
            ; Start = the usual values
            ; Limits = xlow,xhigh,ylow,yhigh
            ; Title = desired title $
```

The function can be a log likelihood or any other function. In Figure E6.29, we have plotted the function that was examined in the previous section,

$$F(\rho,\beta) = \rho\log\beta - \log\Gamma(\rho) - 3\beta + \rho - 1.$$

The commands are

```
SAMPLE      ; 1 $
CPLOT       ; Fcn = ro*Log(beta) - Lgm(ro) - 3*beta + ro - 1
            ; Labels = beta,ro
            ; Plot(beta,ro)
            ; Start = 2,5
            ; Limits = .1,5,1,16
            ; Pts = 150
            ; Title = Contour Plot of Log Likelihood $
```

This procedure does a very large amount of computation and the plot takes a while to draw. The function is computed Pts^2 times to produce the grid of values. Then, the plot itself is quite involved. The first example below involves a simple function of two parameters, but no data, so the sample is set to one observation at the outset. Figure E6.29 takes about ten seconds to produce. The second example below is computed for the log likelihood of a probit model in a sample of about 4,000 observations and six parameters. This procedure takes several minutes to complete. Note, finally, contour plots are difficult to plot when the function has more than one model. The implementation here will not perform well if the function has hills and valleys. It is constructed to handle a unimodal function such as the straightforward log likelihood functions shown in the examples.

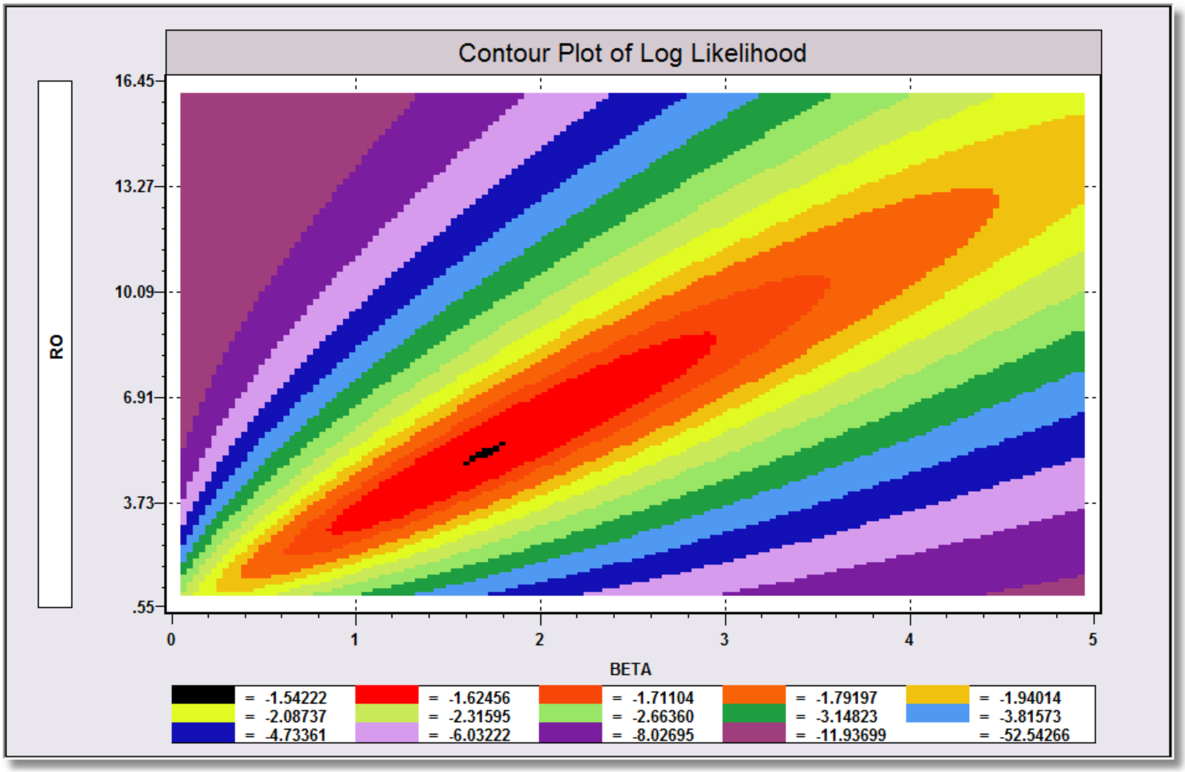


Figure E6.29 Contour Plots

The computations for a probit model are as follows: The estimator maximizes the log likelihood over six parameters. The contour plot examines the log likelihood function in the space of the coefficients on female and educ, which for convenience have been made the first and second parameters in the function.

```
INCLUDE     ; New ; year = 1991 $
PROBIT      ; Lhs = doctor
            ; Rhs = female,educ,one,age,married,hhninc $
CPLOT       ; Fcn = Log(Phi((2*doctor-1)*
                  (b1*female+b2*educ+b3+b4*age+b5*married+b6*hhninc)))
            ; Start = b
            ; Limits = .3,.5,-.03,.01
            ; Pts = 25
            ; Plot(b1,b2)
            ; Labels = b1,b2,b3,b4,b5,b6
            ; Title = Log Likelihood for Probit Model $
```

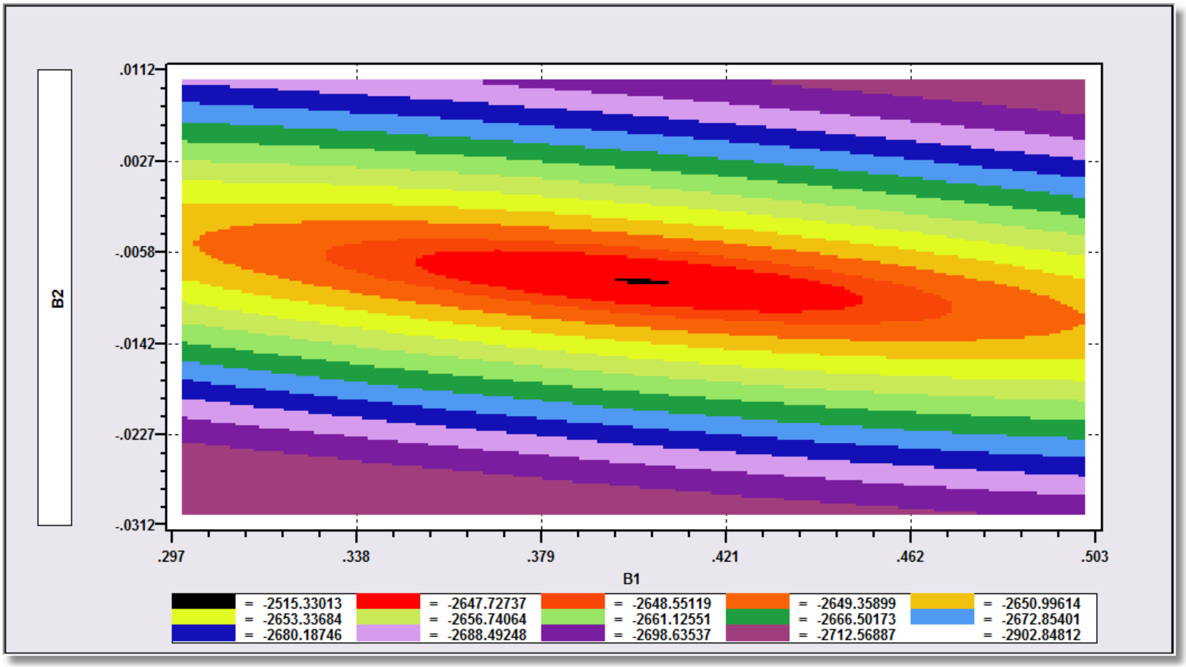


Figure E6.30 Contour Plot for Probit Log Likelihood

E7: Linear Regression – Estimation

E7.1 Introduction

This chapter will detail estimation of the single equation, linear regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{iK}\beta_K + \varepsilon_i$$
$$= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i,\ i = 1,\ldots,n.$$

The full set of observations is denoted for present purposes as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

The initial stochastic assumptions are the most restrictive for the linear model:

$$E[\varepsilon_i | \mathbf{X}] = 0 = E[\varepsilon_i]\ \forall\ i \quad \text{(zero mean)}$$
$$\text{Var}[\varepsilon_i | \mathbf{X}] = \text{Var}[\varepsilon_i] = \sigma^{2,}\ \forall\ i \quad \text{(homoscedastic)}$$
$$\text{Cov}[\varepsilon_i, \varepsilon_j | \mathbf{X}] = \text{Cov}[\varepsilon_i, \varepsilon_j] = 0\ \forall\ i, j \quad \text{(nonautocorrelation).}$$

More general models are described in the chapters to follow.

E7.2 Least Squares Regression Command

The basic command for the classical linear regression model is

REGRESS ; Lhs = dependent variable
; Rhs = regressors $

A constant term is *not* automatically included in the Rhs. If your model should contain a constant, you must include *one* among the Rhs variables. (Unless the model specifically dictates that there should be no constant term (as in certain time series settings), you should always include it.)

NOTE: Remember that *LIMDEP* does not automatically include a constant term in the equation. If you want one, be sure to include *one* among the Rhs variables.

The Rhs list may also include lagged variables, logs of variables, interaction terms, powers of variables, and so on. This is discussed further in Section E7.5. This command requests a linear ordinary least squares regression of the Lhs variable on the set of Rhs variables. The standard output from the procedure is listed in the next section.

This is the basic regression model. The limit on the number of parameters which may appear in the model is about 148 if you use no other specifications. If you use any of the optional procedures listed below, reduce this maximum to 146 to allow for the additional space needed for the computations.

The Main page of the command builder for the basic regression model appears below. This command builder is obtained by selecting Model:Linear Models/Regression. The minimum information provided from this dialog box is the Lhs and Rhs variables, as shown in Figure E7.1. Other options are discussed below.

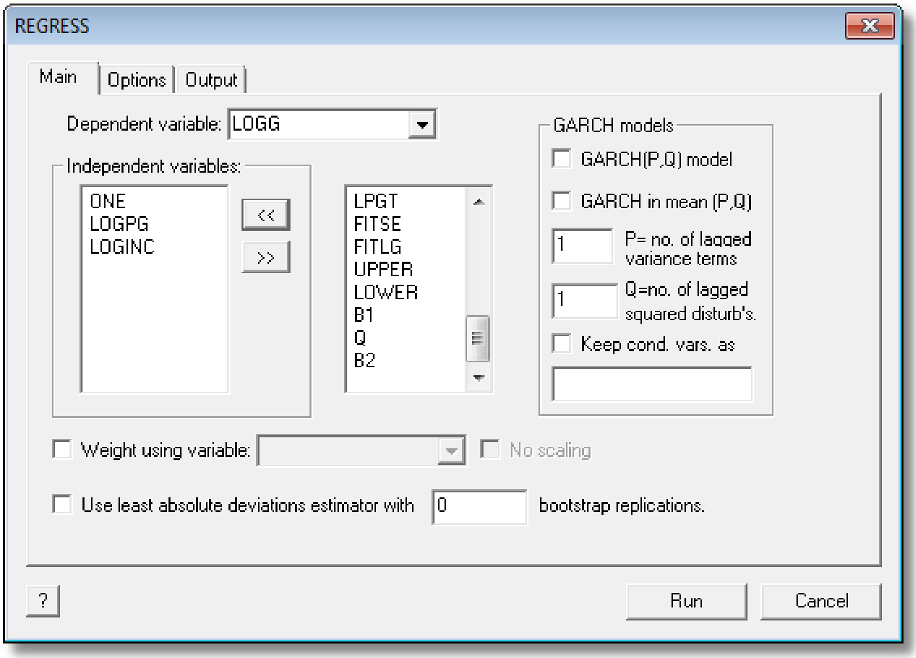


Figure E7.1 Main Page of the Command Builder for REGRESS

E7.3 Computing the Least Squares Coefficients

Least squares regression is based on the central estimation results

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$
$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

and the subsequent results that are derived from these. However, in order to maximize the accuracy of the computation, *LIMDEP* does not compute the least squares regression coefficients directly using the standard matrix algebra formula when it can be avoided. Consider the general case, ignoring practical considerations for the moment. The data matrix, **X**, can be written in its *QR* decomposition as

$$\mathbf{X} = \mathbf{Q}\mathbf{R}$$

where **Q** is $n \times K$, **R** is $K \times K$, $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and **R** is upper triangular. Using the *QR* decomposition of **X** in order to compute the least squares coefficients, rather than using an inversion method to apply the matrix formula allows extremely accurate solutions. (See the NIST benchmarks in Section E7.11.) Note, however, that using this method requires replication of the data matrix, so if **X** is very large, this may be impractical; *LIMDEP* uses the *QR* method if $n \times K \leq 33{,}000$ and if $n \leq 6{,}000$. If either of these constraints is exceeded, *LIMDEP* computes the moment matrix and inverts it using the *LU* decomposition method. This is an accurate inversion method, as good as or better than the common Cholesky method (which is used elsewhere in *LIMDEP*). (This does not place a limit on the number of observations you can use. If your data set has several million observations, you can still use all of *LIMDEP*'s estimation programs.)

Before attempting to compute a linear regression, *LIMDEP* makes one all out attempt to prevent you from using bad data to compute a linear model. We search for the condition of more than one variable with no variation in the model – one such variable, the constant, is to be expected.

More than one means trouble. The set of variables is examined. If more than one variable shows a variance less than 10^{-20}, we conclude that the regression is not estimable. In this case, a diagnostic such as the following (produced by a query from a user) will appear:

```
Variable MU1      always =     6.43751. No variation!
Variable D_C1     always =    -2.35884. No variation!
Regression cannot be computed. Collinearity
```

NOTE: *LIMDEP never* decides to just drop a few variables for you and compute the regression using those that remain. That decision is up to you, not the software. If your data are perfectly collinear because variables are identical, or because you have variables that have no variation, *LIMDEP* stops the estimation at that point, with a diagnostic of the problem.

E7.3.1 Results Produced by REGRESS

The linear regression produces a set of results such as the one below for the gasoline data used at several points in the earlier chapters: The command and results are as follows:

REGRESS ; Lhs = logg
; Rhs = one, logpg,loginc,Log(pnc),Log(puc),Log(ppt) $

NOTE: You can suppress all results in the command by including **; Quietly**. Why would you do this? You might be interested in producing only the retrievable results, but not actually seeing the surrounding regression results. For example, if you are computing a bootstrap estimator by computing the same regression for, say, 1,000 different random subsets of your sample, chances are, you are not interested in the visible results of the 1,000 regressions. Rather, only the sample variance of the 1,000 vectors of coefficients is of interest.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =        -.25713
             Standard deviation   =         .23849
             No. of observations  =             52  Degrees of freedom
Regression   Sum of Squares       =        2.79379          5
Residual     Sum of Squares       =        .107004         46
Total        Sum of Squares       =        2.90080         51
             Standard error of e  =         .04823
Fit          R-squared            =         .96311  R-bar squared =   .95910
Model test   F[  5,    46]        =      240.20584  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =       87.05475  Akaike I.C.   = -5.95537
             Restricted (b=0)     =        1.25792  Bayes  I.C.   = -5.73022
             Chi squared [  5]    =      171.59365  Prob C2 > C2* =   .00000
Model was estimated on May 08, 2011 at 11:40:49 PM
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*          Interval
--------+--------------------------------------------------------------------
Constant|  -11.5997***     1.48817   -7.79  .0000    -14.5165   -8.6829
   LOGPG|   -.03438         .04202    -.82  .4174     -.11673    .04797
  LOGINC|   1.31597***      .14198    9.27  .0000     1.03769   1.59425
  logPNC|   -.11964         .20384    -.59  .5601     -.51916    .27989
  logPUC|    .03754         .09814     .38  .7038     -.15481    .22990
  logPPT|   -.21514*        .11656   -1.85  .0714     -.44359    .01331
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

The statistics reported are as follows:

- The model framework – linear least squares regression
- The date and time when the estimates were computed
- Name of the dependent variable
- Mean of Lhs variable $\bar{y} = (1/n)\Sigma_i y_i,$
- Standard deviation of Lhs variable $s_y = \left\{[1/(n-1)]\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}^{1/2}$
- Name of the weighting variable if one was specified
- Number of observations $= n,$
- Number of parameters in regression $= K,$
- Degrees of freedom $= n\text{-}K,$
- Sum of squared residuals $\mathbf{e}'\mathbf{e} = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}(y_i-\mathbf{x}_i'\mathbf{b})^2$
- Standard error of e $s = \sqrt{\mathbf{e}'\mathbf{e}/(n-K)}$
- R^2 $R^2 = 1 - \mathbf{e}'\mathbf{e}/\sum_{i=1}^{n}(y_i-\mathbf{x}_i'\mathbf{b})^2$
- Adjusted R^2 $\bar{R}^2 = 1 - (n\text{-}1)/(n\text{-}K)[1 - R^2]$
- F statistic $F[K\text{-}1,n\text{-}K] = [R^2/(K\text{-}1)] / [(1\text{-}R^2)/(n\text{-}K)]$
- Prob value for F $Prob_F = \text{Prob}[F(K\text{-}1,n\text{-}k)] > \text{observed } F$
- Log likelihood $\log L = \text{-}n/2[1 + \log 2\pi + \log(\mathbf{e}'\mathbf{e}/n)]$
- Restricted log likelihood $\log L_0 = \text{-}n/2[1 + \log 2\pi + \log(s_y^2\ (\text{n-1})/\text{n})]$
- Chi squared[K-1] $\chi^2 = 2(\log L - \log L_0)$
- Prob value for chi squared $Prob_{\chi 2} = \text{Prob}[\chi^2(K\text{-}1)] > \text{observed chi squared}$
- Akaike Information Criterion $AIC = (\log L - K)/(n/2) - (1 + \log 2\pi)$
- Bayes Information Criterion $= \log[\text{e}'\text{e/n}) + \text{k} \log(\text{n})/\text{n}$

In time series settings, the results will also contain

- Durbin-Watson $dw = \Sigma_{t=2}^{T}(e_t - e_{t-1})^2 / \Sigma_{t=1}^{T}e_t^2$
- Autocorrelation $r = 1 - dw/2.$

The R^2 and related statistics are problematic if your regression does not contain a constant term. For the linear model, *LIMDEP* will check your specification, and issue a warning in the output, as shown below. In the results below, we have used the same **REGRESS** command, but omitted the constant term.

REGRESS ; Lhs = logg
; Rhs = logpg,loginc,Log(pnc),Log(puc),Log(ppt) $

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                  =        -.25713
             Standard deviation    =         .23849
             Number of observs.    =             52
Model size   Parameters            =              5
             Degrees of freedom    =             47
Residuals    Sum of squares        =        .248331
             Standard error of e   =         .07269
Fit          R-squared             =         .91439
             Adjusted R-squared    =         .90711
Model test   F[  4,    47] (prob) =   125.5(.0000)
Diagnostic   Log likelihood        =       65.16529
             Restricted(b=0)       =        1.25792
             Chi-sq [  4]  (prob) = 127.8(  .0000)
Info criter. Akaike Info. Criter. =       -5.15193
Not using OLS or no constant. Rsqrd & F may be < 0
Model was estimated on May 08, 2011 at 11:44:44 PM
-----------------------------------------------------------------------------
```

Note that the analysis of variance computations are now omitted.

Some additional notes about the standard least squares computations: (These are among our FAQs.) *The log likelihood can be positive!* It will be if $\mathbf{e'e}/n \le 0.058549$, and nothing in the model prevents this. Log likelihoods are only guaranteed to be negative for discrete choice models. If your model does not contain a constant term, then the restricted log likelihood that assumes only a constant term is meaningless in your model, and you should not use it as a basis for likelihood ratio tests.

Finally, the main table of regression results contains, for each Rhs variable in the regression:

- Name of the variable,
- Coefficient b_k,
- Standard error of coefficient estimate = se_k = the square root of the kth diagonal element of $s^2(\mathbf{X'X})^{-1}$
- t ratio for the coefficient estimate $t_k = b_k / se_k$
- Significance level of each t ratio based on the t distribution with $[n-K]$ degrees of freedom = p value = Prob$[t(n-K)] >$ observed t_k
- Confidence interval for coefficient. The default confidence level is 95%. You can change this with **; Clevel = value**, where *value* ranges from .05 to .99.

Footnotes to the table will often document computations that are not obvious or might not be well known, such as how the RESET test is computed. (See Section E7.9.2.)

E7.3.2 Retrievable Results

The retrievable results which are saved automatically by the **REGRESS** command are

Matrices: *b* = slope vector = $(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$
varb = estimated covariance matrix = $[\mathbf{e}'\mathbf{e}/(n\text{-}K)](\mathbf{X}'\mathbf{X})^{-1}$

Scalars:

ssqrd	=	$\mathbf{e}'\mathbf{e}/(n\text{-}K)$
rsqrd	=	R^2
s	=	s
sumsqdev	=	sum of squared deviations, $\mathbf{e}'\mathbf{e}$
rho	=	autocorrelation coefficient, r
degfrdm	=	$n\text{-}K$
sy	=	sample standard deviation of Lhs variable
ybar	=	sample mean of Lhs variable
kreg	=	number of independent variables, K
nreg	=	number of observations used to compute the regression, n Note, this may differ from the sample size if you have skipped missing values.
logl	=	log likelihood
exitcode	=	0.0 unless the data were collinear or OLS gives a perfect fit

Last Model: *b_name* where the names are the Rhs variables.
(See **WALD** in Chapter R14.)

Last Function: Conditional mean = $\mathbf{b}'\mathbf{x}$

The results listed above are all replaced by each regression. For example, after a **REGRESS** command is given, the names *b* and *varb* can be used in subsequent **MATRIX** and **CALC** commands for any computation. For example,

```
NAMELIST    ; x = x1,x2,x3,x4,one $
REGRESS     ; Lhs = y ; Rhs = x $
```

produces a 5×1 vector *b* and a 5×5 matrix *varb*. A subsequent **MATRIX** command might be

```
MATRIX      ; c = b(1:4) ; vc = Varb(1:4,1:4) ; wald = c'<vc>c $
```

This computes a Wald statistic for testing the hypothesis that the first four elements of β are zero. The last function noted above is used by **SIMULATE** to compute predictions (see the end of Section E7.6), by **PARTIALS** to compute partial effects (see Section E7.5) and by **DECOMPOSE** for Oaxaca decompositions (see Chapter R12).

E7.3.3 Results that Can Be Computed with MATRIX and CALC

Many of the particular statistics listed above can be computed with the **MATRIX** and **CALC** commands without producing all of the visible regression results. These functions can be used, for example, in computing regression results as parts of other kinds of programs.

Any of the regression results shown above can be reproduced with **MATRIX**. Here are a few of the computations. All depart from the definitions of the sample and regressor matrix:

SAMPLE ; ... or
INCLUDE ; ... or
PERIOD ; ... $
NAMELIST ; x = one,... $
CREATE ; y = ... the dependent variable $ Now, we have a common notation.
NAMELIST ; ny = y $ (Some functions require namelists, even if only one variable.)

You can compute the following regression statistics with **MATRIX** and **CALC**:

- Coefficients:

  MATRIX ; slopes = **Xlsq(x, y) or <x’x> * x’y**

- Sum of squares:

  MATRIX ; sumofsqs = **y’y - y’x * <x’x> * x’y**.
  MATRIX ; sumofsqs = **Rcpm(x,ny)**

- Covariance matrix for least squares slopes:

  MATRIX ; vc = {1/(n-Col(x))} * Rcpm(x,ny) * <x’x> $ (Uses **CALC** in { }.)

Other results can be computed with **MATRIX** as well, but since these are all scalars, it is likely to be easier to compute them with **CALC**, which also has several regression based statistics built into its functions. The ‘*x*’ and ‘*y*’ in the functions below are the namelist and variable defined above. For the statistics below to have their usual meaning, *x* should contain a constant term.

Rsq(*x*,*y*)	= R^2 in regression of *y* on *x*,	$1- \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 / \sum_{i=1}^{n}(y_i-\overline{y})^2$
Tss(*x*,*y*)	= total sum of squares,	$\sum_{i=1}^{n}(y_i-\overline{y})^2$
Ess(*x*,*y*)	= error, or residual sum of squares,	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$
Xss(*x*,*y*)	= explained sum of squares,	$\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$
Ser(*x*,*y*)	= standard error of regression,	$\left(\frac{1}{n-K}\sum_{i=1}^{n}e_i^2\right)^{1/2}$
Lik(*x*,*y*)	= log likelihood function,	$\frac{-n}{2}\left[1+\ln 2\pi+\frac{1}{n}\sum_{i=1}^{n}e_i^2\right]$

You can use any of these in subsequent commands. For example, to replicate the F statistic for testing the hypothesis that all coefficients are zero, you could use

CALC ; fstat = (n - Col(x)) * Xss(x,y) / Ess(x,y) / (Col(x) - 1) $

Of course, **MATRIX**, **CALC**, and **CREATE** commands may all be combined in programs that do any regression based computation you'd like.

E7.3.4 Beta Coefficients

Researchers are sometimes interested in 'beta coefficients' instead of the original regression coefficients. Beta coefficients are the linear regression coefficients that would result if data were standardized – centered around the mean then divided by the standard deviation – before computing the regression. In principle, there is no need to compute this regression separately; these coefficients can be computed from the original regression results by multiplying each regression coefficient by the ratio of the standard deviation of the dependent variable to the standard deviation of the respective independent variable. However, standardizing the data is a minor operation that produces the appropriate standard errors as well. The following example demonstrates for a small model:

CREATE ; s_logg = Std(logg) ; s_logp = Std(logpg) ; s_loginc = Std(loginc) $
REGRESS ; Lhs = s_logg ; Rhs = s_logp, s_loginc, one $

We include the redundant constant term so that the estimator will report the analysis of variance results.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=S_LOGG   Mean                    =        .00000
             Standard deviation      =       1.00000
             No. of observations     =            52  Degrees of freedom
Regression   Sum of Squares          =       47.8899          2
Residual     Sum of Squares          =       3.11011         49
Total        Sum of Squares          =       51.0000         51
             Standard error of e     =        .25194
Fit          R-squared               =        .93902  R-bar squared =   .93653
Model test   F[  2,    49]           =     377.25464  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =       -.55355  Akaike I.C.   = -2.70120
             Restricted (b=0)        =     -73.27993  Bayes  I.C.   = -2.58863
             Chi squared [  2]       =     145.45276  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 06:28:19 AM
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  S_LOGG|  Coefficient       Error        t    |t|>T*          Interval
--------+--------------------------------------------------------------------
  S_LOGP|    -.48753***      .10787    -4.52  .0000      -.69894   -.27611
S_LOGINC|    1.41654***      .10787    13.13  .0000      1.20512   1.62795
Constant|       0.0          .03494      .00 1.0000 -.68476D-01  .68476D-01
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E7.4 Stepwise Regression

LIMDEP will compute a least squares regression by the 'forward stepwise' method. The command is

REGRESS ; Lhs = ...
; Rhs = ... as usual (one may be omitted)
; Alg = stepwise $

Stepwise regression can also be selected from the Options page of the **REGRESS** command builder as shown in Figure E7.7. This command treats all variables in the Rhs list as candidates for entrance to or deletion from the equation. The constant term is always inserted automatically, so you can omit *one* from the Rhs list. (This is one of only a small handful of cases in which a constant term is assumed for you.) You can force a set of variables to be included in the equation with

; Rh2 = list of variables

With this specification included, the final equation will include all variables in the Rh2 list and those in the Rhs list which pass the tests for inclusion. The limits on the total number of variables in the Rhs and Rh2 lists are 70 in each.

Other options available for this form of the command are the usual for the linear regression model, including **; List**, **; Res = name**, and **; Keep = name**. These are described in detail in the following sections. The results which are saved by this command are exactly the same as those for the regression command described earlier

LIMDEP will pause between steps and ask you if you want to continue (take another step). You can suppress this dialogue and instruct the program to continue until all variables that satisfy the entry criterion have been entered by adding

; Output = 5

The stepwise regression method (see Kennedy and Gentle (1980)) is as follows: With a given set of variables already in the equation,

Step 1. Find the variable not yet entered which will raise R^2 by the largest amount. Enter it if the squared t ratio on that variable after it is entered exceeds the critical F ratio from the table for the appropriate degrees of freedom.

Step 2. Next, among the set of variables already entered, including the one from Step 1, if *one* was entered, find the one which leads to the smallest reduction in R^2 when it is deleted. Delete it if its squared t ratio is less than the critical F ratio.

After Steps 1 and 2, exit if any of the following criteria are met:

- The same variable was entered then deleted.
- All variables are entered.
- No variable can enter or exit.
- Fifty cycles are attempted.

The diagnostic statistics presented at each step are the standard analysis of variance results. The Mallows C_p criterion is

$$C_p = \mathbf{e'e} / s^2 + 2(P+1) - T$$

where **e′e** is based on the P included variables at the time and s^2 is $\mathbf{e'e}/(n\text{-}K)$ with all variables included. The limiting value, with all variables included is $2(P+1)$.

In the example below, we reestimated the gasoline consumption equation with all price variables. The command forces the log of the gasoline price and the log of income to remain in the equation, and uses the stepwise method to select among the micro- and macroeconomic price indices. (Note, this example illustrates the danger of relying on the mechanical selection of variables. If **; Rh2** is not used to force the gasoline price to remain in the model, it does not appear in the final specification.)

REGRESS ; Lhs = logg
; Rhs = logpnc,logpuc,logppt,logpn,logpd,logps
; Rh2 = logpg,loginc
; Alg = stepwise
; Output = 5 $

```
+--------------------------------------------------------------------------+
| Stepwise Regression                                                      |
| Dependent variable     = LOGG                                            |
| Number of observations =       52                                        |
| Number of regressors   =        8                                        |
| Degrees of freedom     =       43                                        |
| Predictor variables are:                                                 |
| LOGPNC   LOGPUC   LOGPPT   LOGPN    LOGPD    LOGPS    LOGPG    LOGINC    |
+--------------------------------------------------------------------------+
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23849
             No. of observations  =            52  Degrees of freedom
Regression   Sum of Squares       =       2.80737           3
Residual     Sum of Squares       =    .934290E-01         48
Total        Sum of Squares       =       2.90080          51
             Standard error of e  =        .04412
Fit          R-squared            =        .96779  R-bar squared =   .96578
Model test   F[  3,    48]        =     480.77026  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      90.58193  Akaike I.C.   = -6.16795
             Restricted (b=0)     =       1.25792  Bayes  I.C.   = -6.01786
             Chi squared [  3]    =     178.64802  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 07:02:01 AM
Mallows Cp statistic              =        39.011
-----------------------------------------------------------------------------
Analysis of Variance for the Current Regression
Source       Deg.Fr.      Sum of squares      Mean Square      F
Regression      3                2.80737          .93579 480.77
Residual       48                 .09343          .00195
Total          51                2.90080          .05688
Variable entered this step = LOGPS    , Deleted =
Note: First  2 variables are forced in.
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
   LOGPG|     .02995         .04145      .72  .4734     -.05128    .11119
  LOGINC|    1.46684***      .09339    15.71  .0000     1.28380   1.64987
   LOGPS|    -.36557***      .05583    -6.55  .0000     -.47499   -.25616
Constant|   -13.0339***      .76701   -16.99  .0000    -14.5372  -11.5306
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
(Intermediate steps omitted)
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =        -.25713
             Standard deviation   =         .23849
             No. of observations  =             52  Degrees of freedom
Regression   Sum of Squares       =        2.83882           5
Residual     Sum of Squares       =    .619739E-01          46
Total        Sum of Squares       =        2.90080          51
             Standard error of e  =         .03671
Fit          R-squared            =         .97864  R-bar squared =   .97631
Model test   F[  5,    46]        =      421.42166  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      101.25461  Akaike I.C.   = -6.50152
             Restricted (b=0)     =        1.25792  Bayes  I.C.   = -6.27637
             Chi squared [  5]    =      199.99337  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 07:02:06 AM
Mallows Cp statistic              =         14.430
-----------------------------------------------------------------------------
Analysis of Variance for the Current Regression
Source       Deg.Fr.      Sum of squares      Mean Square      F
Regression      5              2.83882           .56776 421.42
Residual       46               .06197           .00135
Total          51              2.90080           .05688
Variable entered this step =         , Deleted =
Note: First  2 variables are forced in.
**********> This is the final equation <**********
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
   LOGPG|    -.06744         .04067    -1.66  .1041     -.14716    .01227
  LOGINC|    1.62664***      .09183    17.71  .0000     1.44664   1.80663
   LOGPS|    -.51615***      .08632    -5.98  .0000     -.68533   -.34696
   LOGPN|     .57218***      .12261     4.67  .0000      .33187    .81249
  LOGPUC|    -.20629**       .08545    -2.41  .0198     -.37378   -.03881
Constant|   -15.1666    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

E7.5 Interactions and Partial Effects

When the regression contains nonlinearities and interaction terms, such as logs of variables, squares or cross products of variables, the raw coefficients in the model do not reveal the actual relationship between the dependent variable   Consider the regression model

$$\log(income) = \beta_1 + \beta_2\, age + \beta_3\, age^2 + \beta_4\, educ + \beta_5\, female + \beta_6\, educ{\times}female + \varepsilon.$$

In this regression model, the coefficients are semi-elasticities. However, none of them give the relevant effect of a variable on log(*income*). For example, the effect of *age* is not β_2, it is $\beta_2 + 2\beta_3 age$, the female differential is not β_5, it is $\beta_5 + \beta_6 educ$, and the impact of *educ* is not β_4, it is $\beta_4 + \beta_6\, female$. Each of these could be easily computed after the linear regression (with a hand calculator if necessary). However, if one wants to compute standard errors and/or confidence intervals for the coefficients, that is rather more complicated. For example, the estimator of the standard error for the impact of *age* on log(*income*) is

$$\text{Est.Std.Err}(age \text{ effect}) = [\text{Var}(b_2) + 4age^2\text{Var}(b_3) + 4age\text{Cov}(b_2,b_4)]^{1/2},$$

which is likely to be quite inconvenient. The **PARTIALS** command is provided for this purpose, and fully automates the computation.

Using **PARTIALS** to compute effects of nonlinearities is done in a second step after the regression. The first step involves specifying the regression with the nonlinearities specified explicitly in the equation. For the example, the following are two ways to compute the regression, where we use the health satisfaction data used in earlier examples for the illustration:

```
SAMPLE        ; All $
INCLUDE       ; New ; year = 1991 $
CREATE        ; educ_fem = educ*female ; agesq = age*age $
CREATE        ; loginc = Log(hhninc) $
? First method
REGRESS       ; Lhs = loginc
              ; Rhs = one,age,agesq,educ,female,educ_fem $
? Second method
REGRESS       ; Lhs = Log(hhninc)
              ; Rhs = one,age,age*age,educ,female,educ*female $
```

The two methods of computing the regression give identical results save for a slight difference in labeling, as can be seen below. However, using the second method allows the **PARTIALS** command to detect that the model contains the nonlinearities and interaction terms and to compute the partial effects for you. That is, the regression program has no way to know that a variable named *agesq* is the square of one named *age* that appears elsewhere in the list of variables in the model. But, in the second specification, that relationship appears specifically.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGINC   Mean                     =     -1.00062
             Standard deviation       =       .46494
             No. of observations      =         4340  Degrees of freedom
Regression   Sum of Squares           =      119.784            5
Residual     Sum of Squares           =      818.169         4334
Total        Sum of Squares           =      937.953         4339
             Standard error of e      =       .43449
Fit          R-squared                =       .12771  R-bar squared =   .12670
Model test   F[  5,  4334]            =    126.90421  Prob F > F*   =   .00000
Diagnostic   Log likelihood           =  -2537.41550  Akaike I.C.   = -1.66580
             Restricted (b=0)         =  -2833.90545  Bayes  I.C.   = -1.65698
             Chi squared [  5]        =    592.97990  Prob C2 > C2* =   .00000
First Method
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGINC|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -3.28158***      .10929  -30.03  .0000    -3.49578   -3.06738
     AGE|     .08189***      .00492   16.65  .0000      .07225     .09153
   AGESQ|    -.00091***   .5578D-04  -16.38  .0000     -.00102    -.00080
    EDUC|     .05087***      .00366   13.90  .0000      .04370     .05805
  FEMALE|     .04460         .06630     .67  .5012     -.08536     .17455
EDUC_FEM|    -.00656         .00568   -1.15  .2484     -.01770     .00458
--------+--------------------------------------------------------------------
Second Method
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
logHHNIN|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -3.28158***      .10929  -30.03  .0000    -3.49578   -3.06738
     AGE|     .08189***      .00492   16.65  .0000      .07225     .09153
 AGE*AGE|    -.00091***   .5578D-04  -16.38  .0000     -.00102    -.00080
    EDUC|     .05087***      .00366   13.90  .0000      .04370     .05805
  FEMALE|     .04460         .06630     .67  .5012     -.08536     .17455
        |Interaction EDUC*FEMALE
Intrct03|    -.00656         .00568   -1.15  .2484     -.01770     .00458
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

After estimation, the **PARTIALS** command can be used as follows:

PARTIALS ; Effects: age / educ / female ; Summary $

```
---------------------------------------------------------------------
Partial Effects for Linear Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .00320     .00059    5.46       .00205       .00435
      EDUC           .04770     .00286   16.68       .04210       .05331
   *  FEMALE        -.03081     .01339    2.30      -.05706      -.00456
---------------------------------------------------------------------
```

The results shown provide the partial effect for each variable averaged over the sample observations. Note, in particular, how b_5, the coefficient on *female* in the original regression, which equals +.045, is misleading regarding the female differential when the nonlinear effect of *educ* is taken into account.

There are many optional features and specifications available for partial effects described in Chapter R11. One possibility that shows clearly the implication of the quadratic specification in *age*, is to plot the partial effect of *age* at a range of values, as shown below.

PARTIALS ; Effects: age & age = 20(5)70 ; Plot(ci) $

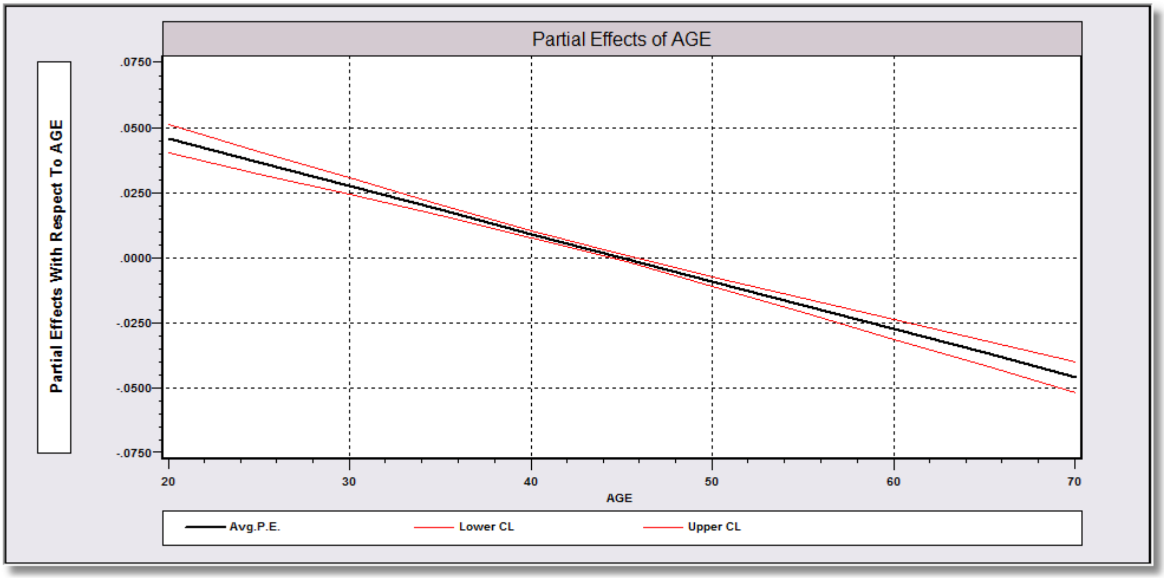


Figure E7.2 Partial Effects for Quadratic Regression Model

E7.6 Predictions and Residuals

To obtain a list of the residuals and fitted values from a linear regression model, add the specification

; List

to the command. The residuals and predicted values may be kept in your data area by using the specifications

; Res = name to retain residuals

and **; Keep = name** to retain predictions

If you are not using the full sample or all of the rows of your data matrix, some of the cells in these columns will be marked as missing. If you have data on the regressors but not the dependent variable, you can use

; Fill

to obtain predictions for the missing data. Remember, though, that the prediction is -999 (missing) for any observation for which any of the *x*s are missing. (You can use this procedure as an alternative to the multiple imputation method discussed in Chapter R20.)

The command builder dialog box for these and several other options are on the Output page of the **REGRESS** command builder, as shown below in Figure E7.3.

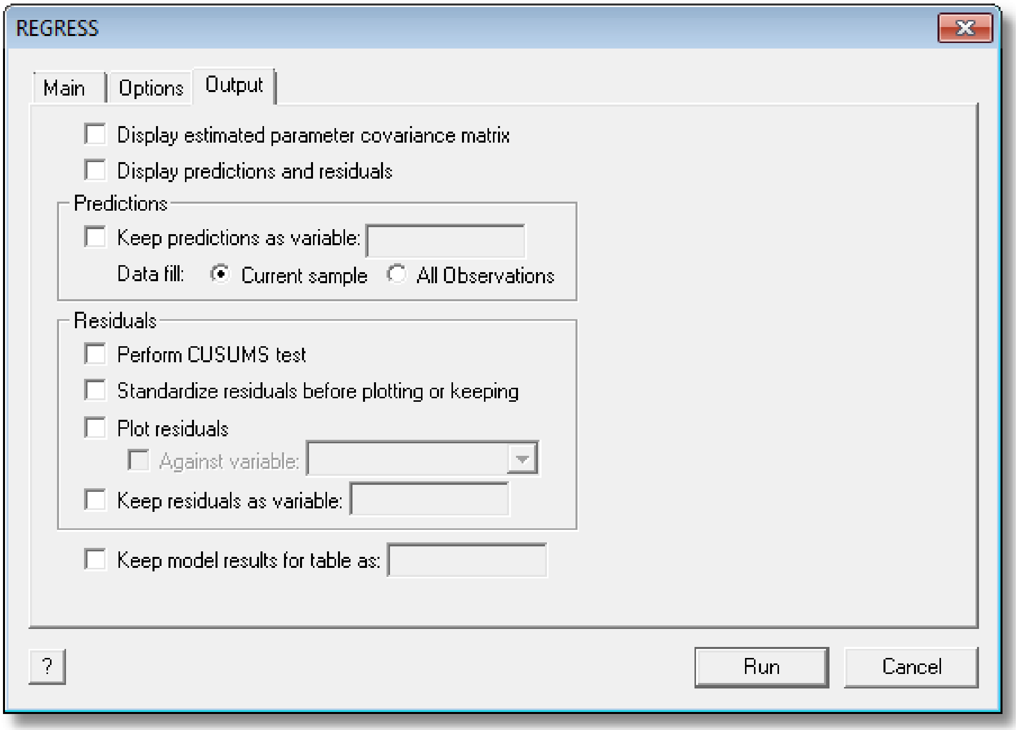


Figure E7.3 Command Builder for Regression Residuals and Predictions

The listing will also contain a 95% confidence interval for the forecast of the dependent variable. (See the example below.) The confidence limits are not kept. But, listed below are commands which can be used to obtain this result.

For the model estimated earlier with the gasoline, we now use **; List** to obtain a list of fitted values.

REGRESS ; Lhs = logg
; Rhs = one,logpnc,logpuc,logppt,logpn,logpd,logps,logpg,loginc
; List $

```
(Regression results omitted)
Predicted Values          (* => observation was not in estimating sample.)
Observation         Observed Y   Predicted Y   Residual     95% Forecast Interval
  1953              -.7988548    -.7486919    -.0501629    -.8337218    -.6636619
  1954              -.7851131    -.7433575    -.0417556    -.8232908    -.6634243
  1955              -.7192742    -.6736580    -.0456163    -.7538397    -.5934763
  1956              -.6782199    -.6102414    -.0679785    -.6876777    -.5328050
  1957              -.6584227    -.6296708    -.0287519    -.7053592    -.5539823
 (Rows 1958 - 2001 omitted)
  2002               .0080191     .0124173    -.0043982    -.0664631     .0912976
  2003               .0050948     .0035076     .0015871    -.0793875     .0864027
  2004              -.0080513     .0256079    -.0336592    -.0606169     .1118326
```

The following commands could be used for computing forecast standard errors. This routine uses the matrices *b* (the coefficients) and *varb* (estimated covariance matrix) kept by the regression and scalar *ssqrd* which is s-squared from the regression. The forecast standard errors are the values computed by the Sqr function in the **CREATE** command.

NAMELIST ; x = the set of regressors $
REGRESS ; Lhs = y ; Rhs = x ; Keep = yhat $
CALC ; ct = Ttb(.975,degfrdm) $
CREATE ; lowerbnd = yhat - ct * Sqr(ssqrd + Qfr(x,varb))
; upperbnd = yhat + ct * Sqr(ssqrd + Qfr(x,varb)) $

The built in simulator may also be used to obtain predictions and confidence intervals for predictions. The command

SIMULATE ; List $

immediately after the regression command produces the following results.

```
---------------------------------------------------------------------
Model Simulation Analysis for Linear Regression Function
---------------------------------------------------------------------
Simulations are computed by average over sample observations
---------------------------------------------------------------------
User Function      Function   Standard
(Delta method)      Value      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
Obs. =    1        -.74869     .02232  -33.55     -.79243     -.70495
Obs. =    2        -.74336     .01707  -43.55     -.77681     -.70990
Obs. =    3        -.67366     .01735  -38.82     -.70767     -.63965
Obs. =    4        -.61024     .01395  -43.73     -.63759     -.58289
Obs. =    5        -.62967     .01135  -55.46     -.65193     -.60742
(Rows 6 - 49 omitted)
Obs. =   50         .01242     .01582     .78     -.01859      .04342
Obs. =   51         .00351     .02025     .17     -.03618      .04319
Obs. =   52         .02561     .02342    1.09     -.02029      .07151
Avrg. Function     -.25713     .00496   51.83     -.26685     -.24741
```

Note that although the predictions are the same as those produced by the regression, the confidence intervals are narrower. The reason is that **SIMULATE** produces a confidence interval for the simulated value of the dependent variable, not a forecast interval. The forecast variance used by **REGRESS ; List…$** equals Var[*forecast*] = $s^2 + \mathbf{x}'[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{x}$ while the variance for the simulated value omits the leading s^2 term. In general, the latter will be smaller than the former. No generality is possible save that the larger is R^2 the closer will be the two values.

E7.6.1 Plotting Residuals

A plot of the residuals from your regression can be requested by adding

; Plot

to the command. Residuals are plotted against observation number (i.e., simply listed). For example, the following would generate the residual plot for the preceding gasoline market example.

```
DATES       ; Undated $
SAMPLE      ; 1-52 $
REGRESS     ; Lhs = logg
            ; Rhs = one,logpg,loginc,logpnc,logpuc,logppt,logpn,logpd,logps
            ; Plot $
```

You will get a time series style plot if the data have been identified as time series data with a **DATES** command. The same plot preceded by

```
DATES       ; 1953 $
PERIOD      ; 1953-2004 $
```

appears as follows in the second panel of Figure E7.4.

If you would like to plot the residuals against another variable, change the preceding to

; Plot(variable name)

The variable can be any existing variables. It need not have been used in the regression. The residuals are sorted according to the variable you name and plotted against it. In the third panel, we used

```
REGRESS     ; Lhs = logg
            ; Rhs = one,logpnc,logpuc,logppt,logpn,logpd,logps,logpg,loginc
            ; Plot(pnc) $
```

The plot will show the residuals graphed against either the observation number, the date for time series data, or the variable you specify using the **; Plot(variable)** option described above.

- If there are outliers in the data, this may severely cramp the figure, since the vertical axis is scaled so that every observation will appear.

- The mean residual bar may not appear at zero because the residuals may not have zero mean. They will not if you do not have a constant term in your regression or if you are plotting two stage least squares residuals. Since 0.0 will generally not be the midpoint between the high and low residual, the zero bar will not be in the center of your screen even when you do have a constant term in the model.

You can plot up to 5,000 observations in the figure with this option.

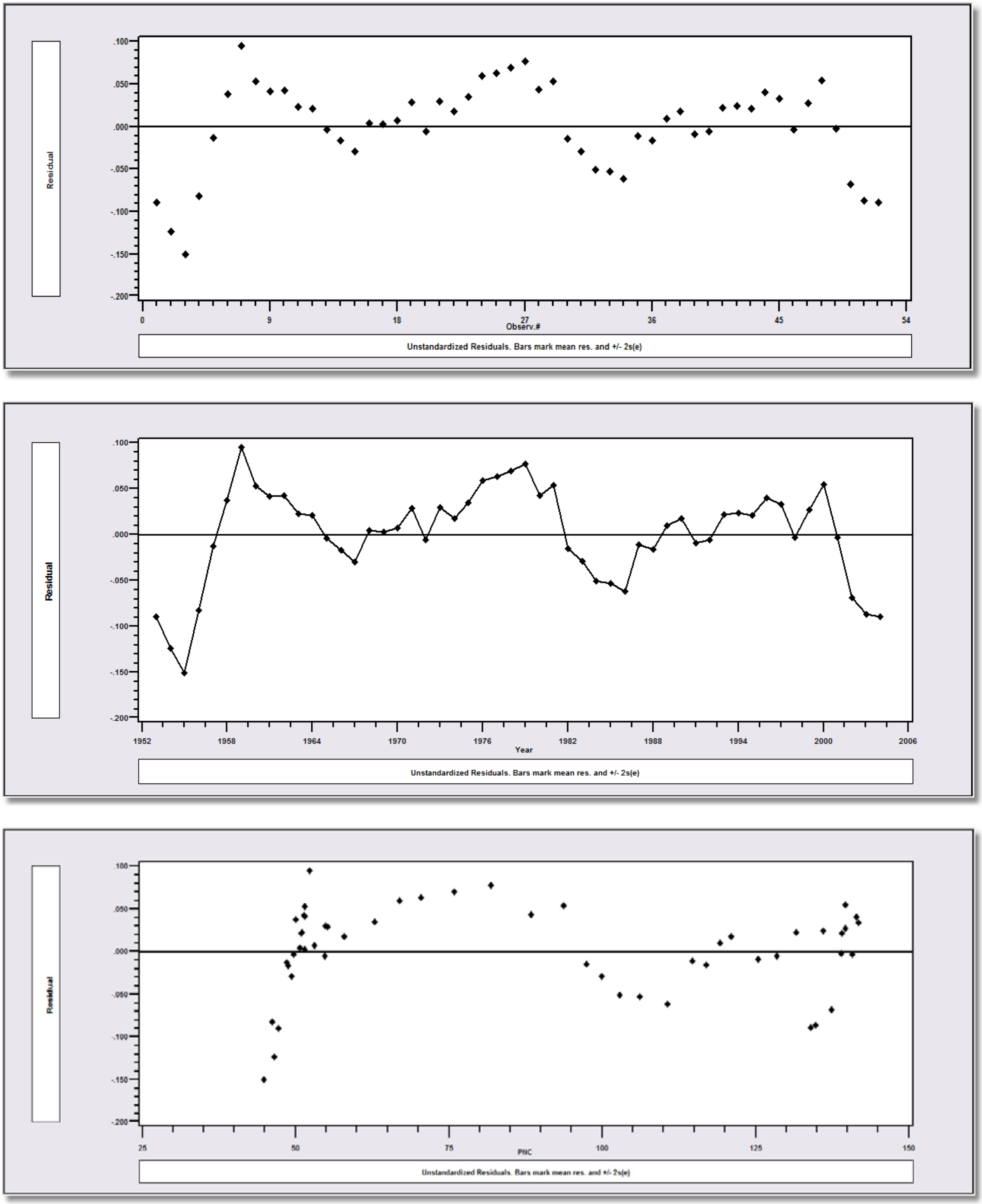


Figure E7.4 Residual Plots

E7.6.2 Standardized Residuals and Regression Diagnostics

In the linear regression model, the variance of a least squares residual is not σ^2, but

$$\text{Var}[e_i] = \sigma^2[1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i] = \sigma^2 (1 - h_{ii}).$$

Belsley, Kuh, and Welsch (1980) suggest that the standardized residuals,

$$u_i = e_i / \text{Est.Var}[e_i]^{1/2}$$

be plotted instead of the raw residuals as a more useful diagnostic tool. Values of u_i in excess of two indicate possible outliers. You can plot the standardized residuals if the regression command includes

; Plot ; Standardized

To retain the standardized residuals, just use

; Res = name ; Standardized

NOTE: These residuals have a mean and variance that will be close to 0.0 and 1.0, respectively. But, unlike the ordinary OLS residuals, they do not have a mean identically equal to 0.0 and they are only approximately orthogonal to the regressors. In fact, Est.Var[e_i] = s^2 - o(1/n), so as the sample size increases, the standardized residuals will converge to the OLS residuals.

The following is a plot of the standardized residuals. This corresponds to the center frame in Figure E7.4. Even with only 52 observations, save for the scale, the residuals are essentially the same.

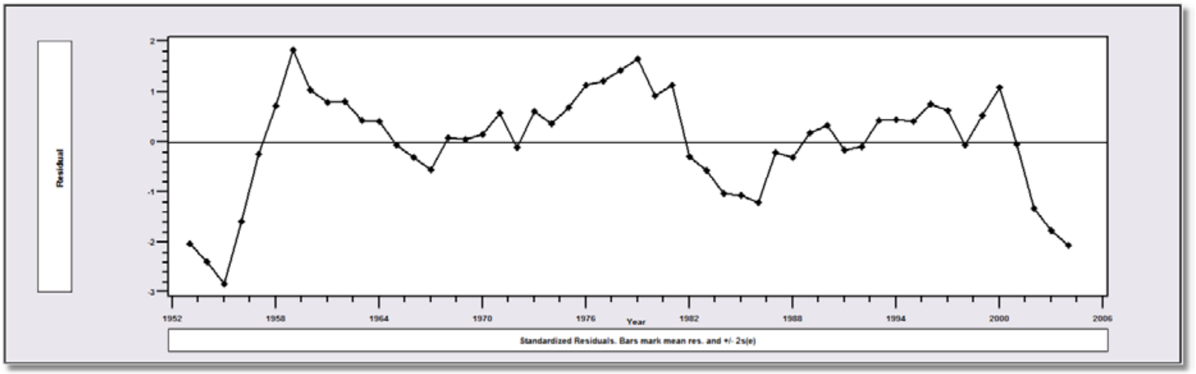


Figure E7.5 Standardized Residuals

An additional quantity of interest is the 'leverage' value,

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i.$$

Note that Var[e_i] = σ^2(1 - h_{ii}). Belsley et al. suggest values of h_{ii} greater than $2K/n$ signal points worthy of attention. To obtain them, we require the 'hat matrix,' i.e., the projection matrix into the column space of $\mathbf{X}$,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

This is an $n \times n$ matrix, which will be quite unmanageable if n is large. But, in fact, we only require the diagonal elements. To do this computation for a particular data set, you can use

```
NAMELIST    ; x = ... variables in X $
MATRIX      ; xxi = <x'x>  $
CREATE      ; hii = Qfr(x,xxi)  $
CALC        ; big = 2 * Col(x) / n $
CREATE      ; outlier = ( hii > big ) $
```

A list of the variable named *outlier* will flag the important observations with values of one. (Other observations get a zero.) When we apply this to the regression immediately above, the diagnostic identifies three years, 1978-1980, as outliers.

As a further refinement, Belsley et al. suggest that for each residual, the coefficient vector, **b**, and the residual variance, s^2, be reestimated without that observation. In principle, this requires that the regression be recomputed for each observation. But, there are some shortcuts which make the computation quite simple. To describe this procedure, we require some regression algebra. Let **X** be the $n \times K$ matrix of regressors. The least squares estimator is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The standardized residuals were computed above

$$u_i = e_i / [s^2(1 - h_{ii})]^{1/2}.$$

This does not recompute s^2 without the ith observation. Adding in Belsley et al.'s refinement takes a bit more work. First, they show that if the regression is recomputed without observation 'i,' that the resulting slope estimator is

$$\mathbf{b}(i) = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i / (1 - h_{ii}).$$

Therefore, the residual vector from this regression (where, for the moment, we include the ith observation in the residual vector) is

$$\begin{aligned}\mathbf{e}(i) &= \mathbf{y} - \mathbf{X}\mathbf{b}(i)\\ &= \mathbf{e} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i / (1 - h_{ii}).\end{aligned}$$

Multiplying it out, and remembering that $\mathbf{X}'\mathbf{e} = \mathbf{0}$, we get the sum of squared residuals for the full sample based on $\mathbf{b}(i)$,

$$\mathbf{e}(i)'\mathbf{e}(i) = \mathbf{e}'\mathbf{e} + e_i^2 h_{ii} / (1 - h_{ii})^2.$$

Now, we have to subtract out the square of the ith residual. This is

$$\begin{aligned}y_i - \mathbf{x}_i'\mathbf{b}(i) &= e_i[1 + h_{ii}/(1 - h_{ii})]\\ &= e_i / (1 - h_{ii}).\end{aligned}$$

Subtracting the square of this from $\mathbf{e}(i)'\mathbf{e}(i)$ produces

$$\mathbf{e}_*(i)'\mathbf{e}_*(i) = \mathbf{e}'\mathbf{e} - e_i^2/(1 - h_{ii}).$$

This shows the 'shortcut.' The regression need not be recomputed. Finally, the estimator of σ^2 is

$$s^2(i) = \mathbf{e}_*(i)'\mathbf{e}_*(i) / (n - K - 1).$$

Combining terms, we obtain the desired standardized residual

$$u(i)_i = [e_i/(1-h_{ii})] / [(\mathbf{e}'\mathbf{e} - e_i^2/(1-h_{ii}))/(n-K-1)]^{1/2}$$

A related computation is Belsley et al.'s '*dfit*' which is an observation specific measure which attempts to capture the influence, or leverage effect as well as the effect of the residual, itself. The calculation is

$$dfit_i = u(i)_i \times \sqrt{h_{ii}}\ .$$

The following commands will obtain these standardized residuals. (The algebra is far more complicated than the actual computation.)

```
NAMELIST   ; ... define x  $
CREATE     ; ... define y $
REGRESS    ; Lhs  = y ; Rhs = x ; Res = ei $
MATRIX     ; xxi   =  <x'x> $
CREATE     ; mii  = (1 - Qfr(x,xxi))
           ; uii    = ei/mii / Sqr((sumsqdev - ei*ei/mii) / (degfrdm - 1))
           ; dfiti = uii * Sqr(1 - mii) $
```

E7.7 Multicollinearity

If there is a linear combination of the independent variables which produces a column of zeros – i.e., at least one column of **X** can be represented as a linear combination of other columns of **X** – then the least squares regression coefficient cannot be computed by inverting the moment matrix. In this case, the minimizer of the sum of squared residuals is not unique. As a general rule, *LIMDEP* does not proceed any further if it detects that your data are collinear. (Some other programs will successively drop variables from the equation until a noncollinear set remains, as if to report that your desired model was inestimable, so the program found some other model that was. While users differ in their preference for this kind of program driven specification, *LIMDEP* adheres to a strict rule of always waiting for the user to specify the model to be estimated.)

In the case of linear regression, sometimes multicollinearity cannot be detected even when it is present. Recall that in most cases, *LIMDEP* is not using **X′X** to compute the regression, so the presence of multicollinearity may not be obvious. (Even when the moment matrix is being used, the assessment of multicollinearity is only to within some tolerance.) Because the presence of internal rounding error may leave some variation in the representation of the raw data, data may be multicollinear in theory, but only approximately so and, therefore, not, internally. The *QR* method can then become unstable, and report coefficients which appear nonsensical. In some extreme cases, *LIMDEP* will report the condition number for **X'X** with a warning that the data are highly collinear.

NOTE: The condition number is the square root of the ratio of the largest to the smallest characteristic root of $(1/n)\mathbf{X}^{0\prime}\mathbf{X}^0$ in which the first column of $\mathbf{X}^0$ is a column of ones and the remaining columns are the original data not including the constant, in deviations from their means.

For example, the Filippelli data and example discussed in Section E7.11 are a notoriously collinear test data set. Although estimation can still proceed, the following warning is produced by the least squares estimator for this problem:

```
REGRESS      ; Lhs = y ; Rhs = one,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10 $
WARNING: Badly conditioned X. Condition value =    .2999482D+10
```

E7.8 Variance Inflation Factors

Authors sometimes analyze multicollinearity in terms of the effect of the intercorrelation of the regressors on the variances of the least squares coefficient estimators. The *variance inflation factor* is a measure of this effect;

$$VIF_k = \frac{1}{1 - R_{k.}^2}$$

where $R_{k.}^2$ is the R^2 obtained when the kth regressor is regressed on the remaining variables. The optimal value for this statistic is 1.0, which occurs when the R^2 is zero, or this variable is orthogonal to the other variables. Some fairly straightforward algebra reveals that, if the model contains a constant term – that is, one of the columns of **X** is a column of ones – then,

$$VIF_k = \left(\sum_{i=1}^{n} \left(x_{ik} - \overline{x}_k\right)^2\right) \times (\mathbf{X}'\mathbf{X})^{kk}$$

where $(\mathbf{X}'\mathbf{X})^{kk}$ is the kth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. Thus, these auxiliary regressions need not be computed to obtain these factors. The vector of variance inflation factors for the entire coefficient vector can be computed and displayed with the following matrix command:

```
NAMELIST   ; x = ... the list of variables $
MATRIX     ; List ; xm0x = {n-1}*Xvcm(x)
           ; vif = Diag(<x'x>) * Vecd(xm0x) $
```

There is no consensus on what values of the variance inflation factor merit attention, or on what one should do with the results. Some authors (Chatterjee and Price (1991)) suggest that values in excess of 10 are problematic. Others suggest 30 or 40 as a benchmark value. In any event, it is less than obvious what one should do upon finding a large value (or some other indicator of a ‘multicollinearity problem’). As noted earlier, *LIMDEP* leaves this up to the user.

The example below applies the preceding to gasoline data used in the preceding example in which **X** = [*one,logpnc,logpuc,logppt,logpn,logpd,logps,logpg,loginc*]

```
     VIF|             1
--------+--------------
       1|       .000000
       2|       658.212
       3|       189.992
       4|       795.683
       5|       313.746
       6|       1593.98
       7|       5152.04
       8|       72.0155
       9|       220.726
```

Though the diagnostics seem to suggest a high degree of multicollinearity, the regression seems completely routine.

E7.9 Specification Analysis

Several devices are used to assess the adequacy of the model specification. Four that are automated are the tests for heteroscedasticity, functional form, omitted variables and autocorrelation. Others can be programmed with the command language, mainly **CREATE**, **CALC** and **MATRIX**.

E7.9.1 Breusch and Pagan Test for Heteroscedasticity

The Breusch and Pagan (B-P) test for heteroscedasticity is narrowly defined for the hypothesis of homoscedasticity in linear regression with normally distributed disturbances. The full setup for the test is $y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$ where $\varepsilon_i \sim N[0, g(\sigma^2 + \boldsymbol{\alpha}'\mathbf{z}_i)]$, so that the hypothesis of homoscedasticity is equivalent to $\boldsymbol{\alpha} = \mathbf{0}$. The variables in **z** may be the **x** in the original regression, or other variables that might appear in the model. The test statistic, which has a limiting chi squared distribution with degrees of freedom equal to the number of elements in **z**, is computed as one half the regression sum of squares in the linear regression of $w_i = [e_i^2/(\mathbf{e}'\mathbf{e}/n) - 1]$ on $[\mathbf{z},1]$. Let $\mathbf{Z}$ be the $n \times P$ matrix that has ith row $[\mathbf{z}_i',1]$. Then, the statistic is, $BP = \frac{1}{2}\, \mathbf{w}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{w}$. Several decades of research have suggested that the test has power to detect heteroscedasticity if the normality assumption is weakened.

Request this test by adding

; BPT

or **; BPT = list of variables in z**.

If the list is omitted, then the test is carried out assuming that **z** = **x** not including the constant term. The result of the test will be displayed with the regression results, as shown in the example below.

REGRESS ; Lhs = logg
; Rhs = one,logpg,loginc,logpnc,logpuc,logppt,logpn,logpd,logps
; BPT $

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                    =       -.25713
             Standard deviation      =        .23849
             No. of observations     =            52  Degrees of freedom
Regression   Sum of Squares          =       2.84577           8
Residual     Sum of Squares          =    .550250E-01         43
Total        Sum of Squares          =       2.90080          51
             Standard error of e     =        .03577
Fit          R-squared               =        .98103  R-bar squared =    .97750
Model test   F[  8,    43]           =     277.98326  Prob F > F*   =    .00000
Diagnostic   Log likelihood          =     104.34671  Akaike I.C.   = -6.50506
             Restricted (b=0)        =       1.25792  Bayes  I.C.   = -6.16734
             Chi squared [  8]       =     206.17758  Prob C2 > C2* =    .00000
B-P test     Chi squared [  8]       =       8.73211  Prob C2 > C2* =    .36540
Model was estimated on May 09, 2011 at 07:15:15 PM
-----------------------------------------------------------------------------
(Regression results omitted)
-----------------------------------------------------------------------------
```

The B-P test is also carried out automatically assuming that **z** = **x** when you request the heteroscedasticity robust covariance matrix described in Section E7.10.1. The command specification is

; Heteroscedasticity (or ; Het)

In this case, the regression results contain the test statistic as well as the results based on the robust covariance matrix.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23849
             Number of observs.   =            52
Model size   Parameters           =             9
             Degrees of freedom   =            43
Residuals    Sum of squares       =    .550250E-01
             Standard error of e  =        .03577
Fit          R-squared            =        .98103
             Adjusted R-squared   =        .97750
Model test   F[  8,     43] (prob) =   278.0(.0000)
White heteroscedasticity robust covariance matrix.
Br./Pagan LM Chi-sq [  8]  (prob) =   8.73 (.3654)
Model was estimated on May 09, 2011 at 07:23:12 PM
-----------------------------------------------------------------------------
(Regression results omitted)
-----------------------------------------------------------------------------
```

E7.9.2 RESET Specification Test

The regression specification error test (RESET) (Ramsey, 1969) is a general specification test of the adequacy of the linear functional form in the model

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i.$$

The test is carried out in various ways in the literature, all asymptotically equivalent to a linear regression of the regression residuals, e_i on powers of the regression predictions, $(\mathbf{b}'\mathbf{x}_i)^2$, $(\mathbf{b}'\mathbf{x}_i)^3$, etc. The logic of the test is that if the regression is adequately specified by the linear functional form, then addition of the powers of $\mathbf{b}'\mathbf{x}_i$ should not provide additional explanatory power. The test is carried out in a second step after the regression is computed by regressing the least squares residuals on a constant term and the second, third and fourth powers of the predicted values. The test statistic is a Wald statistic based on the three coefficients in this second regression.

The RESET test is requested by adding

; RESET

to the **REGRESS** command. Results of the test will appear in the diagnostic header for the regression model, as shown in the example below.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                    =      -.25713
             Standard deviation      =       .23849
             No. of observations     =           52  Degrees of freedom
Regression   Sum of Squares          =      2.84577          8
Residual     Sum of Squares          =   .550250E-01        43
Total        Sum of Squares          =      2.90080         51
             Standard error of e     =       .03577
Fit          R-squared               =       .98103  R-bar squared =   .97750
Model test   F[  8,    43]           =    277.98326  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =    104.34671  Akaike I.C.   = -6.50506
             Restricted (b=0)        =      1.25792  Bayes  I.C.   = -6.16734
             Chi squared [  8]       =    206.17758  Prob C2 > C2* =   .00000
RESET test   Chi squared [  3]       =      2.69733  Prob C2 > C2* =   .44068
-----------------------------------------------------------------------------
(Regression results omitted)
-----------------------------------------------------------------------------
```

E7.9.3 Omitted Variables

A common application is examining the effect of including an additional variable in a regression after the regression is estimated without that variable. You may provide an additional set of variables with the following specification:

; Rh2 = other variable(s).

The usual regression is computed for '*y*' on the regressors. Then, for each variable in the Rh2 list, the following are computed for that variable, if it *alone* were added to the regression:

- What its coefficient would be
- What the new R^2 would be
- How much R^2 would increase
- Partial R^2 (squared correlation of *y* with this *x*, net of the effects of the other variables)
- Partial F statistic

The partial F statistic is the F ratio for the regression of *y* on this *x*, net of the included variables. This is the square of what would be the t ratio if this variable were included. Do note, this is not a means of carrying out a joint test of whether the group of variables would contribute significantly to the fit of the model. In order to carry out this test, you would use one of the procedures described in Section E8.2.

The following illustrates this analysis applied to the six price indexes used in the preceding examples:

```
REGRESS    ; Lhs = logg
           ; Rhs = one,logpg,loginc
           ; Rh2 = logpnc,logpuc,logppt,logpn,logpd,logps $
```

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=LOGG    Mean                 =        -.25713
            Standard deviation   =         .23849
            No. of observations  =             52  Degrees of freedom
Regression  Sum of Squares       =        2.72390           2
Residual    Sum of Squares       =        .176898          49
Total       Sum of Squares       =        2.90080          51
            Standard error of e  =         .06008
Fit         R-squared            =         .93902  R-bar squared =   .93653
Model test  F[  2,    49]        =      377.25464  Prob F > F*   =   .00000
Diagnostic  Log likelihood       =       73.98430  Akaike I.C.   = -5.56804
            Restricted (b=0)     =        1.25792  Bayes  I.C.   = -5.45547
            Chi squared [  2]    =      145.45276  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 08:16:15 AM
Effects of additional variables on the regression below: -------------
Variable Coefficient  New R-sqrd  Chg.R-sqrd   Partial-Rsq   Partial F
LOGPNC        -.3395       .9603       .0213         .3493      25.772
LOGPUC        -.1947       .9553       .0163         .2670      17.485
LOGPPT        -.2598       .9628       .0238         .3906      30.762
LOGPN         -.1941       .9452       .0062         .1016       5.426
LOGPD         -.4977       .9616       .0226         .3705      28.246
LOGPS         -.3656       .9678       .0288         .4718      42.883
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.99007***     .58201   -15.45  .0000   -10.13078   -7.84936
   LOGPG|    -.17124***     .03789    -4.52  .0000     -.24550    -.09698
  LOGINC|     .96865***     .07376    13.13  .0000      .82408    1.11322
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Note that the six Rh2 variables are collinear with the constant, so the full regression cannot be computed. But, the dummies could be added one at a time with no problem, as the results indicate.

E7.9.4 The CUSUM Test of Model Stability

Brown, Durbin, and Evans' (1972) CUSUM and CUSUM of squares tests are procedures for testing the stability of a model over time. The procedure is requested in *LIMDEP* by adding

; Cusum

to the **REGRESS** command. The procedure is based on the following: Let

$$e_t = y_t - \mathbf{x_t}'\mathbf{b}_{(t\text{-}1)}$$

where $\mathbf{b}_{(t\text{-}1)}$ is the least squares coefficient vector computed using all observations up to *but not including* $[y_t,\mathbf{x}_t]$. The set of scaled residuals

$$w_r = e_r / [1 + \mathbf{x}_r'(\mathbf{X}_{(r\text{-}1)}'\mathbf{X}_{(r\text{-}1)})^{-1}\mathbf{x}_r]^{1/2}$$

are independent and, assuming normality of the original disturbances, normally distributed with mean zero and variance

$$\sigma_{fr}^2 = \sigma^2[1 + \mathbf{x}_r'(\mathbf{X}_{(r\text{-}1)}'\mathbf{X}_{(r\text{-}1)})^{-1}\mathbf{x}_r].$$

The CUSUM test is based on the cumulated sum of residuals

$$W_t = \sum_{r=K+1}^{r=t} w_r \,/\, \hat{\sigma}$$

where
$$\hat{\sigma}^2 = \tfrac{1}{T-K-1}\sum_{r=K+1}^{r=T}(w_r - \overline{w})^2,$$

and
$$\overline{w} = \tfrac{1}{T-K}\sum_{r=K+1}^{r=T} w_r.$$

The series of values are plotted against time. Upper and lower confidence bounds are the lines connecting the points

$$\textit{Upper}:\ K + a(T\text{-}K)^{1/2} \text{ to } K + 3a(T\text{-}K)^{1/2}$$
$$\textit{Lower}:\ K - a(T\text{-}K)^{1/2} \text{ to } K - 3a(T\text{-}K)^{1/2}.$$

The values of '*a*' that correspond to various significance levels are given by the authors. Those for 95% and 99% significance are .948 and 1.143, respectively. *LIMDEP* uses the 95% significance point. The plot is obtained by plotting against

$$m = t\text{-}K = 1,\ldots,M = T\text{-}K$$

the three series W_m, *Upper* = .948 + 1.896*m*, and *Lower* = -*Upper*.

The CUSUM of squares test is based on

$$S_t = \frac{\sum_{r=K+1}^{r=t} w_r^2}{\sum_{r=K+1}^{r=T} w_r^2}$$

The expected value of S_t is approximately $(t\text{-}K)/(T\text{-}K)$. The series of values are also plotted against time. Upper and lower confidence limits are obtained by comparing S_t to $E[S_t] \pm c_0$, where c_0 is a function of both $(T\text{-}K)$ and the significance level. We use an approximation to obtain c_0:

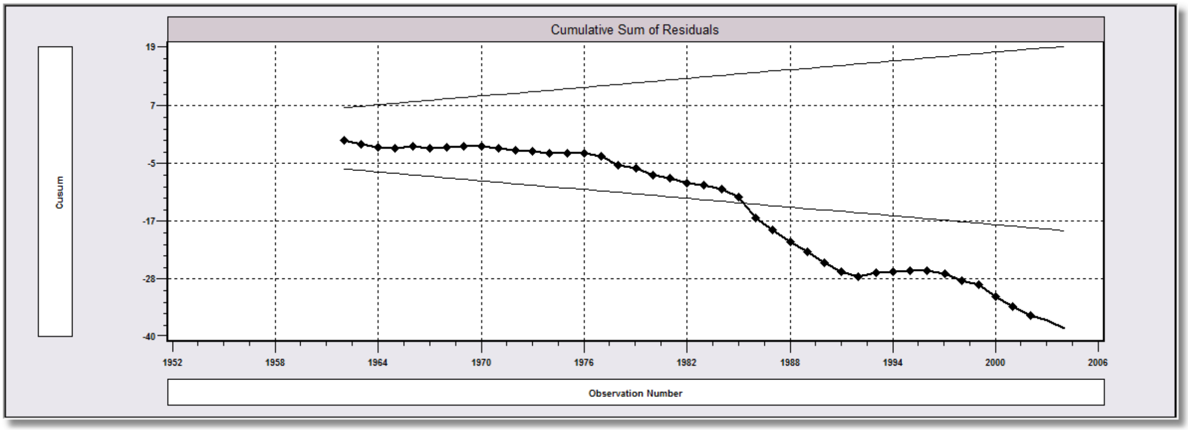
$$c_0 \approx .31 \ \text{ if } L < 10$$
$$c_0 \approx .594017 - .142897\log(L) + .00560574\log^2(L) + .000616272\log^3(L)$$
$$c_0 \approx .11 \ \text{ if } L > 100$$

where $L = (T\text{-}K)/2 - 1$. This produces an R^2 exceeding .9999 for the table of values provided by Harvey (1988).

The following example is the result of computing the CUSUM and CUSUM of squares tests as part of the basic regression computed above for the gasoline market data. The usual regression results are given first, followed by a listing of the CUSUM and CUSUM of squares values. Those outside the confidence limits are marked with an asterisk in the output.

```
DATE      ; 1953 $
PERIOD    ; 1953-2004 $
REGRESS   ; Lhs = logg
          ; Rhs = one,logpnc,logpuc,logppt,logpn,logpd,logps,logpg,loginc
          ; Cusum $
```

Period	Cusum	CusumSq	Period	Cusum	CusumSq	Period	Cusum	CusumSq
1962	-.473	.0029	1963	-1.131	.0086	1964	-1.874	.0158
1965	-1.945	.0159	1966	-1.697	.0167	1967	-2.000	.0179
1968	-1.895	.0180	1969	-1.506	.0200	1970	-1.601	.0201
1971	-2.034	.0226	1972	-2.313	.0236	1973	-2.542	.0243*
1974	-2.968	.0266*	1975	-2.917	.0267*	1976	-2.956	.0267*
1977	-3.661	.0332*	1978	-5.381	.0719*	1979	-6.071	.0781*
1980	-7.542	.1064*	1981	-8.090	.1103*	1982	-9.087	.1233*
1983	-9.500	.1255*	1984	-10.185	.1317*	1985	-11.901	.1702*
1986	-16.213*	.4133	1987	-18.633*	.4899	1988	-20.898*	.5569
1989	-22.908*	.6098	1990	-25.206*	.6789	1991	-26.971*	.7196
1992	-28.035*	.7344	1993	-27.298*	.7415	1994	-27.037*	.7424
1995	-26.764*	.7433	1996	-26.844*	.7434	1997	-27.496*	.7490
1998	-28.795*	.7710	1999	-29.653*	.7807	2000	-32.126*	.8606
2001	-34.075*	.9103	2002	-35.829*	.9505	2003	-36.844*	.9640
2004	-38.504*	1.0000						



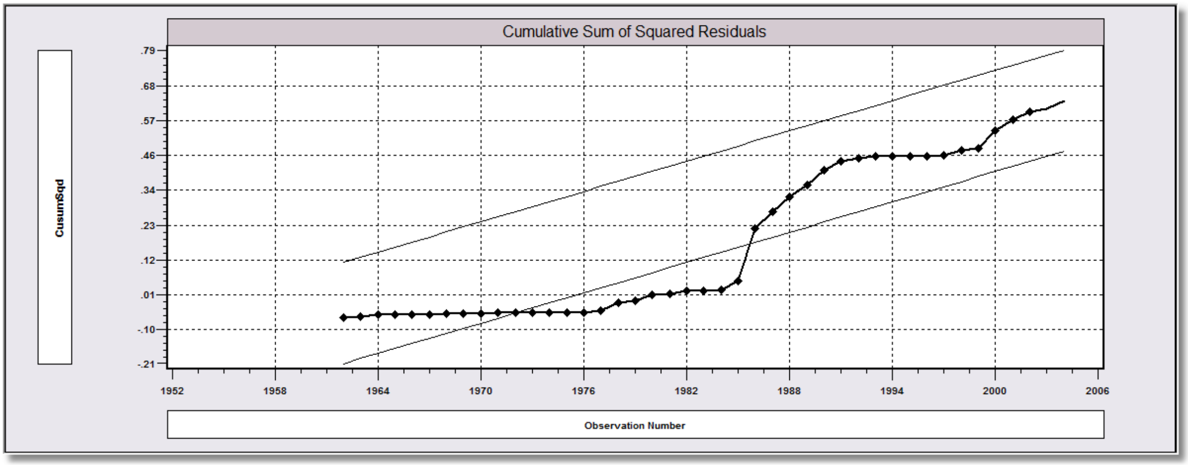


Figure E7.6 CUSUM and CUSUM of Squares Plots

E7.10 Robust Covariance Matrix Estimation

REGRESS will compute robust estimators for the covariance matrix of the least squares estimator for both heteroscedastic and autocorrelated disturbances. Although OLS is generally quite robust, some researchers have advocated other estimators for finite sample purposes. **REGRESS** can also be used to compute the least absolute deviations estimator.

Robust covariance matrix estimators are specified in the command line or selected from the Options page of the command builder, as shown below.

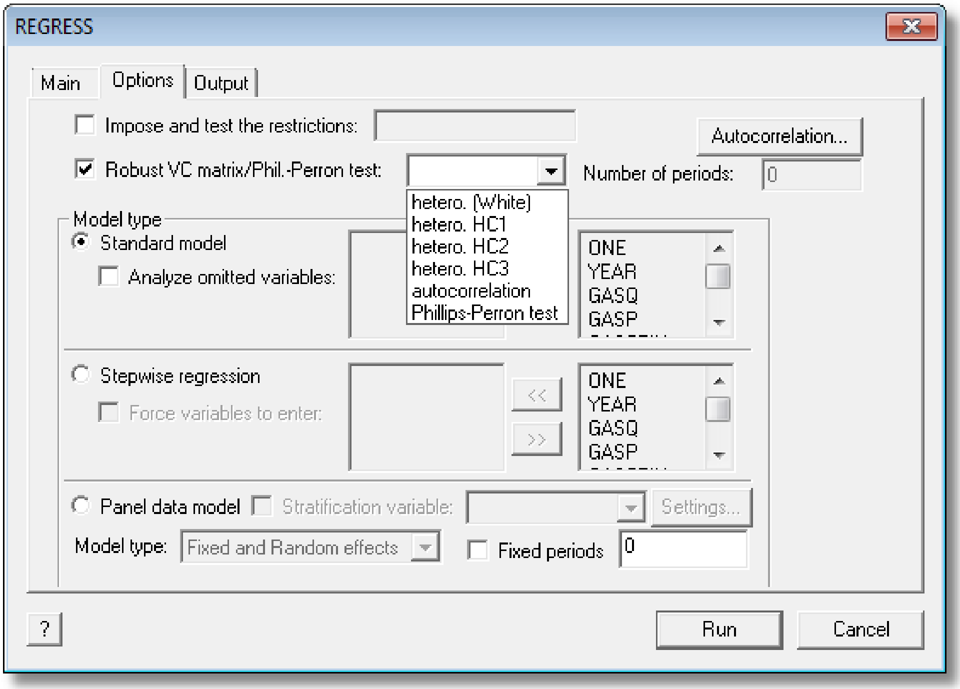


Figure E7.7 Command Builder for Robust Regression

E7.10.1 Heteroscedasticity – The White Estimator

For the heteroscedasticity corrected (White) estimator, use

; Heteroscedasticity

in the **REGRESS** command. The White estimator is

$$\text{Est.Var}[\mathbf{b}] \;=\; (\mathbf{X'X})^{-1} \times \sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i{}' \times (\mathbf{X'X})^{-1}$$

Davidson and MacKinnon (1993) have recommended three alternative forms of the estimator which appear to perform well in small to moderate sized samples. Use

; Het ; Hc1 to change e_i^2 to $ne_i^2/(n\text{-}K)$

; Het ; Hc2 to change e_i^2 to e_i^2 / $[1 - \mathbf{x}_i'(\mathbf{X'X})^{-1}\,\mathbf{x}_i]$

; Het ; Hc3 to change e_i^2 to e_i^2 / $[1 - \mathbf{x}_i'\,(\mathbf{X'X})^{-1}\,\mathbf{x}_i]^2$

To illustrate, we will use the first five firms in the widely used Grunfeld data set. (There are ten firms in the whole data set.) We will use these data in several examples to follow. (These data are also found in http://pages.stern.nyu.edu/~wgreene/Text/Edition7/TableF10-4.txt – Table F10-4 in the website for Greene (2011).) The following results based on the Grunfeld data show the results of ordinary least squares and the four different heteroscedasticity robust covariance matrix estimators.

REGRESS ; Lhs = i ; Rhs = one,f,c $

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -63.6112***    22.37624    -2.84  .0055   -107.4678  -19.7546
       F|      .11844***     .00948    12.49  .0000      .09985    .13703
       C|      .25648***     .03894     6.59  .0000      .18015    .33281
--------+--------------------------------------------------------------------
```

REGRESS ; Lhs = i ; Rhs = one,f,c ; Het $

```
--------+--------------------------------------------------------------------
Constant|   -63.6112***    21.62394    -2.94  .0041   -105.9933  -21.2290
       F|      .11844***     .00734    16.13  .0000      .10404    .13283
       C|      .25648***     .05318     4.82  .0000      .15225    .36070
--------+--------------------------------------------------------------------
```

REGRESS ; Lhs = i ; Rhs = one,f,c ; Het ; Hc1 $

```
--------+--------------------------------------------------------------------
Constant|   -63.6112***    21.95579    -2.90  .0047   -106.6437  -20.5786
       F|      .11844***     .00746    15.88  .0000      .10382    .13305
       C|      .25648***     .05399     4.75  .0000      .15065    .36230
--------+--------------------------------------------------------------------
```

REGRESS ; Lhs = i ; Rhs = one,f,c ; Het ; Hc2 $

```
--------+--------------------------------------------------------------------
Constant|   -63.6112***    23.93932    -2.66  .0092   -110.5314  -16.6910
       F|      .11844***     .00766    15.45  .0000      .10341    .13346
       C|      .25648***     .05950     4.31  .0000      .13985    .37310
--------+--------------------------------------------------------------------
```

REGRESS ; Lhs = i ; Rhs = one,f,c ; Het ; Hc3 $

```
--------+--------------------------------------------------------------------
Constant|   -63.6112**     26.72238    -2.38  .0192   -115.9861  -11.2363
       F|      .11844***     .00802    14.77  .0000      .10272    .13415
       C|      .25648***     .06705     3.83  .0002      .12506    .38790
--------+--------------------------------------------------------------------
```

E7.10.2 Autocorrelation – The Newey-West Estimator

The Newey-West robust estimator for the covariance matrix of the least squares estimator in the presence of autocorrelation is

$$\text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \sum\nolimits_{t=1}^{T} e_t^2 \mathbf{x}_t \mathbf{x}_t' \times (\mathbf{X}'\mathbf{X})^{-1}$$
$$+ (\mathbf{X}'\mathbf{X})^{-1} \times \left\{ \frac{1}{T} \sum\nolimits_{j=1}^{L} \sum\nolimits_{t=j+1}^{T} \left(1 - \frac{j}{L+1}\right) e_t e_{t-j} \left[\mathbf{x}_t \mathbf{x}_{t-j}' + \mathbf{x}_{t-j} \mathbf{x}_t'\right] \right\} \times (\mathbf{X}'\mathbf{X})^{-1}$$

You (the analyst) must provide the value of L, the number of lags for which the estimator is computed. Then, request this estimator by adding

; Pds = ... the value for *L*

to the **REGRESS** command. No finite sample improvement for this estimator has been devised, so there is no counterpart to the Davidson and MacKinnon variants for the heteroscedasticity estimator.

An application based on the preceding example for the gasoline market follows.

```
REGRESS   ; Lhs = logg
          ; Rhs = one,logpg,logy,logpnc,logpuc,logppt,logpn,logpd,logps
          ; Pds = 10 $
```

The uncorrected least squares results are shown below those based on the robust estimator.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23849
             Number of observs.   =            52
Model size   Parameters           =             9
             Degrees of freedom   =            43
Residuals    Sum of squares       =    .550250E-01
             Standard error of e  =        .03577
Fit          R-squared            =        .98103
             Adjusted R-squared   =        .97750
Model test   F[  8,    43] (prob) =   278.0(.0000)
Robust VC    Newey-West, Periods  =            10
Model was estimated on May 09, 2011 at 09:04:30 AM
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    LOGG|  Coefficient      Error       t    |t|>T*        Interval
--------+--------------------------------------------------------------------
Constant|   -18.1296***    3.04773    -5.95  .0000    -24.1031   -12.1562
   LOGPG|     .04124        .10120      .41  .6857    -.15711     .23959
  LOGINC|    1.91886***     .31175     6.16  .0000    1.30784    2.52988
  LOGPNC|     .42763        .45241      .95  .3498    -.45909    1.31434
  LOGPUC|    -.29824***     .09769    -3.05  .0039    -.48971    -.10678
  LOGPPT|     .15858        .11311     1.40  .1681    -.06312     .38027
   LOGPN|     .57537***     .19268     2.99  .0046     .19772     .95302
   LOGPD|    -.28216        .32880     -.86  .3956    -.92661     .36228
   LOGPS|    -.81314***     .28842    -2.82  .0072   -1.37843    -.24786
--------+--------------------------------------------------------------------
Uncorrected Least Squares Results
--------+--------------------------------------------------------------------
Constant|   -18.1296***    2.26108    -8.02  .0000    -22.5613   -13.6980
   LOGPG|     .04124        .06261      .66  .5136    -.08146     .16394
  LOGINC|    1.91886***     .21338     8.99  .0000    1.50065    2.33708
  LOGPNC|     .42763        .29421     1.45  .1533    -.14901    1.00427
  LOGPUC|    -.29824***     .09365    -3.18  .0027    -.48180    -.11469
  LOGPPT|     .15858        .15996      .99  .3271    -.15495     .47210
   LOGPN|     .57537***     .17445     3.30  .0020     .23346     .91729
   LOGPD|    -.28216        .31248     -.90  .3716    -.89461     .33028
   LOGPS|    -.81314*       .42566    -1.91  .0628   -1.64743     .02114
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E7.10.3 Clustering

An estimator which has become popular for data which are 'clustered' (loosely like a panel), and which accommodates some kinds of correlation within groups of observations is the cluster robust estimator,

$$\text{Est.Asy.Var}\left[\mathbf{b}\right] = s^2 \ (\mathbf{X'X})^{-1} \times \frac{C}{C-1}\sum\nolimits_{c=1}^{C} \mathbf{g}_c\mathbf{g}_c{}' \times s^2 \ (\mathbf{X'X})^{-1}$$

where C is the number of clusters, n_c is the number of observations in a particular cluster, 'ic' indicates observation i in cluster c, and

$$\mathbf{g}_c = \sum\nolimits_{i=1}^{n_c} \frac{e_{ic}}{s^2}\mathbf{x}_{ic}$$

Note that $\mathbf{g}_c$ is a derivative from the normal likelihood function. (If there is one observation in each group, then this is the $(n/(n\text{-}1))$ times the White estimator.) This is requested with

; Cluster = specification.

The specification is either the fixed group size, or the name of a variable which gives the group a particular identifier – i.e., a stratification variable, such as a group number, firm number, country identifier, etc. The following applies to the Grunfeld data used earlier.

```
REGRESS    ; Lhs = i
           ; Rhs = one,f,c
           ; Cluster = firm $
```

```
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of    100 observations contained      5 clusters defined by  |
| variable FIRM     which identifies by a value a cluster ID.         |
+---------------------------------------------------------------------+

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                    =     253.74220
             Standard deviation      =     265.19714
             No. of observations     =           100  Degrees of freedom
Regression   Sum of Squares          =    .529952E+07            2
Residual     Sum of Squares          =    .166310E+07           97
Total        Sum of Squares          =    .696262E+07           99
             Standard error of e     =     130.94025
Fit          R-squared               =        .76114  R-bar squared =   .75621
Model test   F[  2,    97]           =     154.54699  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =    -627.84500  Akaike I.C.   = 9.77902
             Restricted (b=0)        =    -699.43868  Bayes  I.C.   = 9.85718
             Chi squared [  2]       =     143.18736  Prob C2 > C2* =   .00000
B-P test     Chi squared [  2]       =      12.14252  Prob C2 > C2* =   .00231
Model was estimated on May 09, 2011 at 09:11:45 AM
```

```
--------+-------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
       I|  Coefficient       Error       t     |t|>T*         Interval
--------+-------------------------------------------------------------------
Constant|   -63.6112       56.07716    -1.13  .2594   -173.5204   46.2980
       F|     .11844***      .01194     9.92  .0000      .09504    .14184
       C|     .25648***      .09299     2.76  .0070      .07421    .43874
--------+-------------------------------------------------------------------
These are the uncorrected results.
--------+-------------------------------------------------------------------
Constant|   -63.6112***    22.37624    -2.84  .0055   -107.4678  -19.7546
       F|     .11844***      .00948    12.49  .0000      .09985    .13703
       C|     .25648***      .03894     6.59  .0000      .18015    .33281
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

An extension of this robust covariance matrix estimator allows for two level, stratified and clustered data sets. Use

; Stratum = the specification

The specification provides either the fixed number of observations in a stratum or a variable that provides the identifier for strata in the data. Specifics on the use of **; Stratum** for complex survey data appear in Section R10.3.

E7.11 Accuracy in Linear Regression – NIST Benchmarks

Continuing the analysis in Chapter E2, we now examine the accuracy of the linear regression routine in *LIMDEP*. The NIST/StRD benchmarks are a suite of eleven data sets and linear regression problems designed to test the accuracy of linear regression programs. McCullough (1998) has analyzed various algorithms and programs with these data sets, and tabulated the highest accuracy he could achieve with three algorithms, Cholesky inversion, *QR* decomposition, and singular value decomposition (SVD). *QR*, the method used in *LIMDEP* is the most accurate in nine of eleven runs. The following displays full results for a few of the NIST linear regression benchmarks. The first, the Norris data, is a low level, fairly simple test. The Longley set is a moderately difficult test, but is a de facto benchmark which no respectable commercial package should fail. The Wampler (5) test is known to be one of the most difficult of the standard benchmarks. The Filippelli data are the most difficult data; this regression is not computable with many packages, and, for example, does not solve at all, if one is using Cholesky or any other direct inversion method. As shown below, *LIMDEP* achieves high accuracy on all of these problems, including the Filippelli problem.

The data sets are primarily structured to test two features of the solver, its ability to solve a problem when the data are highly collinear and its ability to handle a data set with widely differing orders of magnitude. The Filippelli data are the most difficult of the first of these. The Wampler data sets test the second. Note that in spite of the huge condition numbers reported for some of these data matrices, the solutions all agree closely with the benchmarks.

Many of the NIST datasets and test programs are included with the *LIMDEP* program, and can be found in the C:\LIMDEP10\Command Files folder created with program installation and also in the NIST Benchmarks book of the Help file. (The initial statement of each problem is the verbatim text of the NIST/StRD posting on their website.)

```
File Name:      NIST-Regression-Norris11.lim
Dataset Name:   Norris (NIST-Norris11.dat)
Procedure:      Linear Least Squares Regression
Reference:      Norris, J., NIST.
                Calibration of Ozone Monitors.
Data:           1 Response Variable (y)
                1 Predictor Variable (x)
                36 Observations
                Lower Level of Difficulty
                Observed Data
Model:          Linear Class
                2 Parameters (B0,B1)
                y = B0 + B1*x + e
                Certified Regression Statistics
                                           Standard Deviation
     Parameter          Estimate              of Estimate
        B0        -0.262323073774029     0.232818234301152
        B1         1.00211681802045      0.429796848199937E-03
     Residual
     Standard Deviation   0.884796396144373
     R-Squared            0.999993745883712
                Certified Analysis of Variance Table
Source of Degrees of    Sums of             Mean
Variation  Freedom      Squares            Squares           F Statistic
Regression    1     4255954.13232369   4255954.13232369   5436385.54079785
Residual     34     26.6173985294224   0.782864662630069
```

```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                 =      419.80278
             Standard deviation   =      348.71113
             No. of observations  =             36  Degrees of freedom
Regression   Sum of Squares       =    .425595E+07           1
Residual     Sum of Squares       =        26.6174          34
Total        Sum of Squares       =    .425598E+07          35
             Standard error of e  =         .88480
Fit          R-squared            =         .99999  R-bar squared =   .99999
Model test   F[  1,    34]        =  5436385.54078  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      -45.64662  Akaike I.C.   =  -.19084
             Restricted (b=0)     =     -261.32749  Bayes  I.C.   =  -.10287
             Chi squared [  1]    =      431.36175  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 10:14:01 AM
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       t    |t|>T*         Interval
--------+-------------------------------------------------------------------
Constant|    -.26232         .23282    -1.13  .2677     -.71864    .19399
       X|    1.00212***      .00043  2331.61  .0000     1.00127   1.00296
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

MATRIX ; Peek ; b $

```
Display of all internal digits of matrix B
B[0001] = -.26232307377404140D+00
B[0002] = .10021168180204550D+01
```

```
File Name:      NIST-Regression-Wampler5.lim
Dataset Name:   Wampler-5 (NIST-Wampler5.dat)
Procedure:      Linear Least Squares Regression
Reference:      Wampler, R. H. (1970).
                A Report of the Accuracy of Some Widely-Used Least
                Squares Computer Programs.
                Journal of the American Statistical Association, 65, pp. 549-565.
Data:           1 Response Variable (y)
                1 Predictor Variable (x)
                21 Observations
                Higher Level of Difficulty
                Generated Data
Model:          Polynomial Class
                6 Parameters (B0,B1,...,B5)
                y = B0 + B1*x + B2*(x**2) + B3*(x**3)+ B4*(x**4) + B5*(x**5)
                Certified Regression Statistics
                                           Standard Deviation
     Parameter          Estimate              of Estimate

        B0        1.00000000000000         21523262.4678170
        B1        1.00000000000000         23635517.3469681
        B2        1.00000000000000         7793435.24331583
        B3        1.00000000000000         1014755.07550350
        B4        1.00000000000000         56456.6512170752
        B5        1.00000000000000         1123.24854679312
     Residual
     Standard Deviation   23601450.2379268
     R-Squared            0.224668921574940E-02
                Certified Analysis of Variance Table
Source of Degrees of    Sums of                 Mean
Variation  Freedom      Squares                Squares           F Statistic
Regression    5    18814317208116.7      3762863441623.33 6.7552445824012241E-03
Residual     15    0.835542680000000E+16 557028453333333.

WARNING: Badly conditioned X. Condition value =    .4330261D+07
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                    =   623960.33333
             Standard deviation      = 20462454.78579
             No. of observations     =             21  Degrees of freedom
Regression   Sum of Squares          =    .188143E+14            5
Residual     Sum of Squares          =    .835543E+16           15
Total        Sum of Squares          =    .837424E+16           20
             Standard error of e     = 23601450.23793
Fit          R-squared               =         .00225  R-bar squared =  -.33034
Model test   F[  5,    15]           =         .00676  Prob F > F*   =   .99999
Diagnostic   Log likelihood          =     -382.77794  Akaike I.C.   = 34.18859
             Restricted (b=0)        =     -382.80156  Bayes  I.C.   = 34.48703
             Chi squared [  5]       =         .04723  Prob C2 > C2* =   .99997
Model was estimated on May 09, 2011 at 10:17:45 AM
```

```
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       t     |t|>T*         Interval
--------+-------------------------------------------------------------------
Constant|        1.0       .2152D+08      .00 1.0000 -.42185D+08  .42185D+08
      X1|    1.00000       .2364D+08      .00 1.0000 ***********  ***********
      X2|        1.0       .7793D+07      .00 1.0000 -.15275D+08  .15275D+08
      X3|    1.00000       .1015D+07      .00 1.0000 ***********  ***********
      X4|        1.0        56456.65      .00 1.0000 -.11065D+06  .11065D+06
      X5|    1.00000        1123.249      .00  .9993 -2200.52670  2202.52670
--------+-------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

MATRIX ; Peek ; b $

```
Display of all internal digits of matrix B
B[0001] = .10000000890577210D+01
B[0002] = .99999982105423900D+00
B[0003] = .10000000650753040D+01
B[0004] = .99999999135189290D+00
B[0005] = .10000000004789460D+01
B[0006] = .99999999999061120D+00
```

```
File Name:     NIST-Regression-Filippelli.lim
Dataset Name:  Filippelli (NIST-Filippelli.dat)
Procedure:     Linear Least Squares Regression
Reference:     Filippelli, A., NIST.
Data:          1 Response Variable (y)
               1 Predictor Variable (x)
               82 Observations
               Higher Level of Difficulty
               Observed Data
Model:         Polynomial Class
               11 Parameters (B0,B1,...,B10)
               y = B0 + B1*x + B2*(x**2) + ... + B9*(x**9) + B10*(x**10) + e
               Certified Regression Statistics
                                                 Standard Deviation
     Parameter         Estimate                     of Estimate
        B0       -1467.48961422980         298.084530995537
        B1       -2772.17959193342         559.779865474950
        B2       -2316.37108160893         466.477572127796
        B3       -1127.97394098372         227.204274477751
        B4       -354.478233703349         71.6478660875927
        B5       -75.1242017393757         15.2897178747400
        B6       -10.8753180355343         2.23691159816033
        B7       -1.06221498588947         0.221624321934227
        B8       -0.670191154593408E-01    0.142363763154724E-01
        B9       -0.246781078275479E-02    0.535617408889821E-03
        B10      -0.402962525080404E-04    0.896632837373868E-05

     Residual
     Standard Deviation   0.334801051324544E-02
     R-Squared            0.996727416185620
               Certified Analysis of Variance Table
Source of Degrees of     Sums of                     Mean
Variation  Freedom        Squares                   Squares              F Statistic
Regression   10     0.242391619837339     0.242391619837339E-01 2162.43954511489
Residual     71     0.795851382172941E-03 0.112091743968020E-04
```

```
WARNING: Badly conditioned X. Condition value =    .2999482D+10
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                 =          .84958
             Standard deviation   =          .05479
             No. of observations  =              82  Degrees of freedom
Regression   Sum of Squares       =         .242392          10
Residual     Sum of Squares       =     .795851E-03          71
Total        Sum of Squares       =         .243187          81
             Standard error of e  =          .00335
Fit          R-squared            =          .99673  R-bar squared =   .99627
Model test   F[ 10,    71]        =      2162.43959  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =       356.90255  Akaike I.C.   =-11.27452
             Restricted (b=0)     =       122.29336  Bayes  I.C.   =-10.95167
             Chi squared [ 10]    =       469.21839  Prob C2 > C2* =   .00000
Model was estimated on May 09, 2011 at 10:21:47 AM
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient      Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -1467.49***    298.0845    -4.92  .0000    -2051.72   -883.25
      X1|   -2772.18***    559.7799    -4.95  .0000    -3869.33  -1675.03
      X2|   -2316.37***    466.4776    -4.97  .0000    -3230.65  -1402.09
      X3|   -1127.97***    227.2043    -4.96  .0000    -1573.29   -682.66
      X4|   -354.478***    71.64787    -4.95  .0000    -494.905  -214.051
      X5|   -75.1242***    15.28972    -4.91  .0000   -105.0915  -45.1569
      X6|   -10.8753***     2.23691    -4.86  .0000    -15.2596   -6.4911
      X7|   -1.06222***      .22162    -4.79  .0000    -1.49659   -.62784
      X8|    -.06702***      .01424    -4.71  .0000     -.09492   -.03912
      X9|    -.00247***      .00054    -4.61  .0000     -.00352   -.00142
     X10|-.40296D-04***   .8966D-05    -4.49  .0000 -.57870D-04  -.22723D-04
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

MATRIX ; Peek ; b $

```
Display of all internal digits of matrix B
B[0001] =-.14674896451936570D+04
B[0002] =-.27721796507612570D+04
B[0003] =-.23163711311320980D+04
B[0004] =-.11279739653118660D+04
B[0005] =-.35447824142799430D+03
B[0006] =-.75124203396297770D+02
B[0007] =-.10875318278760100D+02
B[0008] =-.10622150100252390D+01
B[0009] =-.67019117009404680D-01
B[0010] =-.24678108409567800D-02
B[0011] =-.40296253478694550D-04
```

E8: Linear Regression – Hypothesis Tests and Restrictions

E8.1 Introduction

This chapter will detail hypothesis testing and restricted estimation in the single equation, linear regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{iK}\beta_K + \varepsilon_i$$
$$= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i,\ i = 1,\ldots,n.$$

The full set of observations is denoted for present purposes as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

The initial stochastic assumptions are the most restrictive for the linear model:

$$E[\varepsilon_i | \mathbf{X}] = 0 = E[\varepsilon_i]\ \forall\ i \quad \text{(zero mean)}$$
$$\text{Var}[\varepsilon_i | \mathbf{X}] = \text{Var}[\varepsilon_i] = \sigma^{2,}\ \forall\ i \quad \text{(homoscedastic)}$$
$$\text{Cov}[\varepsilon_i, \varepsilon_j | \mathbf{X}] = \text{Cov}[\varepsilon_i, \varepsilon_j] = 0\ \forall\ i,j \quad \text{(nonautocorrelation).}$$

Estimation of β and σ^2 and computation of appropriate standard errors were detailed in Chapter E7. This chapter will present methods of testing hypotheses about coefficients and how to estimate the regression model subject to restrictions on the coefficients.

E8.2 Hypothesis Tests in the Linear Regression Model

There are several built in procedures for inference. In addition, the **REGRESS** and **MATRIX** commands can be used to test a variety of hypotheses. For purposes of a running example, we will use the Christensen and Greene (1976) electricity data, which are in Table F4-4 in Greene (2011) – http://pages.stern.nyu.edu/~wgreene/Text/Edition7/TableF4-4.txt. The data are set up with

```
IMPORT $
id   year   cost    q     pl        sl      pk       sk      pf        sf
1    1970   .2130   8.0   6869.470  .3291   64.945   .4197   18.0000   .2512
157 additional observations.  Only the first 123 are used in the study.
SAMPLE      ; 1-123 $
CREATE      ; lnpk = Log(pk) ; lnpl = Log(pl) ; lnpf = Log(pf) $
CREATE      ; lncost = Log(cost) ; lnq = Log(q) ; lnqsq = lnq*lnq $
CREATE      ; lnpk_pf = Log(pk/pf) ; lnlp_pf = Log(pl/pf) ; lncostpf = Log(cost/pf)$
```

E8.2.1 Testing Significance of Individual Coefficients

The standard regression results contain the results of hypothesis tests that individual coefficients are equal to zero. The following results illustrate.

REGRESS ; Lhs = lncost ; Rhs = one,lnpk,lnpl,lnpf,lnq $

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LNCOST   Mean                    =       3.00917
             Standard deviation      =       1.56241
             No. of observations     =           123  Degrees of freedom
Regression   Sum of Squares          =       292.338           4
Residual     Sum of Squares          =       5.47915         118
Total        Sum of Squares          =       297.817         122
             Standard error of e     =        .21548
Fit          R-squared               =        .98160  R-bar squared =   .98098
Model test   F[  4,   118]           =    1573.95986  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =      16.81149  Akaike I.C.   = -3.02993
             Restricted (b=0)        =    -228.91353  Bayes  I.C.   = -2.91562
             Chi squared [  4]       =     491.45003  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LNCOST|  Coefficient       Error       t     |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.06007***     1.30600    -6.17  .0000   -10.61979  -5.50035
    LNPK|     .17131         .13498     1.27  .2069     -.09324    .43586
    LNPL|     .12860         .13233      .97  .3331     -.13076    .38796
    LNPF|     .70487***      .07506     9.39  .0000      .55776    .85197
     LNQ|     .83024***      .01095    75.85  .0000      .80879    .85169
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The statistical significance (10%, 5%, 1%) is indicated in the table. The test is based on the t statistic, the P value or whether the confidence interval contains zero. All produce the same conclusion. A test of whether a coefficient equals some particular nonzero value can, in principle, be carried out using the procedure in the next section. However, the same significance test results just by assessing whether the confidence interval contains the indicated value. For example, based on the results below, they hypothesis that β_3 (the coefficient on *lnpl*) equals 0.5 would be rejected because the interval (-0.13076,0.38796) does not contain 0.5. By the same construction, the hypothesis that β_5 (the coefficient on *lnq*) equals 0.84 would not be rejected, as 0.84 is contained in (0.80879,0.85169).

E8.2.2 Linear Function of Coefficients

A test of a hypothesis based on a linear function of the coefficients is requested by adding

; Test: value * name ± value * name ± … = value

to the **REGRESS** command. When value equals one it (and the *) may be omitted. 'Name' is the name of a variable in the equation. For example, a restriction on the model parameters that is normally imposed as part of the cost function model is that the log price coefficients sum to one. We can test that as a hypothesis here with

REGRESS ; Lhs = lncost ; Rhs = one,lnpk,lnpl,lnpf,lnq
; Test: lnpk + lnpl + lnpf = 1 $

This produces the results below

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LNCOST   Mean                    =       3.00917
             Standard deviation      =       1.56241
             No. of observations     =           123  Degrees of freedom
Regression   Sum of Squares          =       292.338           4
Residual     Sum of Squares          =       5.47915         118
Total        Sum of Squares          =       297.817         122
             Standard error of e     =        .21548
Fit          R-squared               =        .98160  R-bar squared =   .98098
Model test   F[  4,    118]          =    1573.95986  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =      16.81149  Akaike I.C.   = -3.02993
             Restricted (b=0)        =    -228.91353  Bayes  I.C.   = -2.91562
             Chi squared [  4]       =     491.45003  Prob C2 > C2* =   .00000
Wald Test:   Chi-squared [  1]       =        .00061  Prob C2 > C2* =   .98025
F Test:      F ratio [ 1, 118]       =        .00061  Prof F  > F*  =   .98030
--------+--------------------------------------------------------------------
```

(The estimates are the same.) The test statistic is reported in the diagnostic results above the estimates. When the model is fit by least squares with no modifications to accommodate nonnormality or other violations of the standard assumptions (such as a robust covariance matrix), then two versions of the statistic are presented. The standard F statistic for testing the J restrictions,

$$H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$$

is

$$F[J,n-K] = [\mathbf{Rb} - \mathbf{q}]'[s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{Rb} - \mathbf{q}]/J$$
$$= [(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e})/J] \,/\, [\mathbf{e}'\mathbf{e}/(n-K)]$$

Under the assumption of normality of disturbances, this is distributed as $F[J,n-K]$ under the null hypothesis. The critical value is taken from the standard F table. When the denominator degrees of freedom are larger than 10,000, this critical value will be indistinguishable from 1/J times the counterpart from the chi squared, so in this case, only the Wald statistic is reported. For a linear regression the Wald chi squared statistic is exactly $J \times F$. The critical value for the Wald statistic is based on large sample results whereas that for the F statistic is based on the actual sample size (degrees of freedom). The result is that the P value for the Wald test will always be lower than that for the F value – the F statistic is more conservative.

NOTE: This syntax for testing restrictions is new with Version 10 of *LIMDEP*. Earlier versions would use **; Test: value * b(index) ± … ± value * b(index) = value**. The example above would be b(2) + b(3) + b(4) = 1. The older syntax is still usable (and even necessary on occasion, as shown in the next section). The syntax based on variable names rather than index positions will be more convenient, as the terms in the restriction are independent of the order or position of variables in the model specification.

E8.2.3 Linear Function with Interaction Terms and Nonlinearities

A slight change in the command is needed if the model contains interaction terms or nonlinear functions of the variables. For example, the model

REGRESS ; Lhs = lncost ; Rhs = one,lnpk,lnpl,lnpf,lnq,lnq*lnq $

contains a quadratic term in log output. To test a hypothesis about terms such as this, it is necessary to revert to the earlier form of restrictions. For example, to test the hypothesis that the coefficient on *lnq***lnq* equals zero, the syntax

; Test: lnq*lnq = 0

will not work; *lnq***lnq* is not the name of a variable, and *lnq***lnq* looks deceptively like, say, 2**lnq* to the program. The solution is to use the earlier format. For this simple hypothesis, we must use

; Test: b(6) = 0.

It is permissible to mix the two forms. For example, to test the two previous hypotheses at the same time, we would use

; Test: lnpk + lnpl + lnpf = 1, b(6) = 0

which produces

```
-----------------------------------------------------------------------------
... (Results omitted)

Wald Test:   Chi-squared [  2]    =      145.56748  Prob C2 > C2* =   .00000
F Test:      F ratio [ 2, 117]    =       72.78374  Prof F  > F*  =   .00000
--------+--------------------------------------------------------------------
```

E8.2.4 More Than One Linear Restriction

As shown in the example immediately above, when you have more than one restriction to test at the same time, you separate the restrictions with commas. To carry out separate tests, each of which can involve more than one restriction, separate the hypotheses with a '|' character. For example, to test the two restrictions above as separate hypotheses, rather than as one joint hypothesis, we would use

; Test: lnpk + lnpl + lnpf = 1 | b(6) = 0.

The results would be as follows:

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LNCOST   Mean                 =       3.00917
             Standard deviation   =       1.56241
             No. of observations  =           123  Degrees of freedom
Regression   Sum of Squares       =       295.375           5
Residual     Sum of Squares       =       2.44152         117
Total        Sum of Squares       =       297.817         122
             Standard error of e  =        .14446
Fit          R-squared            =        .99180  R-bar squared =   .99145
Model test   F[  5,   117]        =    2830.93377  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      66.52372  Akaike I.C.   = -3.82200
             Restricted (b=0)     =    -228.91353  Bayes  I.C.   = -3.68482
             Chi squared [  5]    =     590.87450  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LNCOST|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -7.04431***      .87956    -8.01  .0000    -8.76821  -5.32041
    LNPK|     .05419         .09100      .60  .5527     -.12418    .23255
    LNPL|     .24309***      .08922     2.72  .0074      .06823    .41795
    LNPF|     .66279***      .05044    13.14  .0000      .56394    .76164
     LNQ|     .39105***      .03713    10.53  .0000      .31827    .46383
 LNQ*LNQ|     .03123***      .00259    12.07  .0000      .02616    .03630
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Chi squared tests of linear restrictions. Degrees of freedom shown
in [.]. Equals zero is implied if no specific value was given.
 1. Restriction:LNPK+LNPL+LNPF=1
    Chi squared[ 1] =         .095, P value =  .7575
 2. Restriction:B(6)=0
    Chi squared[ 1] =      145.566, P value =  .0000
-----------------------------------------------------------------------------
```

E8.2.5 Testing Nonlinear Restrictions

Chapter R14 describes how to use the **WALD** command for testing and analyzing nonlinear restrictions and nonlinear functions of parameters. The procedure described there applies to the linear regression model as well as the others, so we need not add new details here. We demonstrate the use of the feature with a simple example.

Consider the equation based, once again, on the gasoline market data:

$$logg = \beta_1 + \beta_2 logpg + \beta_3 logy + \beta_4 logpnc + \beta_5 logpuc + \beta_6 logppt + \varepsilon.$$

Consider the nonlinear hypothesis

$$H_0: \beta_2/\beta_4 + \beta_2/\beta_5 = 0.$$

The following could be used to test this (admittedly meaningless) hypothesis:

NAMELIST ; x = one,logpg,logy,logpnc,logpuc,logppt $
REGRESS ; Lhs = logg ; Rhs = x $
WALD ; Fn1 = b_logpg / b_logpnc + b_logpg / b_logpuc $

Note that the restriction is implicitly **; Fn1 = … = 0**. The '**= 0**' may be omitted. Of course, if some other constant is needed, it must be included in the form **; Fn1 = … - value**. For example,

; Fn1 = b_K + b_L + b_F - 1

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23849
             No. of observations  =            52  Degrees of freedom
Regression   Sum of Squares       =       2.79379           5
Residual     Sum of Squares       =       .107004          46
Total        Sum of Squares       =       2.90080          51
             Standard error of e  =        .04823
Fit          R-squared            =        .96311  R-bar squared =   .95910
Model test   F[  5,    46]        =     240.20584  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      87.05475  Akaike I.C.   = -5.95537
             Restricted (b=0)     =       1.25792  Bayes  I.C.   = -5.73022
             Chi squared [   5]   =     171.59365  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*          Interval
--------+--------------------------------------------------------------------
Constant|   -11.5997***    1.48817     -7.79  .0000     -14.5165   -8.6829
   LOGPG|    -.03438        .04202      -.82  .4174      -.11673    .04797
  LOGINC|    1.31597***     .14198      9.27  .0000      1.03769   1.59425
  LOGPNC|    -.11964        .20384      -.59  .5601      -.51916    .27989
  LOGPUC|     .03754        .09814       .38  .7038      -.15481    .22990
  LOGPPT|    -.21514*       .11656     -1.85  .0714      -.44359    .01331
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The chi squared statistic for carrying out the test is given in the information at the top of the results, as shown below.

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic              =      .08431
Prob. from Chi-squared[ 1] =      .77154
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
WaldFcns|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
 Fncn(1)|    -.62840       2.16422     -.29  .7715    -4.87019   3.61338
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Although the difference is large, the standard error is extremely large, and the Wald test fails to reject the hypothesis.

When necessary, you can use the more general form of the **WALD** command to provide the parameters, covariance matrix and labels for the test. The preceding could also be obtained with

```
WALD      ; Parameters = b
          ; Covariance = varb
          ; Labels = 6_b
          ; Fn1 = b2/b4 + b2/b5 $
```

If more than one nonlinear function is specified in the **WALD** command, the overall chi squared given is used to test whether all of the functions equal to zero at the same time. The individual results given in the table can be used to test whether the individual functions equal zero. The following example proposes three 'hypotheses'

```
WALD      ; Parameters = b
          ; Covariance = varb
          ; Labels = 6_b
          ; Fn1 = b2/b4 + b2/b5
          ; Fn2 = b3
          ; Fn3 = b6 $
```

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic              =   327.46324
Prob. from Chi-squared[ 3] =      .00000
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
WaldFcns|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
 Fncn(1)|    -.62840       2.16422     -.29  .7715    -4.87019   3.61338
 Fncn(2)|    1.31597***     .14198     9.27  .0000     1.03769   1.59425
 Fncn(3)|    -.21514*       .11656    -1.85  .0649     -.44359    .01331
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The overall chi squared tests all three hypotheses at the same time. The individual results would be used to test the restrictions individually.

TIP: If you are testing a hypothesis that a function equals something other than zero, just subtract that value from the function specification. For example, **; Fn1 = b1 + b2 + b3 - 1** might be appropriate.

E8.2.6 Tests of Structural Change

We consider again the regression of gasoline consumption on price, income and three related price indexes. We are interested in testing the hypothesis that there is a structural break in 1974. The following does the standard Chow test. The **CALC** command then does a likelihood ratio test of the same hypothesis. Finally, we use a Wald test. The Wald test differs from the Chow test in that it allows for the disturbance variance to change across the periods.

```
DATES      ; 1953 $
PERIOD     ; 1953 - 2004 $
CREATE     ; d = Ind(1974,2004) ; dp = d*logpg
           ; dy = d*logy ; dpnc = d*logpnc
           ; dpuc = d*logpuc ; dppt = d*logppt $
NAMELIST   ; x = one,logpg,logy,logpnc,logpuc,logppt
           ; xd = d,dp,dy,dpnc,dpuc,dppt $
REGRESS    ; Lhs = logg ; Rhs = x,xd
           ; Test: d = 0, dp = 0, dy = 0, dpnc = 0, dpuc = 0, dppt = 0 $
CALC       ; List ; c = 2* (Lik(x,xd,logg) - Lik(x,logg))
           ; lrtest = 1 - Chi(c,(Col(xd))) $
```

The following shows the unconstrained and constrained regressions. (A few lines are omitted from the results.) Both the F and likelihood ratio statistics reject the null hypothesis of no structural change. The world did change in 1973-1974.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =      -.25713
             Standard deviation   =       .23849
             No. of observations  =           52  Degrees of freedom
Regression   Sum of Squares       =      2.89363          11
Residual     Sum of Squares       =   .716442E-02         40
Total        Sum of Squares       =      2.90080          51
             Standard error of e  =       .01338
Fit          R-squared            =       .99753  R-bar squared =   .99685
Model test   F[ 11,    40]        =   1468.68819  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    157.35187  Akaike I.C.   = -8.42833
             Restricted (b=0)     =      1.25792  Bayes  I.C.   = -7.97805
             Chi squared [ 11]    =    312.18791  Prob C2 > C2* =   .00000
Model was estimated on May 12, 2011 at 08:32:09 PM
Wald Test:   Chi-squared [  6]    =    557.41682  Prob C2 > C2* =   .00000
F Test:      F ratio [ 6,  40]    =     92.90280  Prof F  > F*  =   .00000
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t     |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -12.0786***     1.96984    -6.13  .0000    -15.9394   -8.2177
   LOGPG|      .06832        .18643      .37  .7159     -.29707    .43371
  LOGINC|      .98085***     .21124     4.64  .0000      .56683   1.39488
  LOGPNC|      .70430***     .24651     2.86  .0068      .22114   1.18746
  LOGPUC|    -.21467***      .07542    -2.85  .0069     -.36249   -.06685
  LOGPPT|      .06275        .11682      .54  .5941     -.16622    .29172
       D|     5.13886**     2.05052     2.51  .0164     1.11991   9.15780
      DP|    -.17914         .18708     -.96  .3440     -.54581    .18753
      DY|    -.23873         .21941    -1.09  .2831     -.66876    .19130
    DPNC|    -.58783**       .25676    -2.29  .0274    -1.09106   -.08460
    DPUC|      .22101**      .08504     2.60  .0130      .05433    .38769
    DPPT|    -.19422         .12615    -1.54  .1315     -.44146    .05302
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
[CALC] C       =    140.5942491
[CALC] LRTEST  =       .0000000
Calculator: Computed   2 scalar results
```

This command set carries out the Wald test. The last line lists the critical value for the chi squared distribution. The table value is about 12 while the statistic is over 582, so again, the hypothesis is rejected. Note that the earlier F statistic was 92.9. The number of restrictions times F, 6×92.9 = 557.4 should roughly equal the chi squared value (582.46), which it does.

```
REGRESS    ; If[year < 1974]   ; Quietly ; Lhs = logg ; Rhs = x $
MATRIX     ; b1 = b ; v1 = varb $
REGRESS    ; If[year > = 1974] ; Quietly ; Lhs = logg ; Rhs = x $
MATRIX     ; b2 = b ; v2 = varb $
MATRIX     ; db = b1 - b2 ; vdb = v1 + v2 ; List ; Wald = db'<vdb>db $
CALC       ; List ; Ctb(.95,6) $
CALC       ; List ; Ctb(.95,6) $
```

```
    WALD|             1
--------+--------------
       1|       582.461
 [CALC] *Result*=     12.5915872
```

There is a convenient, automated method of carrying out the Chow test when the hypothesis involves dividing the sample into two subsamples. The following carries out the same Wald test as shown above. *Post*1973 is a dummy variable that partitions the sample. (You might do this in a cross section, for example, to compare men and women, or two countries.) The **REGRESS** command begins with a loop that computes the regression (or any model) three times, with the full sample and with the two subsamples. The **DECOMPOSE** command then produces several results, including the chi squared statistic.

```
CREATE     ; Post1973 = year > 1973 $
REGRESS    ; For [Post1973 = *,0,1] ; Quietly ; Lhs = logg ; Rhs = x $
DECOMPOSE $
```

The following output results. (If **; Quietly** is omitted from the **REGRESS** command, then a full set of results is displayed for each of the three iterations.)

```
-----------------------------------------------------
Setting up an iteration over the values of POST1973
The model command will be executed for     2 values
of this variable.  In the current sample of      52
observations, the following counts were found:
Subsample   Observations    Subsample  Observations
POST1973 =   0         21    POST1973=   1         31
POST1973 =****         52
----------------------------------------------------
Actual subsamples may be smaller if missing values
are being bypassed.  Subsamples with 0 observations
will be bypassed.
----------------------------------------------------

-----------------------------------------------------------------
Subsample analyzed for this command is POST1973 =       0
-----------------------------------------------------------------
Subsample analyzed for this command is POST1973 =       1
-----------------------------------------------------------------
Full pooled sample is used for this iteration.
-----------------------------------------------------------------
Decomposition of Changes in Average Functions
Model Used in Computations = Linear Regression Function
-------------------------------------------------------------------
               Sample is POST1973= 0         POST1973= 1     Sample
Estimates Based on                (0)                 (1)      Size
POST1973 = 0 (a)       -.492872 (a,0)       .551191 (a,1)        21
POST1973 = 1 (b)       -.305223 (b,0)     -.097432 (b,1)        31
Pooled   =** (*)       -.486167 (*,0)     -.101974 (*,1)        52
-------------------------------------------------------------------
Wald Test of Difference in the Two Coefficient Vectors
Chi squared[  6] =  582.4610        , P Value =  .0000 <--
-------------------------------------------------------------------
Total Change in Function    (a,0) - (b,1) =     -.395440 ( 100.00%)
-------------------------------------------------------------------
Oaxaca    | Due to data is  (a,0) - (a,1) =     -1.04406 ( 264.03%)
Blinder   | Due to beta is  (a,1) - (b,1) =      .648623 (-164.03%)
-------------------------------------------------------------------
Daymont   | Due to data is  (b,0) - (b,1) =     -.207791 (  52.55%)
Andrisani | Due to beta is  (a,0) - (b,0) =     -.187649 (  47.45%)
-------------------------------------------------------------------
Daymont   | Due to data is  (b,0) - (b,1) =     -.207791 (  52.55%)
Andrisani | Due to beta is  (a,1) - (b,1) =      .648623 (-164.03%)
(3 Fold)  | Due to function (a,0) - (b,0) -
          |                 (a,1) - (b,1) =     -.836271 ( 211.48%)
-------------------------------------------------------------------
Ransom    | Due to data is  (*,0) - (*,1) =     -.384194 (  97.16%)
Oaxaca    | Due to beta is  (a,0) - (*,0) +     -.011246 (   2.84%)
Neumark   |                 (*,1) - (b,1)
-------------------------------------------------------------------
```

E8.2.7 Homogeneity Test

The following command set can be used to test for homogeneity of a set of *NG* subsamples of the sample. The initial **NAMELIST** command defines the variables in the regression equation. The **CREATE** command defines the dependent variable and the index variable that is used to partition the sample into *NG* groups. The rest of the command set is generic and need not be changed. The only displayed output is the test statistic and the critical value from the F table.

```
? Define the set of Rhs variables in the regression model
NAMELIST    ; x = the relevant x vector $
CREATE      ; y = the dependent variable ; Group = the index 1,2,… $
? Lines below here are generic and do not need to be changed.
SAMPLE      ; All $
CALC        ; ng = max(group) ; sspool = Ess(x,y) ; sssum = 0 ; nt = n $
PROC $
INCLUDE     ; New ; Group = i $
CALC        ; sssum = sssum + Ess(x,y) $
ENDPROC $
EXEC        ; i = 1,ng $
CALC        ; List ; f = ((sspool-sssum)/(kreg*(ng-1))) / (sssum/(nt-ng*kreg)) $
CALC        ; List ; Ftb(.95, (kreg*(ng-1)), (nt-ng*kreg)) $
```

Using the regression defined in the previous example, we defined the group variable by three periods,

```
CREATE      ; Group = 1 + year > 1073 + year > 1985 $
```

The result of the test for this partitioning of the period 1953 to 2004 is

```
[CALC] F       =     72.7744898
[CALC] *Result*=      2.0500398
```

Since the sample F statistic is greater than the critical value, the homogeneity hypothesis is rejected.

E8.2.8 J Tests for Nonnested Hypotheses

We suppose that the dependent variable is **y** and there are two competing sets of regressors, **X** and **Z**, which are nonnested. Which is the right one? Davidson and MacKinnon propose a simple method of testing the hypothesis in the linear case. We regress **y** on **X** and compute the fitted values, then regress **y** on **Z** and these fitted values. If **Z** is the correct regressor vector, the coefficient on the fitted values should be close to zero by a conventional t test. We then reverse the roles of **X** and **Z** and repeat (and hope the results are consistent). The commands are:

```
NAMELIST    ; z = ...  $
NAMELIST    ; x = ...  $
REGRESS     ; Lhs = y ; Rhs = x ; Keep = yfx $
```

In this regression, we examine the coefficient on *yfx*.

```
REGRESS     ; Lhs = y ; Rhs = z, yfz $
REGRESS     ; Lhs = y ; Rhs = z ; Keep = yfz $
```

In this regression, we examine the coefficient on *yfz*.

```
REGRESS     ; Lhs = y ; Rhs = x, yfz $
```

E8.3 Restricted Least Squares

This section describes procedures for estimating the restricted regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

subject to $\mathbf{R}\boldsymbol{\beta} \geq \mathbf{q}$.

R is a $J \times K$ matrix assumed to be of full row rank. That is, we impose J linearly independent restrictions. They may be equality restrictions, inequality restrictions, or a mix of the two.

E8.3.1 Equality Restrictions

If $\mathbf{X}'\mathbf{X}$ is nonsingular, the constrained ordinary least squares estimator is

$$\mathbf{b}_c = \mathbf{b} - [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\ [\mathbf{Rb}\text{-}\mathbf{q}]$$

where $\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}$

is the unrestricted least squares estimator. The estimator of the variance of the constrained estimator is

$$\text{Est.Var}[\mathbf{b}_c] = s^2[\mathbf{X}'\mathbf{X}]^{-1} - s^2[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}[\mathbf{X}'\mathbf{X}]^{-1}.$$

where $s^2 = (\mathbf{y} - \mathbf{X}\,\mathbf{b}_c)'\,(\mathbf{y} - \mathbf{X}\,\mathbf{b}_c)/(n\text{-}K\text{+}J)$.

The syntax for imposing linear restrictions is the same as that for testing linear restrictions described in the previous section. As before, there are two general forms available, depending on whether the regression involves only linear terms or explicit terms such as interactions, quadratics, or logs. You may impose as many restrictions as you wish with this estimator; simply separate the restrictions with commas.

To continue the earlier example, consider the hybrid cost function in Section E8.2.3,

$$\text{Logcost} = \beta_1 + \beta_2\text{logPk} + \beta_3\text{logPl} + \beta_4\text{logPf} + \gamma_1\text{logq} + \gamma_2(\text{logq})^2 + \varepsilon.$$

Linear homogeneity in the input prices requires $\beta_2 + \beta_3 + \beta_4 = 1$. The Cobb-Douglas cost function results if the restriction γ_2 equals zero is imposed in addition to the linear homogeneity restriction. The restricted regression is obtained with

```
REGRESS  ; Lhs = lncost
         ; Rhs = one,lnpk,lnpl,lnpf,lnq,lnqsq
         ; CLS: lnpk + lnpl + lnpf = 1, lnqsq = 0 $
```

In this case, both unrestricted and restricted regressions are reported. In the second, the F statistic (or Wald statistic if a robust covariance matrix is used) for testing the restrictions is also reported.

```
----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=LNCOST  Mean                 =        3.00917
            Standard deviation   =        1.56241
            No. of observations  =            123  Degrees of freedom
Regression  Sum of Squares       =        295.375           5
Residual    Sum of Squares       =        2.44152         117
Total       Sum of Squares       =        297.817         122
            Standard error of e  =         .14446
Fit         R-squared            =         .99180  R-bar squared =   .99145
Model test  F[  5,    117]       =     2830.93377  Prob F > F*   =   .00000
Diagnostic  Log likelihood       =       66.52372  Akaike I.C.   = -3.82200
            Restricted (b=0)     =     -228.91353  Bayes  I.C.   = -3.68482
            Chi squared [  5]    =      590.87450  Prob C2 > C2* =   .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LNCOST|  Coefficient      Error       t     |t|>T*        Interval
--------+-------------------------------------------------------------------
Constant|   -7.04431***      .87956    -8.01  .0000    -8.76821  -5.32041
    LNPK|     .05419         .09100      .60  .5527     -.12418    .23255
    LNPL|     .24309***      .08922     2.72  .0074      .06823    .41795
    LNPF|     .66279***      .05044    13.14  .0000      .56394    .76164
     LNQ|     .39105***      .03713    10.53  .0000      .31827    .46383
   LNQSQ|     .03123***      .00259    12.07  .0000      .02616    .03630
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------

----------------------------------------------------------------------------
Restricted  least squares regression ............
LHS=LNCOST  Mean                 =        3.00917
            Standard deviation   =        1.56241
            No. of observations  =            123  Degrees of freedom
Regression  Sum of Squares       =        292.338           3
Residual    Sum of Squares       =        5.47918         119
Total       Sum of Squares       =        297.817         122
            Standard error of e  =         .21458
Fit         R-squared            =         .98160  R-bar squared =   .98114
Model test  F[  3,    119]       =     2116.38681  Prob F > F*   =   .00000
Diagnostic  Log likelihood       =       16.81117  Akaike I.C.   = -3.04619
            Restricted (b=0)     =     -228.91353  Bayes  I.C.   = -2.95474
            Chi squared [  3]    =      491.44939  Prob C2 > C2* =   .00000
Restrictions F[  2,    117]      =       72.78374  Prob F > F*   =   .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LNCOST|  Coefficient      Error       t     |t|>T*        Interval
--------+-------------------------------------------------------------------
Constant|   -8.03028***      .50529   -15.89  .0000    -9.02064  -7.03993
    LNPK|     .16885*        .09119     1.85  .0666     -.00987    .34758
    LNPL|     .12647         .10020     1.26  .2094     -.06993    .32287
    LNPF|     .70468***      .07434     9.48  .0000      .55898    .85038
     LNQ|     .83029***      .01067    77.85  .0000      .80939    .85120
   LNQSQ|        0.0    .....(Fixed Parameter).....
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
----------------------------------------------------------------------------
```

E8.3.2 Equality Restrictions and Singularity

LIMDEP does not use precisely the preceding formulation internally. The unrestricted estimator may not exist – $\mathbf{X}'\mathbf{X}$ may be singular. It may still be possible to obtain the restricted estimates, however. The general result is that while the unrestricted model may involve too many parameters to estimate, the restrictions may eliminate enough parameters to leave an estimable model. Instead of the formulas above, we solve the constrained first order conditions for least squares using the Lagrangean multiplier method. That is,

$$\text{Minimize}(\boldsymbol{\beta},\boldsymbol{\lambda})\text{: } ½(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}).$$

The first order conditions are

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{q} \end{bmatrix}.$$

When $\mathbf{X}'\mathbf{X}$ is nonsingular, this produces the formulas above. (See Greene and Seaks (1991).) But, even if $\mathbf{X}'\mathbf{X}$ is singular, this equation system may have a solution. The constrained least squares estimator, if it exists, is the solution to the preceding. If $\mathbf{X}'\mathbf{X}$ does have full rank, this equation system produces the usual constrained estimator. If not, *LIMDEP* just proceeds to this solution. If it exists, no warning is given about the unconstrained estimator – presumably it was of no interest anyway; *LIMDEP* simply produces the constrained estimator.

To illustrate how this works, consider the following obviously badly constructed example:

```
CREATE      ; extra = lnqsq $
REGRESS     ; Lhs = lncost
            ; Rhs = one,lnpk,lnpl,lnpf,lnq,lnqsq,extra
            ; CLS: lnpk + lnpl + lnpf = 1, lnqsq = 0, extra = 0 $
```

The variable *extra* is identical to *lnqsq*, so the unrestricted regression cannot be computed. There is a textbook case of multicollinearity. However, the restrictions include ‘extra = 0,’ so if the restrictions are actually imposed, the regression is fine. The following are the results from this command. The unrestricted regression cannot be computed, but the restricted one can. In fact, it is identical to the one in Section E8.2.3.

```
-----------------------------------------------------------------------------
Restricted   least squares regression ............
LHS=LNCOST   Mean                 =       3.00917
             Standard deviation   =       1.56241
             No. of observations  =           123  Degrees of freedom
Regression   Sum of Squares       =       292.338          3
Residual     Sum of Squares       =       5.47918        119
Total        Sum of Squares       =       297.817        122
             Standard error of e  =        .21458
Fit          R-squared            =        .98160  R-bar squared =   .98114
Model test   F[  3,   119]        =    2116.38681  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =      16.81117  Akaike I.C.   = -3.04619
             Restricted (b=0)     =    -228.91353  Bayes  I.C.   = -2.95474
             Chi squared [  3]    =     491.44939  Prob C2 > C2* =   .00000
Restrictions F[  3,   116]        =        .00000  Prob F > F*   =  1.00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LNCOST|  Coefficient      Error       t     |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.03028***      .50529   -15.89  .0000    -9.02064   -7.03993
    LNPK|     .16885*        .09119     1.85  .0666     -.00987     .34758
    LNPL|     .12647         .10020     1.26  .2094     -.06993     .32287
    LNPF|     .70468***      .07434     9.48  .0000      .55898     .85038
     LNQ|     .83029***      .01067    77.85  .0000      .80939     .85120
   LNQSQ|        0.0    .....(Fixed Parameter).....
   EXTRA|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

The usual method of avoiding the dummy variable trap is to drop one of the dummy variables. Consider, for example, the duration data used in Chapters E58-E60. We have the variables *time*, *sex*, and *married*. For present purposes, consider a linear regression of *time* on a constant, *married*, and *sex*. The dummy variable, *sex* would normally be coded 0/1, and the model would include a constant and this dummy variable. The constant term would give the overall intercept, and the coefficient on sex would give the deviation of the group with *sex* = 1 (male in this case) from this constant. Suppose, instead, we were to attempt to fit the regression

$$time = \beta_1 + \beta_2 male + \beta_3 female + \beta_4 married + \varepsilon.$$

This regression suffers from perfect multicollinearity; the second and third variables sum to the constant term. Therefore, the unrestricted coefficient vector for this model cannot be estimated. We add the constraint

$$\beta_2 + \beta_3 = 0.$$

This is just like the usual model, except that with this restriction, there is an average constant, and the two coefficients will give the difference of each of the two groups from the mean (in this case, a mean of only two items). The following shows the result of estimating this model with *LIMDEP*. The unrestricted regression cannot be computed, but the restricted one can. Notice that *LIMDEP* reports zero for the F test. This is not a substantive restriction. With the restriction, the model becomes just estimable.

The commands are:

READ ; Nobs = 22
; Nvar = 4
; Names = time,status,sex,married
; By Variables $

```
11  3 19 32 2 14 8 21 16 5 2 8 14 18 18 21 10 1 9 23 19 7
 1  1  0  1 1  1 1  1  0 1 1 1  1  1  1  1  0 1 0  1  1 1
 0  0  1  0 1  0 1  1  1 0 0 1  1  0  1  1  0 0 1  1  0 1
 1  1  2  2 1  1 1  1  2 2 2 1  1  1  2  1  2 2 1  1  2 1
```

CREATE ; male = sex ; female = 1 - male $
REGRESS ; Lhs = time
; Rhs = one, male, female, married
; CLS: male + female = 0 $

```
-----------------------------------------------------------------------------
Restricted   least squares regression ............
LHS=TIME     Mean                 =      12.77273
             Standard deviation   =       8.12364
             No. of observations  =            22  Degrees of freedom
Regression   Sum of Squares       =       58.1378          2
Residual     Sum of Squares       =       1327.73         19
Total        Sum of Squares       =       1385.86         21
             Standard error of e  =       8.35944
Fit          R-squared            =        .04195  R-bar squared =  -.05890
Model test   F[  2,    19]        =        .41598  Prob F > F*   =   .66556
Diagnostic   Log likelihood       =     -76.31863  Akaike I.C.   =  4.37291
             Restricted (b=0)     =     -76.79005  Bayes  I.C.   =  4.52169
             Chi squared [  2]    =        .94283  Prob C2 > C2* =   .62412
Restrictions F[  1,    18]        =        .00000  Prob F > F*   =  1.00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    TIME|  Coefficient     Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|    9.14247      5.80682     1.57  .1328     -2.23868  20.52363
    MALE|    1.59946      1.91392      .84  .4143     -2.15175   5.35067
  FEMALE|   -1.59946      1.91392     -.84  .4143     -5.35067   2.15175
 MARRIED|    2.47312      3.87660      .64  .5315     -5.12488  10.07111
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

LIMDEP uses this method to fit the two way fixed effects model – see Section E17.3. The regression model is

$$y_{it} = \mu + \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}., \; i = 1,\ldots,n, \; t = 1,\ldots,T_i.$$

In order to compute the coefficients in this model, it is necessary to impose two restrictions, because both the individual effects and the time effects sum to one, the constant term. To estimate this model, *LIMDEP* drops one of the time constants and imposes that the individual constants sum to zero. (Because the number of periods can vary, it is necessary to create the time dummy variables and insert them into a one way fixed effects model. The template method of using deviations from time means does not give the correct answer when the number of periods varies with *i*.)

E8.3.3 Inequality Restricted Least Squares

You may also impose inequality restrictions. The model is specified as before, with the restrictions now specified as weak inequalities. That is,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

subject to $R_{11}\,\beta_1 + R_{12}\,\beta_2 + ... \geq q_1$ (When $R_{jk} = 1$, it may be omitted. Also, the $\geq$ may be $\leq$.)

$R_{21}\,\beta_1 + R_{22}\,\beta_2 + ... \geq q_2$

...

$R_{J1}\,\beta_1 + R_{J2}\,\beta_2 + ... \geq q_J$

This estimation is formulated as a classical quadratic programming problem. That is, we

Minimize (wrt $\boldsymbol{\beta}$) $\mathbf{y}'\mathbf{y} - (2\mathbf{y}'\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\beta}'(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{a} + \mathbf{c}'\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{H}\boldsymbol{\beta}$
Subject to the equality and inequality constraints.

In order to specify estimation of a model subject to inequality constraints, you use the exact same formulation as if they were equality constraints, save for '<=' for less than or equal to and '>=' for greater than or equal to. Also, you may have any mixture of equality constraints, >= and <= constraints in any model. As in the case of the LAD estimator, there is no well defined result for the asymptotic covariance matrix of the inequality constrained estimator. As before, we suggest using bootstrapping as a method of approximating the appropriate matrix. Use

; Nbt = ... number of bootstrap replications

To illustrate, we continue the sampling experiment computed at the beginning of this section. The following creates data generated by a log quadratic production function

$$\log y = \beta_1 + \beta_2 \log l + \beta_3 \log k + \beta_4 \log^2 l + \beta_5 \log^2 k + \beta_6 \log k \log l + \varepsilon$$

The regression model is fit subject to two constraints: $\beta_2 + \beta_3 = 1$ and $\beta_4 + \beta_5 + \beta_6 \leq 0$. The second constraint is actually binding in our results, as the final results have both constraints imposed as equalities. Standard errors are estimated using 20 bootstrap replications.

```
SAMPLE      ; 1-500 $
CALC        ; Ran(123457) $
CREATE      ; l = Rnu(1,3) ; k = Rnu(.5,2)
            ; ll = Log(l) ; lk = Log(k)
            ; lk2 = lk*lk ; ll2 = ll*ll ; lkl = ll*lk
            ; ly = 3 + .6*ll + .4*lk - .05*ll2 - .15*lk2 + .2*lkl + Rnn(0,4) $
REGRESS     ; Lhs = ly ; Rhs = one,ll,lk,ll2,lk2,lkl
            ; CLS: ll + lk = 1, ll2 + lk2 + lkl <= 0
            ; Nbt = 20 $
REGRESS     ; Lhs = ly ; Rhs = one,ll,lk,ll2,lk2,lkl $
```

```
-----------------------------------------------------------------------------
Inequality restricted least squares...............
Nonlinear     least squares regression ............
LHS=LY        Mean                    =       3.49470
              Standard deviation      =       4.03680
              Number of observs.      =           500
Model size    Parameters              =             6
              Degrees of freedom      =           494
Residuals     Sum of squares          =       8110.93
              Standard error of e     =       4.05202
Fit           R-squared               =        .00254
              Adjusted R-squared      =       -.00756
Model test    F[  5,    494] (prob) =      .3(.9390)
Diagnostic    Log likelihood          =   -1406.05910
              Restricted(b=0)         =   -1406.69478
              Chi-sq [  5]  (prob) =   1.3(  .9379)
Info criter. Akaike Info. Criter. =       2.81036
Not using OLS or no constant. Rsqrd & F may be < 0
Note, with restrictions imposed,  Rsqd may be < 0.
Model test    F[  1,    494] (prob) =     .15 (.6949)
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LY|  Coefficient       Error       z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Covariance matrix based on   20 replications.
Constant|    3.15877***       .36403     8.68  .0000     2.44528   3.87226
      LL|     .82931          .74919     1.11  .2683     -.63908   2.29770
      LK|     .17069          .74919      .23  .8198    -1.29770   1.63908
     LL2|    -.15093          .72644     -.21  .8354    -1.57473   1.27287
     LK2|    -.70691         1.37953     -.51  .6084    -3.41074   1.99692
     LKL|    -.44502         1.02906     -.43  .6654    -2.46194   1.57191
--------+--------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=LY        Mean                    =       3.49470
              Standard deviation      =       4.03680
              No. of observations     =           500  Degrees of freedom
Regression    Sum of Squares          =       23.1777            5
Residual      Sum of Squares          =       8108.40          494
Total         Sum of Squares          =       8131.58          499
              Standard error of e     =       4.05139
Fit           R-squared               =        .00285  R-bar squared =  -.00724
Model test    F[  5,    494]          =        .28242  Prob F > F*   =   .92274
Diagnostic    Log likelihood          =   -1405.98117  Akaike I.C.   =  2.81005
              Restricted (b=0)        =   -1406.69478  Bayes  I.C.   =  2.86062
              Chi squared [  5]       =       1.42720  Prob C2 > C2* =   .92131
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LY|  Coefficient       Error       t     |t|>T*          Interval
--------+--------------------------------------------------------------------
Constant|    3.40854***       .76286     4.47  .0000     1.91336   4.90371
      LL|    -.09423         2.56561     -.04  .9707    -5.12273   4.93428
      LK|    -.05073         1.16722     -.04  .9653    -2.33845   2.23698
     LL2|     .54116         2.06524      .26  .7934    -3.50664   4.58896
     LK2|    -.70218         1.38386     -.51  .6121    -3.41449   2.01012
     LKL|    -.15810         1.55538     -.10  .9191    -3.20659   2.89039
--------+--------------------------------------------------------------------
```

E9: Non- and Semiparametric Regression Models

E9.1 Introduction

This chapter will detail estimation of a single equation, linear regression model

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ i = 1,...,n.$$

The initial stochastic assumptions of the classical regression model depart from

$$E[y_i|\ \mathbf{x}_i\ ] = \boldsymbol{\beta}'\mathbf{x}_i.$$

This chapter considers models that relax this assumption. The least restrictive version is the nonparametric regression model,

$$y_i = m(x_i) + \varepsilon_i \text{ where } E[\varepsilon_i \mid x_i] = 0$$

for a single variable x_i. This makes minimal assumptions about the relationship between y_i and x_i. The LOWESS method described in Section E9.5 is a graphical technique that is based on this principle.

A convenient extension of the nonparametric regression approach is the index function model

$$E[y|\boldsymbol{\beta}'\mathbf{x} = z] = F_{\boldsymbol{\beta}}(z).$$

Finally, we describe a semiparametric approach. The median regression is

$$\text{Med}[\ y_i \mid \mathbf{x}_i\ ] = \boldsymbol{\beta}'\mathbf{x}_i.$$

This is the least absolute deviations estimator. The median is the 50^{th} percentile. Section E9.4 describes an estimator in which any specified quantile may be analyzed – and all may differ;

$$p^{th} \text{ quantile } [\ y_i \mid \mathbf{x}_i\ ] = \boldsymbol{\beta}'\mathbf{x}_i.$$

Since the regression may differ at different quantiles, this draws the model closer to the nonparametric regression.

E9.2 Nonparametric (Kernel Density) Regression Estimation

The basic command for the nonparametric regression estimator is

NPREG ; Lhs = dependent variable
; Rhs = regressor $

NPREG is used to fit a nonparametric regression function. This estimator estimates a smooth, regression function,

$$E[y|x] = F(x)$$

using the method of kernels. A straightforward extension detailed in Section E9.2.2 is to a single index model,

$$E[y|\boldsymbol{\beta}'\mathbf{x} = z] = F_{\beta}(z),$$

for any parameter vector $\boldsymbol{\beta}$ (assumed known or at least given). With an appropriate choice of $\mathbf{x}$ and $\boldsymbol{\beta}$, and by rescaling the response, **NPREG** can estimate any sufficiently smooth univariate regression function with known bounded range. **NPREG** takes as input sample data consisting of n observations $(y_i,\mathbf{x}_i)$ where $\mathbf{x}_i$ is a K-vector of regressors, and y_i is the dependent variable, and a parameter vector $\boldsymbol{\beta}$. ($\boldsymbol{\beta}$ is omitted for the simple univariate model.) **NPREG** also requires a smoothing parameter, h, also called the bandwidth parameter. The simplest nonparametric regression model between a y variable and a single x variable is obtained simply by specifying $\mathbf{x}_i$ = that variable and $\boldsymbol{\beta} = 1$. We consider this case first, then turn to the index function model.

E9.2.1 Nonparametric Regression on a Single Variable

To estimate the regression function

$$E[y|x] = F(x)$$

for a single variable, x, use

NPREG ; Lhs = y variable
; Rhs = x variable $

All aspects of the specification will be taken care of internally. Output consists of a text description of the data followed by the plot of the estimated regression function.

To illustrate, we will use the gasoline market data employed in several previous examples. The first plot showing the nonparametric regression of *logg* on *logpg* shows the model at work, but also demonstrates that the problem of omitted variables impacts the nonparametric regression as well. It is not robust to omitted variables.

NPREG ; Lhs = logg ; Rhs = logpg $

```
+----------------------------------------+
| Nonparametric Regression for LOGG      |
| Observations       =             52    |
| Points plotted     =             52    |
| Bandwidth          =        .273055    |
| Statistics for abscissa values----     |
| Mean               =       3.729303    |
| Standard Deviation =        .678991    |
| Minimum            =       2.813491    |
| Maximum            =       4.819483    |
| ----------------------------------     |
| Kernel Function    =       Logistic    |
| Cross val. M.S.E.  =        .017347    |
| Results matrix     =         KERNEL    |
+----------------------------------------+
```

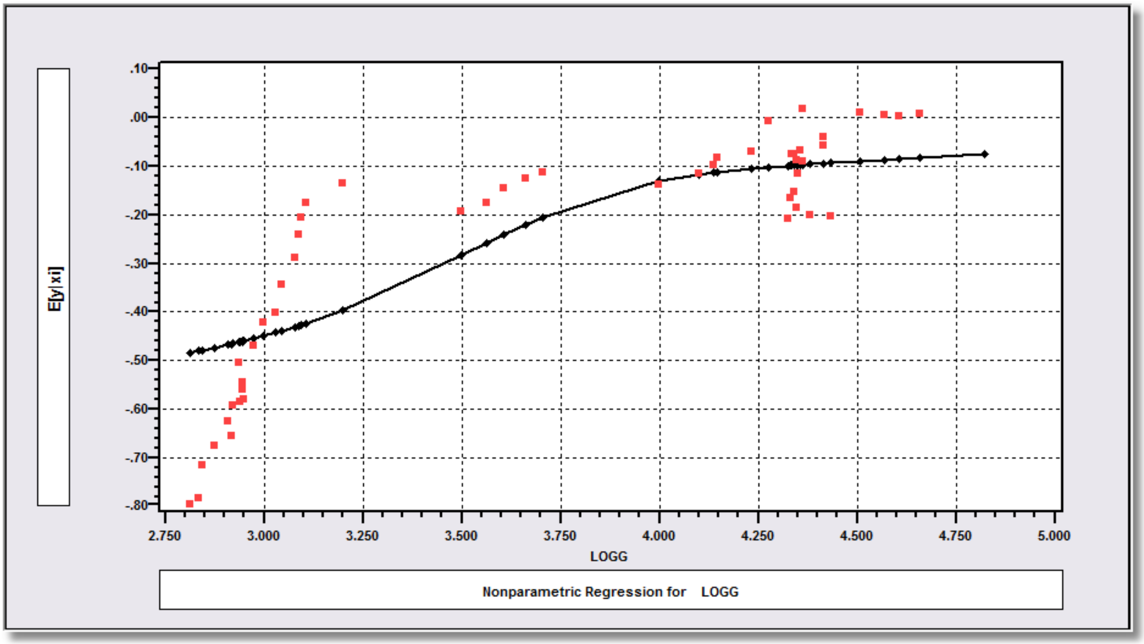


Figure E9.1 Nonparametric Regression

E9.2.2 Estimating a Nonparametric Single Index Regression Function

To analyze a regression function of the form

$$E[y|\boldsymbol{\beta}'\mathbf{x}] = F_{\beta}(\boldsymbol{\beta}'\mathbf{x}),$$

you must provide the values of the parameters as well as the data. Note that the estimator is not estimating the parameters, it is analyzing the regression function based on the index function. You may, of course, provide any parameters you wish. One possibility might be to analyze your linear regression model to see if it is really linear. Keep in mind, the results are only suggestive, as the parameters you would provide are already based on an assumption of linearity.

We first compute $z_i = \boldsymbol{\beta}'\mathbf{x}_i$ for each observation. The sample standard deviation of the observations, s, is computed next. Then, h, s, z_i, and the kernel function, $K[\cdot]$ are used to define a weighting function

$$w_i(z_j) = \frac{1}{h}K\left[\frac{z_i - z_j}{h}\right],$$

then, the regression function is

$$F(z_j) = \frac{\sum_{i=1}^{N} w_i(z_j)y_i}{\sum_{i=1}^{N} w_i(z_j)}.$$

The default kernel function used is the density for the standardized logistic. Several alternatives are available. These are discussed below.

The command for **NPREG** for an index function model is

```
(Commands to obtain the parameter vector)
NPREG          ; Lhs = y variable
               ; Rhs = ... regressors that correspond to the parameters
               ; Parameters = parameter values $
```

(In earlier versions of *LIMDEP*, **; Start = …** would replace **; Parameters = …** You may still use this syntax.) Since the command does no estimation of its own, you must provide the parameter values if you are plotting a regression function with more than one independent variable. Output from this estimator consists of a summary table and a plot of $F(z_j)$ against z_j.

To illustrate the computations, we continue the analysis above, by analyzing

$$logg = \beta_1 + \beta_2 logpg + \beta_3 logincome + \varepsilon$$

The command sequence is

```
REGRESS        ; Lhs = logg ; Rhs = one,logpg,logy $
NPREG          ; Lhs = logg ; Rhs = one,logpg,logy
               ; Parameters = b $
```

(The least squares regression results are omitted.)

```
+---------------------------------------+
| Nonparametric Regression for LOGG     |
| Observations       =            52    |
| Points plotted     =            52    |
| Bandwidth          =       .094374    |
| Statistics for abscissa values----    |
| Mean               =      -.257129    |
| Standard Deviation =       .231106    |
| Minimum            =      -.682513    |
| Maximum            =       .094240    |
| ---------------------------------     |
| Kernel Function    =      Logistic    |
| Cross val. M.S.E.  =       .005011    |
| Results matrix     =        KERNEL    |
+---------------------------------------+
```

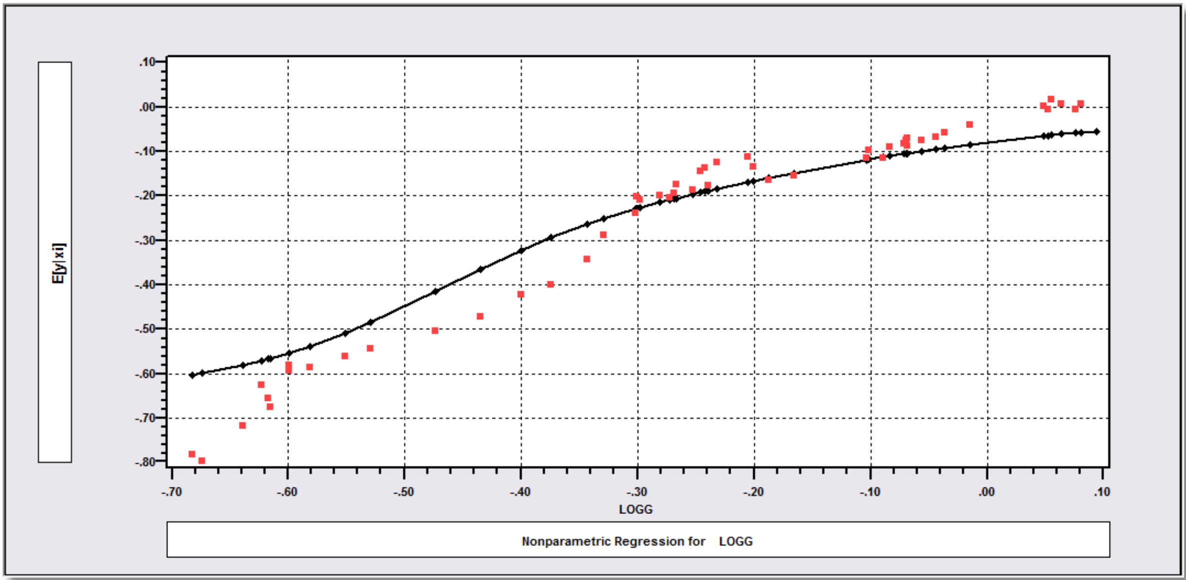


Figure E9.2 Nonparametric Regression for Gasoline Consumption

E9.2.3 Options for NPREG

Kernel Functions

The primary component of the computation is the kernel function, K[.]. Eight alternatives are provided:

1. Epanechnikov:	$K[z]$	=	$.75(1 - .2z^2)$ / Sqr(5) if $|z|$ <= 5, 0 else,
2. Normal:	$K[z]$	=	$\phi(z)$ (normal density),
3. Logit:	$K[z]$	=	$\Lambda(z)[1-\Lambda(z)]$ (default),
4. Uniform:	$K[z]$	=	.5 if $|z| \leq 1$, 0 1 else,
5. Beta:	$K[z]$	=	$(1-z)(1+z)/24$ if $|z| \leq 1$, 0 1 else,
6. Cosine:	$K[z]$	=	$1 + \cos(2\pi z)$ if $|z| < .5$, 0 else,
7. Triangle:	$K[z]$	=	$1 - |z|$, if $|z| \leq 1$, 0 else,
8. Parzen:	$K[z]$	=	$4/3 - 8z^2 + 8|z|^3$ if $|z| \leq .5$, $8(1-|z|)^3$ else.

You may specify the kernel function to be used with

; Kernel = one of the eight types of kernels listed above,

e.g., **; Kernel = normal**

The logit kernel function is used if you do not specify one. Epanechnikov is a popular alternative.

Bandwidth Parameter

The default value for the smoothing parameter,

$$h = .9Q/n^{0.2}$$

where $$Q = \min(\text{std.dev of } \mathbf{b}'\mathbf{x}_i., \text{ data range}/1.5)$$

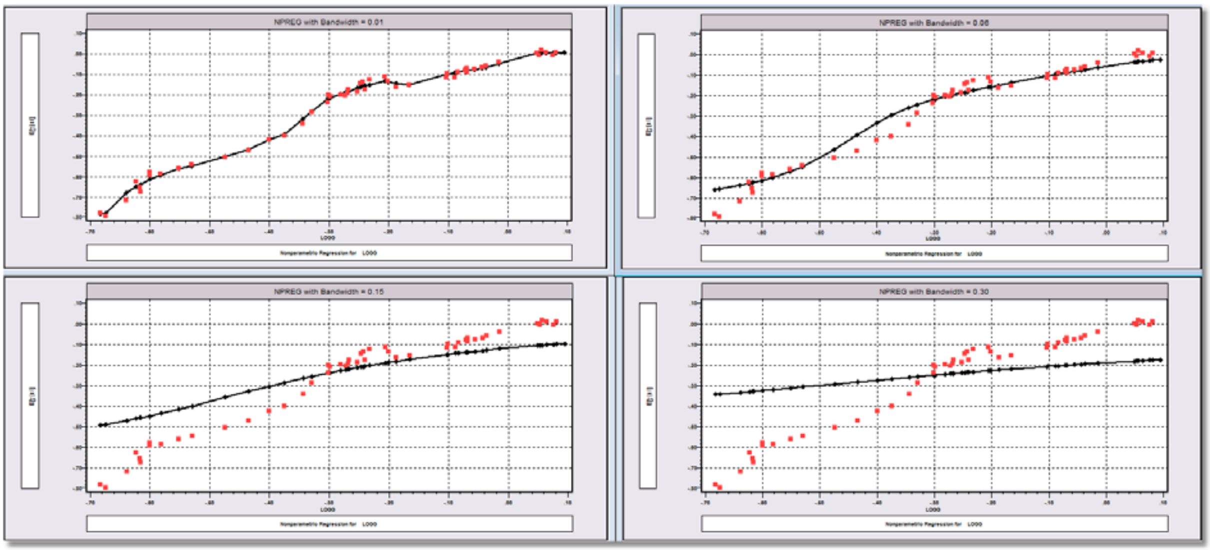
An alternative value is provided with

; Smooth = value

There is no definitive theory for choosing the right smoothing parameter, h. Large values will cause the estimated function to flatten at the average value of y_i. Values close to zero will cause the function to pass through the points z_i,y_i and to become computationally unstable elsewhere. Higher values will smooth the function, but will, in the process, degrade the fit of the function to the data.

The bandwidth parameter is a crucial element of the analysis. For example, note that in the preceding example, the bandwidth parameter is about .094 Figure E9.3 shows the estimated regression functions for four values, .01, .06, .15 and .30. The differences in the estimated function are clearly visible.

Bandwidth = 0.01 **Bandwidth = 0.06**



Bandwidth = 0.15 **Bandwidth = 0.30**

Figure E9.3 Effect of Bandwidth on Kernel Regression

Number of Points to Plot

The default number of points to be plotted is M = 100, or the sample size, n, if $n \leq 5000$ and you do not specify M or the range. Use

; Pts = M

to compute the function at M equally spaced points in the range defined as below by the sample values. **; Limits** and **; Pts** may be given together to specify a grid in a particular range. **; Pts** may be any number up to the number of rows in the data area. If the value you give exceeds the limit of rows, an error will occur and computation will cease.

Range of Estimation

The next set of specifications dictates the points at which the regression function should be computed. The default is to compute the function at the data points, of which there are n. An alternative is

; Limits = lower,upper

to compute the function at M equally spaced points in the range [lower,upper]. The default is n equally spaced points with lower = the sample minimum of $\boldsymbol{\beta}'\mathbf{x}_i - h$ and upper = the maximum + h.

Cross Validation Mean Square Prediction Error

The cross validation mean squared prediction error (*CVMSPE*) is a goodness of fit measure. Each observation, '*i*' is excluded in turn from the sample. Using the reduced sample, the regression function is reestimated at the point z_i in order to provide a point prediction for y_i. The average squared prediction error defines the *CVMSPE*. The calculation is defined by the point predictions,

$$F^*(\boldsymbol{\beta}'\mathbf{x}_i) = \frac{\sum_{j\neq i} y_j K\left[(y_j - \boldsymbol{\beta}'\mathbf{x}_i)/h\right]}{\sum_{j\neq i} K\left[(y_j - \boldsymbol{\beta}'\mathbf{x}_i)/h\right]}.$$

Then, $$CVMSPE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - F^*(\boldsymbol{\beta}'\mathbf{x}_i)\right)^2 .$$

The *CVMSPE* is more or less a counterpart to the sum of squares in regression, which suggests that one could compute a fit measure by using

```
CALC      ; List ; npregfit = 1 - cvmspe/((n-1)*Var(y)) $
```

The usual warning about fit measures in nonlinear regressions applies, however. This number need not be positive.

E9.2.4 Output from NPREG

Results from **NPREG** consist of

- The table shown in the earlier examples,
- An $M \times 2$ matrix named *kernel*, whose first column is the sorted values of $\boldsymbol{\beta}'\mathbf{x}_i$, and the second column is the estimated values from the regression function,
- A scalar named *cvmspe* (described below),
- The plot of the estimated function.

Note that since M may not equal n, there is no necessary correspondence between the observations and the values in the matrix. In addition to the stored result, a plot of the regression function, as shown earlier, is part of the usual output for this estimator as well. You can request a listing of the ordinate values, z_i, and the estimated values of the regression function by including

; List

in the command. A more compact listing in a scrollable window can be obtained by double clicking the matrix *kernel* in the project window. (See Figure E9.4.)

You may provide a title for the figure with

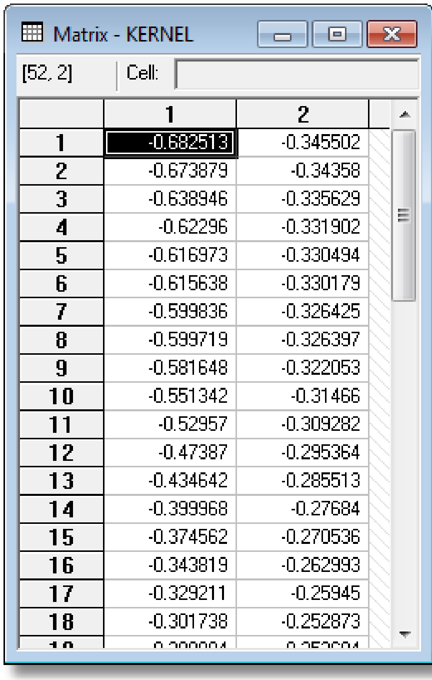
; Title = ... <the title for the figure> ...

If your sample size is 5000 or less and you have not specified the number of points to plot or the range in which to plot, then **NPREG** will have used the actual data to generate the abscissas for the function. In this case, the fitted function will correspond to the actual data, and you can keep the function values as predictions for the corresponding values of the Lhs variable. Use

; Keep = name to retain function values as predictions
; Res = name to retain (actual – function value) as a set of residuals

The full set of results from **NPREG** would appear like the following:

```
+----------------------------------------+
| Nonparametric Regression for LOGG      |
| Observations       =             52    |
| Points plotted     =             52    |
| Bandwidth          =        .200000    |
| Statistics for abscissa values----     |
| Mean               =       4.108115    |
| Standard Deviation =        .400050    |
| Minimum            =       3.390065    |
| Maximum            =       4.704196    |
| ----------------------------------     |
| Kernel Function    =       Logistic    |
| Cross val. M.S.E.  =        .022038    |
| Results matrix     =         KERNEL    |
+----------------------------------------+
```



Matrix - KERNEL

[52, 2] Cell:

	1	2
1	-0.682513	-0.345502
2	-0.673879	-0.34358
3	-0.638946	-0.335629
4	-0.62296	-0.331902
5	-0.616973	-0.330494
6	-0.615638	-0.330179
7	-0.599836	-0.326425
8	-0.599719	-0.326397
9	-0.581648	-0.322053
10	-0.551342	-0.31466
11	-0.52957	-0.309282
12	-0.47387	-0.295364
13	-0.434642	-0.285513
14	-0.399968	-0.27684
15	-0.374562	-0.270536
16	-0.343819	-0.262993
17	-0.329211	-0.25945
18	-0.301738	-0.252873

Figure E9.4. Matrix Result from KERNEL

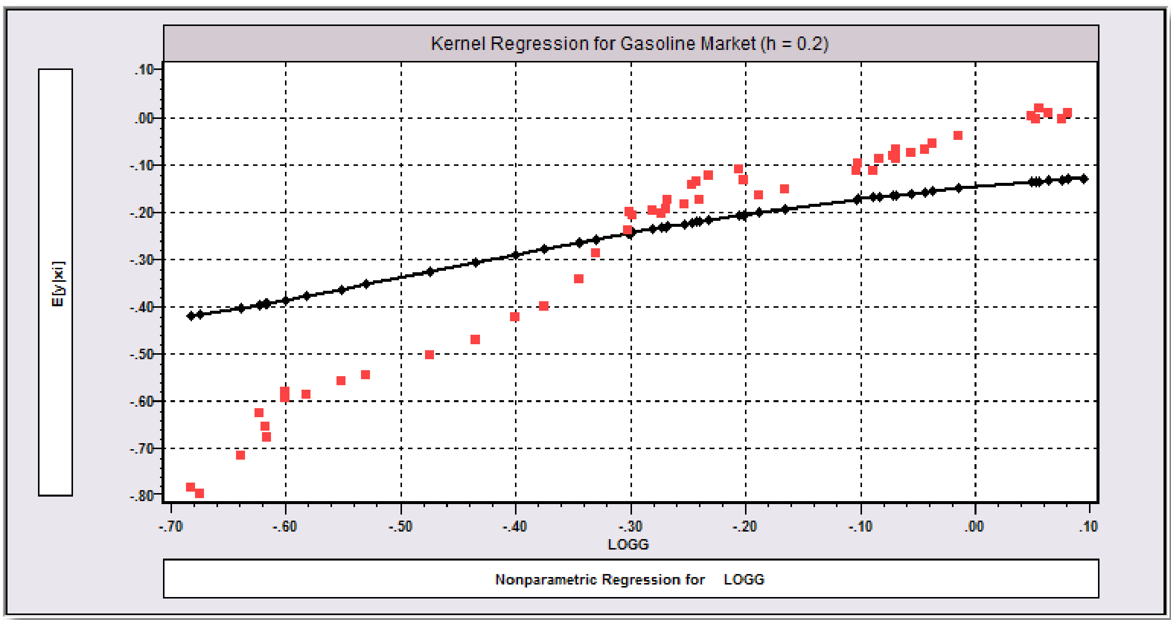


Figure E9.5 Kernel Regression

E9.3 The Least Absolute Deviations Estimator

The basic command for the least absolute deviations estimator is

REGRESS ; Lhs = dependent variable ; Rhs = list of regressors
; Alg = LAD $

The least squares estimator is robust to many variations in the specification. For consistency, it requires only that the data be 'well behaved' and that the conditional mean function, $E[y|\mathbf{x}]$ be linear in $\mathbf{x}$, $\boldsymbol{\beta}'\mathbf{x}$. Still, some researchers have criticized the estimator for the fact that it may be unduly influenced by outlying observations in small samples. The least absolute deviations (LAD) estimator has been advocated (see, e.g., Koenker and Bassett (1982) for discussion) as a preferable alternative. (LAD is a special case of the quantile regression estimator discussed in the next section. The LAD estimator corresponds to the median regression estimator.)

The LAD estimator can be obtained by specifying the regression as usual, and adding

; Alg = LAD

to the **REGRESS** command. The estimator is computed by solving the linear programming problem,

$$\text{Min (wrt } \boldsymbol{\beta}) \ \sum_{i=1}^{n} \left| y_i - \boldsymbol{\beta}'\mathrm{x}_i \right|$$

There is no definitive result for the asymptotic covariance matrix for the LAD estimator. Koenker and Bassett (1982) provide a candidate which may or may not prove useful. Their estimator is

$$\text{Asy.Var}[\mathbf{b}_{LAD}] = (\mathbf{X}'\mathbf{X})^{-1}\, \mathbf{X}'\mathbf{W}^2\mathbf{X}\, (\mathbf{X}'\mathbf{X})^{-1}$$

where $\mathbf{W}$ = Diag[.5 / $f(0)$] and $f(0)$ is the true density of the disturbances evaluated at zero. This requires knowledge of the true density, which is unspecified here. However, one could use the kernel estimator described above to estimate it. Once the set of residuals is in hand, one could use the estimator

$$\hat{f}(0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left[\frac{(y_i - \mathbf{x}_i'\mathbf{b}_{LAD})}{h}\right] = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left[\frac{e_{i,LAD})}{h}\right]$$

One useful special case would be that of the normal distribution. If the disturbances are distributed normally with zero mean and constant variance σ^2, then the result specializes to

$$\text{Asy.Var}[\mathbf{b}_{LAD}|\text{normal}] = \frac{\pi}{2}\sigma^2(\mathbf{X}'\mathbf{X})^{-1},$$

which is a simple multiple of the result for least squares. It would also be simple to compute. (Of course, if the disturbances are known to be normal, we should be using least squares.) We might consider two approaches in this case,

$$\text{Est.Asy.Var}[\mathbf{b}_{LAD}|\text{normal}] = \frac{\pi}{2}\frac{\mathbf{e}'\mathbf{e}}{n}(\mathbf{X}'\mathbf{X})^{-1}$$

or, allowing for possible heteroscedasticity and using White's robust estimator,

$$\text{Est.Asy.Var}[\mathbf{b}_{LAD}|\text{normal}] \;=\; \frac{\pi}{2}\;(\mathbf{X}'\mathbf{X})^{-1} \times \sum\nolimits_{t=1}^{T} e_t^2 \mathbf{x}_t \mathbf{x}_t{}' \times (\mathbf{X}'\mathbf{X})^{-1.}$$

Both of these are based on the normal distribution, which rather defeats the purpose, since one intent of the estimator is to relax distributional assumptions. An alternative approach which moves in this direction is to use a bootstrap estimator. The estimator would be

$$\text{Est.Asy.Var}[\mathbf{b}_{LAD}] \;=\; \frac{1}{R}\sum\nolimits_{r=1}^{R}\left(\mathbf{b}_{LAD,r} - \mathbf{b}_{LAD}\right)\left(\mathbf{b}_{LAD,r} - \mathbf{b}_{LAD}\right)'$$

where R is the number of repetitions, $\mathbf{b}_{LAD,r}$ is the LAD estimator obtained at the rth repetition, and $\mathbf{b}_{LAD}$ is the original LAD estimate. Each repetition is computed using a random sample of n observations, drawn with replacement, from the original sample. To obtain this estimator, add

; Nbt = ... value for *R* ...

to the **REGRESS** command.

If you do not specify bootstrap samples, no estimate of the asymptotic covariance matrix is computed. As shown below, the estimators based on the normal distribution are very simple to compute. The following data limitations are imposed on the LAD estimator:

- Number of observations up to 5,000
- Number of coefficients including the constant term, up to 15

(Both of these restrictions can be relaxed by using the **QREG** command described in the next section.)

In the following application, the LAD estimator is computed with 50 bootstrap replications using the gasoline market data. The three estimators of the asymptotic covariance matrix are computed. At the end of the results, the ordinary least squares estimator is computed, and its sum of absolute deviations is computed to compare to the LAD estimator. (Some of the output is omitted, including the initial table of statistics for the OLS estimator.)

To define the data matrix, use

```
NAMELIST    ; x = one,logpg,loginc,logpnc,logpuc,logppt $
```

This is the LAD estimator with bootstraps:

```
REGRESS     ; Lhs = logg ; Rhs = x
            ; Res = e
            ; Alg = LAD
            ; Nbt = 50 $
```

The covariance matrices are based on the normal distribution. V2 is the White estimator.

```
CREATE      ; abslad = Abs(e) $
MATRIX      ; v1 = {pi/2 * e’e/n} * <x’x>
            ; v2 = {pi/2} * <x’x> * Bhhh(x,e) * <x’x>
            ; Stat(b,v1,x)
            ; Stat(b,v2,x) $
```

Compare the sums of absolute deviations and sum of squares to OLS.

```
CALC        ; sumlad = Sum(abslad)
            ; sumlad2 = e'e $
```

Compute the ordinary least squares regression.

```
REGRESS     ; Lhs = logg ; Rhs = x
            ; Res = e $
CREATE      ; absols = Abs(e) $
```

Compare the residual sums and sums of squares for the two estimators.

```
CALC        ; List
            ; sumols = Sum(absols)
            ; sumlad
            ; sumsqdev
            ; sumlad2 $
```

```
-----------------------------------------------------------------------------
Least absolute deviations estimator...............
Nonlinear    least squares regression ............
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23849
             Number of observs.   =            52
Model size   Parameters           =             6
             Degrees of freedom   =            46
Residuals    Sum of squares       =       .115843
             Standard error of e  =        .05018
Fit          R-squared            =        .96007
             Adjusted R-squared   =        .95572
Model test   F[  5,    46] (prob) =   221.2(.0000)
Diagnostic   Log likelihood       =      84.99102
             Restricted(b=0)      =       1.25792
             Chi-sq [   5]  (prob) = 167.5(  .0000)
Info criter. Akaike Info. Criter. =      -5.87599
Not using OLS or no constant. Rsqrd & F may be < 0
Sum of absolute deviations        =     1.7685530
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Covariance matrix based on   50 replications.
Constant|   -12.0403***     2.14199    -5.62  .0000    -16.2385   -7.8420
   LOGPG|    -.03933         .04735     -.83  .4062    -.13212     .05347
  LOGINC|    1.34498***      .21206     6.34  .0000     .92935    1.76061
  LOGPNC|     .00091         .28775      .00  .9975    -.56308     .56490
  LOGPUC|    -.00616         .10190     -.06  .9518    -.20587     .19356
  LOGPPT|    -.25389         .20172    -1.26  .2082    -.64925     .14147
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Number of observations in current sample =      52
Number of parameters computed here       =       6
Number of degrees of freedom             =      46
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  Matrix|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -12.0403***    1.82527    -6.60  .0000    -15.6177    -8.4628
   LOGPG|    -.03933        .05153     -.76  .4454     -.14033     .06168
  LOGINC|    1.34498***     .17414     7.72  .0000     1.00366    1.68630
  LOGPNC|     .00091        .25002      .00  .9971     -.48911     .49093
  LOGPUC|    -.00616        .12037     -.05  .9592     -.24208     .22977
  LOGPPT|    -.25389*       .14296    -1.78  .0757     -.53409     .02631
--------+--------------------------------------------------------------------
Number of observations in current sample =      52
Number of parameters computed here       =       6
Number of degrees of freedom             =      46
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  Matrix|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -12.0403***    1.28858    -9.34  .0000    -14.5658    -9.5147
   LOGPG|    -.03933        .05061     -.78  .4372     -.13853     .05988
  LOGINC|    1.34498***     .13029    10.32  .0000     1.08961    1.60035
  LOGPNC|     .00091        .16857      .01  .9957     -.32948     .33130
  LOGPUC|    -.00616        .15982     -.04  .9693     -.31940     .30709
  LOGPPT|    -.25389**      .12255    -2.07  .0383     -.49409    -.01369
--------+--------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                   =       -.25713
             Standard deviation     =        .23849
             No. of observations    =            52  Degrees of freedom
Regression   Sum of Squares         =       2.79379           5
Residual     Sum of Squares         =       .107004          46
Total        Sum of Squares         =       2.90080          51
             Standard error of e    =        .04823
Fit          R-squared              =        .96311  R-bar squared =   .95910
Model test   F[  5,    46]          =     240.20584  Prob F > F*   =   .00000
Diagnostic   Log likelihood         =      87.05475  Akaike I.C.   = -5.95537
             Restricted (b=0)       =       1.25792  Bayes  I.C.   = -5.73022
             Chi squared [  5]      =     171.59365  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    LOGG|  Coefficient      Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -11.5997***    1.48817    -7.79  .0000    -14.5165    -8.6829
   LOGPG|    -.03438        .04202     -.82  .4174     -.11673     .04797
  LOGINC|    1.31597***     .14198     9.27  .0000     1.03769    1.59425
  LOGPNC|    -.11964        .20384     -.59  .5601     -.51916     .27989
  LOGPUC|     .03754        .09814      .38  .7038     -.15481     .22990
  LOGPPT|    -.21514*       .11656    -1.85  .0714     -.44359     .01331
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
[CALC] SUMOLS  =      1.8911144
[CALC] SUMLAD  =      1.7685530
[CALC] SUMSQDEV=       .1070036
[CALC] SUMLAD2 =       .1158431
```

E9.4 Quantile Regressions

The command for the quantile regression estimator is

QREG ; Lhs = dependent variable ; Rhs = list of regressors
; Qnt = desired quantile (0.0+ to 1.0-) $

The quantile regression estimator fits a model of the form

$$Q(y_i|\mathbf{x}_i,\theta) = \boldsymbol{\beta}'\mathbf{x}_i,\ 0 < \theta < 1,$$

where $Q(y_i|\mathbf{x}_i,\theta)$ is the θth quantile of the distribution of $y_i|\mathbf{x}_i$. The default value is $\theta = .5$, which implies the median (and will replicate the LAD estimator of the previous section). The estimator is the linear programming method – a discussion may be found in the various papers on Roger Koenker's (University of Illinois) home page.

$$\arg\min_{\beta} \sum_i \rho_\theta\left(y_i - \sum\nolimits_k \beta_k x_{ik}\right)$$

$$\rho_\theta(u) = \begin{cases} u\theta & u \geq 0 \\ u(1-\theta) & \text{else} \end{cases}$$

Further discussion of the estimation method used here and useful computer code (which was modified for our implementation) may be found in Koenker and D'Orey (1987).

The command for requesting the quantile regression estimator is

QREQ ; Lhs = dependent variable
; Rhs = independent variables $

With no other specifications, this sets $\theta = .5$, and estimates the model by median regression, which is least absolute deviations. (The usual limits on model size – about 150 parameters – millions of observations, apply. But, if you have a huge sample, chances are this is not the estimator you should be using.) You can set a specific quantile with

; Quantile = the desired value of θ

(In previous versions of *LIMDEP*, **; Quantile** would be replaced with **; Qnt**. You may still use the earlier syntax.) You may specify several quantile regressions in the same command with

; Quantile = the set of values.

In the example below, we use **; Quantile = .3,.5,.7**. Standard errors are computed using bootstrapping as described in the previous section. You may request the number of bootstrap replications with

; Nbt = desired number

Other standard options for the linear model are available, including **; Res = name** to keep the residuals, **; Keep = name** for fitted values, and so on. Hypothesis testing about the coefficients must be done with Wald statistics using matrices *b* and *varb* after estimation.

In the example below, we continue our examination of the U.S. gasoline market from 1953 to 2004. The model specification is the same as that in the previous section. The quantile regression is fit for the .3, .5, .7 quantiles – this is a typical form of application.

```
-----------------------------------------------------------------------------
Quantile Regression Model. Quantile =      .300000
Linear Programming estimation method
LHS=LOGG     Mean                 =       -.25713
             Standard deviation   =        .23619
             Number of observs.   =            52
             Minimum              =       -.79885
             t= .30000 quantile   =       -.40304
             Maximum              =        .01454
Model size   Parameters           =             3
             Degrees of freedom   =            49
Residuals    Sum of squares       =        .21863
             Standard error of e  =        .05886
Fit          R-squared            =        .93789
             PseudoR2=1-F(0)/F(b) =        .80631
Not using OLS or no constant. Rsquared may be <= 0
Functions F= Sum r(t)[y(i)-x(i)b] =       1.00341
          F0=Sum r(t)[y(i)-Qy(t)] =       5.18051
             r(t)[u]=t*u-u*[u<0].t=       .300000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    LOGG|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -9.43402***     1.66693    -5.66  .0000   -12.70115  -6.16690
  LOGINC|    1.01763***      .21236     4.79  .0000      .60141   1.43385
   LOGPG|    -.18656*        .10690    -1.75  .0809     -.39608    .02296
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

(Results for θ = .5 and .7 are omitted.)

The following exercise compares the predictions from the three quantile regressions.

QREG ; Lhs = logg ; Rhs = one,logpg,loginc ; Quantile = .2 ; Keep = loggf2 $
QREG ; Lhs = logg ; Rhs = one,logpg,loginc ; Quantile = .5 ; Keep = loggf5 $
QREG ; Lhs = logg ; Rhs = one,logpg,loginc ; Quantile = .8 ; Keep = loggf8 $
PLOT ; Lhs = loginc ; Rhs = loggf2,loggf5,loggf8 ; Fill ; Grid
; Title = Predictions from Quantile Regressions
; Vaxis = Predicted Log Consumption $

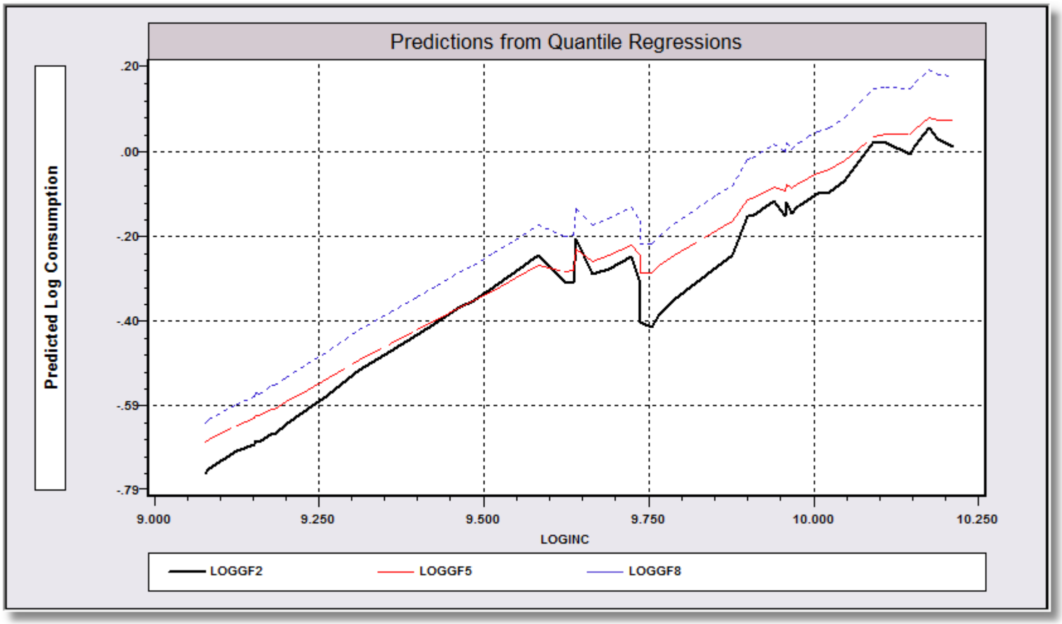


Figure E9.6 Quantile Regressions

E9.5 LOWESS

LOWESS (Locally Weighted Regression and Scatterplot Smoothing) is a nonparametric smoothing technique for examining the relationship between two variables graphically. For the LOWESS regression of a *y* on a single *x*, the technique provides a graphical device for examining the relationship. For example, continuing the gasoline market application, the following is the default results for LOWESS regression of log of consumption on log of income:

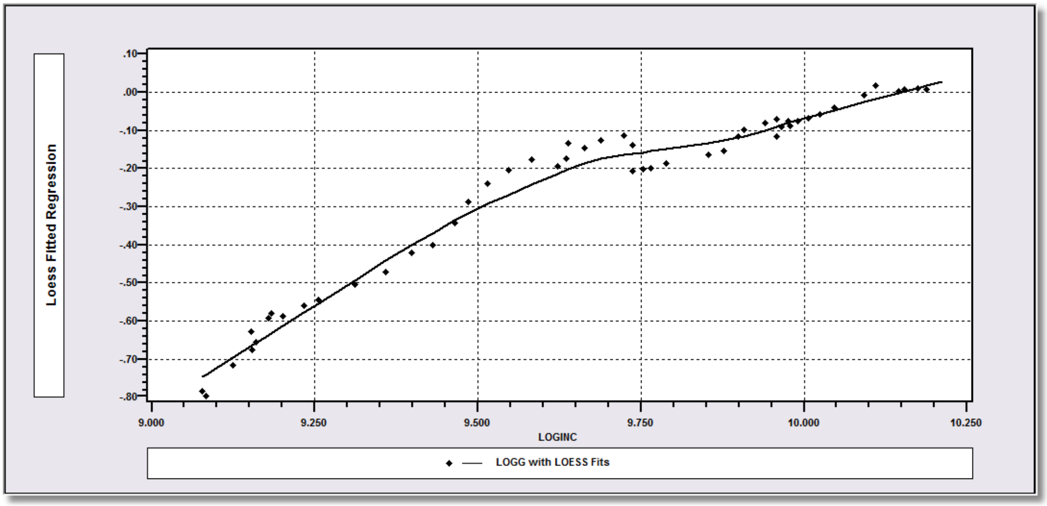


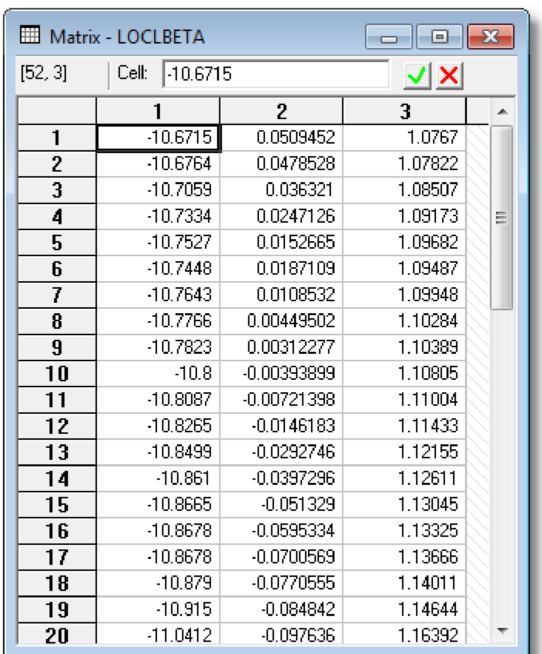
Figure E9.7 LOWESS Smoothed Fit

The technique is also used for a multiple regression, in which case, it produces a 'local' estimate of the parameter vector at each observation, using kernel methods. For the regression of *logg* on (*one*,*logpg*,*loginc*), least squares produces the following:

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                    =        -.25713
             Standard deviation      =         .23849
             No. of observations     =             52  Degrees of freedom
Regression   Sum of Squares          =        2.72390            2
Residual     Sum of Squares          =        .176898           49
Total        Sum of Squares          =        2.90080           51
             Standard error of e     =         .06008
Fit          R-squared               =         .93902  R-bar squared =    .93653
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.99007***      .58201   -15.45  .0000   -10.13078  -7.84936
   LOGPG|    -.17124***      .03789    -4.52  .0000     -.24550   -.09698
  LOGINC|     .96865***      .07376    13.13  .0000      .82408   1.11322
--------+--------------------------------------------------------------------
```

The default results for LOWESS regression produce the following. Various options are provided and post estimation analysis, including graphical methods are used to organize and interpret the findings.

```
+-------------------------------------------------+
| Locally linear weighted regression estimation |
| Sample size                   52                |
| Model size                     3                |
| Band width                .500000               |
| LOESS Sum of Squared Residuals           .02289 |
| OLS   Sum of Squared Residuals           .17690 |
| Derivatives Matrix    LOCLBETA                  |
+-------------------------------------------------+
```


Matrix - LOCLBETA

[52, 3] Cell: -10.6715

	1	2	3
1	-10.6715	0.0509452	1.0767
2	-10.6764	0.0478528	1.07822
3	-10.7059	0.036321	1.08507
4	-10.7334	0.0247126	1.09173
5	-10.7527	0.0152665	1.09682
6	-10.7448	0.0187109	1.09487
7	-10.7643	0.0108532	1.09948
8	-10.7766	0.00449502	1.10284
9	-10.7823	0.00312277	1.10389
10	-10.8	-0.00393899	1.10805
11	-10.8087	-0.00721398	1.11004
12	-10.8265	-0.0146183	1.11433
13	-10.8499	-0.0292746	1.12155
14	-10.861	-0.0397296	1.12611
15	-10.8665	-0.051329	1.13045
16	-10.8678	-0.0595334	1.13325
17	-10.8678	-0.0700569	1.13666
18	-10.879	-0.0770555	1.14011
19	-10.915	-0.084842	1.14644
20	-11.0412	-0.097636	1.16392

Figure E9.8 Matrix Result from LOWESS

E9.5.1 Graphical Smoothing with LOWESS

The command for describing a single variable is

LOWESS ; Lhs = dependent variable
; Rhs = independent variable $

The optional specifications are

; Alg = linear, quadratic or cubic

(see the technical details below),

; Bandwidth = the value

and **; Keep = name** to retain predictions
; Res = name to retain residuals

The bandwidth is used to compute the kernel based estimator. You can analyze up to five variables simultaneously by including them as a set of Lhs variables. Figure E9.9 shows the relationship of three transport related price indices to the price of gasoline

LOWESS ; Lhs = pnc,puc,ppt ; Rhs = gasp $

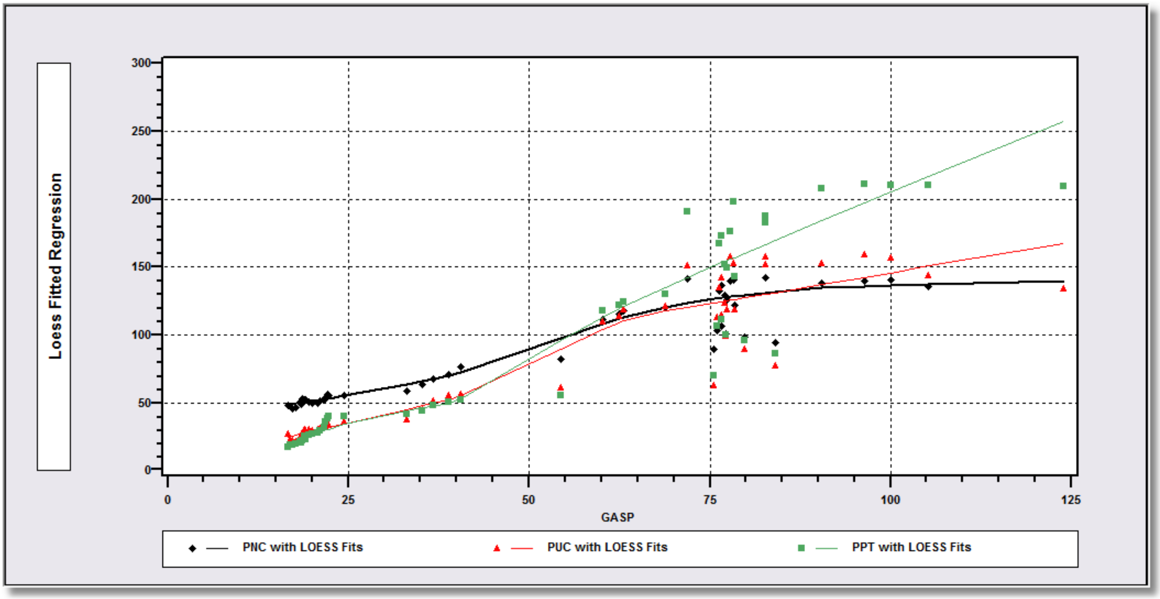


Figure E9.9 LOWESS Fits

The predictions and residuals are not computed when there is more than one Lhs variable.

The bandwidth can be specified as a single value or as a sequence of values using

; Bandwidth = lowest (increment) highest

For example, the analysis below searches for the best fit using the nine values in .1(.1).9. The best fit is found with the lowest value. (That is to be expected). A lower bandwidth, all else equal, will force the kernel estimator to track the data better.)

```
Grid search over bandwidth for lowest sum of squares
Bandwidth =  .10000, LOWESS sum of squares =   .102748E-01
Bandwidth =  .20000, LOWESS sum of squares =   .156447E-01
Bandwidth =  .30000, LOWESS sum of squares =   .309022E-01
Bandwidth =  .40000, LOWESS sum of squares =   .441540E-01
Bandwidth =  .50000, LOWESS sum of squares =   .542459E-01
Bandwidth =  .60000, LOWESS sum of squares =   .615275E-01
Bandwidth =  .70000, LOWESS sum of squares =   .747685E-01
Bandwidth =  .80000, LOWESS sum of squares =   .916489E-01
Bandwidth =  .90000, LOWESS sum of squares =   .125457
```

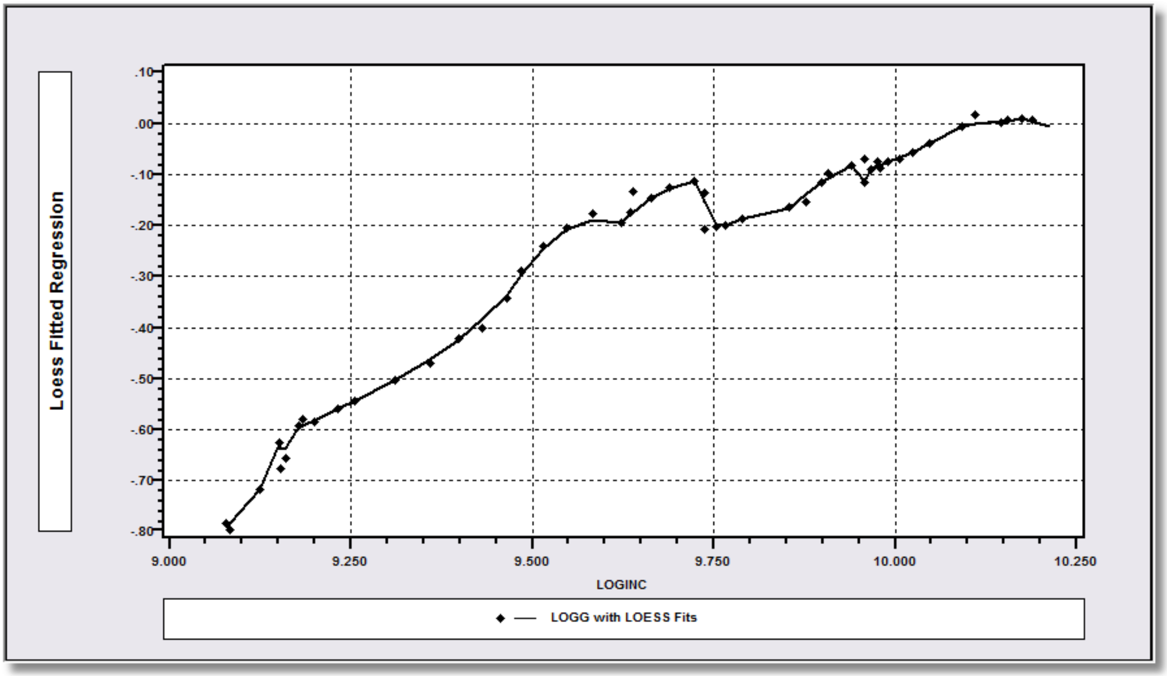


Figure E9.10 Best LOWESS Fit

E9.5.2 Local Multiple Regression

The command for local multiple regression is

LOWESS ; Lhs = dependent variable
; Rhs = list of more than one independent variables $

The options for this procedure are

; Bandwidth = value or lowest (increment) highest

and **; Keep = name** to retain predictions
; Res = name to retain residuals

No graphical output is produced. Results consist of the brief tabular summary of the computation and the matrix *loclbeta* which contains the $n \times K$ matrix of local derivatives of the function relating the dependent variable to the regressors.

We continue (conclude) the gasoline market application with

LOWESS ; Lhs = logg
; Rhs = one,loginc,logpg,logpnc,logpuc,logppt
; Bandwidth=.1(.1).9 $

```
Grid search over bandwidth for lowest sum of squares
Bandwidth =  .10000, LOWESS sum of squares =   .522288E+36
Bandwidth =  .20000, LOWESS sum of squares =   .481805E+19
Bandwidth =  .30000, LOWESS sum of squares =   .294807E-02
Bandwidth =  .40000, LOWESS sum of squares =   .348644E-02
Bandwidth =  .50000, LOWESS sum of squares =   .512996E-02
Bandwidth =  .60000, LOWESS sum of squares =   .944745E-02
Bandwidth =  .70000, LOWESS sum of squares =   .124064E-01
Bandwidth =  .80000, LOWESS sum of squares =   .193199E-01
Bandwidth =  .90000, LOWESS sum of squares =   .236876E-01

+-----------------------------------------------+
| Locally linear weighted regression estimation |
| Sample size                52                 |
| Model size                  6                 |
| Band width            .300000                 |
| LOESS Sum of Squared Residuals         .00295 |
| OLS   Sum of Squared Residuals         .10700 |
| Derivatives Matrix   LOCLBETA                 |
+-----------------------------------------------+
```

E9.5.3 Technical Details for LOWESS Computations

The calculations for LOWESS are presented in Cleveland (1979). The computations differ slightly for the single variable case in Section E9.5.1 (Case 1) and the multiple regressor case in Section E9.5.2 (Case 2). The flow of computations is as follows:

1. For each Lhs variable, the following iterations are computed:

   A. A set of weights, Δ_i is initialized at 1.0

   For each observation i, the following are assembled:
   For each observation j, $D(j|i)$ = the distance between x_i and x_j.
   This is either $D(j|i) = |x_i - x_j|$ for case 1 or $[(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)]^{1/2}$ for case 2.

   h_i = the distance to the nearest neighbor to $\mathbf{x}_i$.
   The bandwidth is used to define the width of the interval for this nearest neighbor calculation.
   For each observation j, $U(j|i) = D(j|i)/h_i$.
   Tricube weights $W(j|i) = [1 - |U(j|i)|^3]^3 \times \Delta_i$.

   B. We now compute the weighted regression of y on either $(1,x,x^2,x^3)$ in Case 1, or on $\mathbf{x}$ in Case 2, with weights $W(j|i)$. The cubic regression is the default in Case 1. You may specify the linear or quadratic regression with **; Alg = linear** or **; Alg = quadratic**. This produces coefficients $\mathbf{b}(i)$. We store the prediction, $\hat{y}_i$ and residual $e_i = y_i - \hat{y}_i$. For case 2, we store $\mathbf{b}(i)$ in row i of *loclbeta*.

   C. Update Δ_i. Let $v_i = |y_i - \hat{y}_i|$. M_v = the median value of v_i then $U_i = e_i/(6M_v)$. Δ_i is replaced with zero or $(1 - U_i^2)^2$ if $|U_i| < 1$. We return to Step A with the updated Δ_i. Cleveland recommends iterating between Steps (A,B) and C. We do a single iteration, then collect the results.

2. For case 1, the result of the computations is a plot of each $\hat{y}_i$ and y_i against x_i. For multiple Lhs variables, the plots are produced in the same figure.

When more than one bandwidth is specified, the entire procedure is computed (silently) for each value, then the results are presented for the bandwidth that results in the lowest sum of squared LOWESS residuals.

E10: Heteroscedasticity and GARCH Models

E10.1 Introduction

This chapter will detail the methods of testing for and estimating with heteroscedasticity in the linear regression model. The underlying model is

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

$$E[\varepsilon_i \mid \mathbf{x}_i] = 0,$$

$$\text{Var}[\varepsilon_i \mid \mathbf{x}_i] = \sigma^2\omega_i,\ i = 1,\ldots,n.$$

E10.2 Correcting the OLS Covariance Matrix

Heteroscedasticity in linear regression is modeled with different forms of

```
REGRESS    ; Lhs = dependent variable
           ; Rhs = independent variables
           ; Heteroscedasticity ; other specifications
           ; Wts = weighting variable $
```

Under the assumptions above, the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

is consistent, and has covariance matrix

$$\text{Var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \boldsymbol{\Sigma}.$$

where $\boldsymbol{\Omega} = \text{diag}[\omega_1,\ldots,\omega_n]$. The usual estimator,

$$\mathbf{V} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

may not be consistent if the variables in $\mathbf{x}\otimes\mathbf{x}$ are correlated with the observation specific variances, ω_i. (See Greene (2011).) White's (1980) consistent estimator of $\boldsymbol{\Sigma}$ is

$$\mathbf{S}_{\text{WHITE}} = \text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \sum_{i=1}^{n} e_i^2 \mathbf{x}_i\mathbf{x}_i' \times (\mathbf{X}'\mathbf{X})^{-1}.$$

For the underlying theory of this estimator, see White (1980) or Greene (2011). *LIMDEP* will produce this estimator as part of the **REGRESS** procedure if the command includes

; Heteroscedasticity (or, just **; Het**)

The usual set of OLS results is given, but with the revised, robust covariance matrix. (Note, this does not change the coefficient estimates. Also, it does not necessarily lead to larger (or smaller) estimated standard errors.)

Davidson and MacKinnon (1993) and Horn, Horn and Duncan (1975) have recommended three alternative forms of the White estimator which appear to perform well in small to moderate sized samples. Use

; Het ; Hc1 to use $\text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \dfrac{n}{n-K} \sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i' \times (\mathbf{X}'\mathbf{X})^{-1}$

; Het ; Hc2 to use $\text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \sum_{i=1}^{n} \dfrac{e_i^2}{\left(1-\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\right)} \mathbf{x}_i \mathbf{x}_i' \times (\mathbf{X}'\mathbf{X})^{-1}$

; Het ; Hc3 to use $\text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \sum_{i=1}^{n} \dfrac{e_i^2}{\left(1-\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\right)^2} \mathbf{x}_i \mathbf{x}_i' \times (\mathbf{X}'\mathbf{X})^{-1}$

(They recommend HC3 as their preferred estimator.)

MacKinnon and White (1985) have recommended a modification of HC3. Define

$$\mathbf{x}_i^* = \frac{e_i}{\left(1-\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\right)} \times \mathbf{x}_i \quad \text{and} \quad \overline{\mathbf{x}}^* = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^*$$

Thus, each row (observation) of **X** is multiplied by e_i / [1 - $\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$]. The estimator is

$$\text{Est.Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} \times \frac{n-1}{n}\left[\mathbf{X}^{*\prime}\mathbf{X}^* - n\overline{\mathbf{x}}^*\overline{\mathbf{x}}^{*\prime}\right] \times (\mathbf{X}'\mathbf{X})^{-1}$$

This estimator is not built in, but it can be computed as follows:

```
NAMELIST   ; x = ... the list of variables, including one $
REGRESS    ; Lhs = ... ; Rhs = x ; Res = e $
MAXRIX     ; xxi = <x'x> $
CREATE     ; u = e / (1 - Qfr(x,xxi)) $
MATRIX     ; v = Bhhh(x,u) - 1/n * x'u * u'x
           ; v = {(n-1)/n} * xxi * v * xxi ; Stat(b,v,x) $
```

All results saved by these procedures are the same as usual with **REGRESS** (see Section E7.2) except:

- v*arb* is the revised estimate,
- The log likelihood function, *logl* should be ignored.

The following data are from an exercise on page 349 of Gujarati (1988). The original source is *The Economic Report of the President*, 1985. Observations pertain to the manufacturing sector of the U.S. economy.

Year	Inventory	Sales	Year	Inventory	Sales
1950	31.1	18.6	1967	84.7	46.5
1951	39.3	21.7	1968	90.6	50.2
1952	41.1	22.5	1969	98.2	53.5
1953	43.9	24.8	1970	101.6	52.8
1954	41.6	23.3	1971	102.6	55.9
1955	45.1	16.5	1972	108.2	63.0
1956	50.6	27.7	1973	124.6	72.9
1957	51.9	28.7	1974	157.8	84.8
1958	50.2	27.2	1975	159.9	86.4
1959	52.9	30.3	1976	175.2	98.8
1960	53.8	30.9	1977	189.2	113.2
1961	54.9	30.9	1978	210.4	126.9
1962	58.2	33.4	1979	240.9	143.9
1963	60.0	35.0	1980	264.1	154.4
1964	63.4	37.3	1981	282.1	168.1
1965	68.2	41.0	1982	264.6	159.2
1966	78.0	44.9	1983	260.4	170.6

Shown below are the results of applying the procedures listed above to the model

$$Inventory = \beta_1 + \beta_2 Sales + \beta_3 Sales^2 + \varepsilon.$$

The consistently larger diagonal elements of the robust estimators suggest that the OLS computations might be somewhat optimistic.

```
CREATE    ; sales2 = sales^2 $
NAMELIST  ; x = one,sales,sales2 $
REGRESS   ; Lhs = invty ; Rhs = x ; Res = e $
MATRIX    ; xxi = <x'x> $
REGRESS   ; Lhs = invty ; Rhs = x ; Het  $
REGRESS   ; Lhs = invty ; Rhs = x ; Het ; Hc1 $
REGRESS   ; Lhs = invty ; Rhs = x ; Het ; Hc2 $
REGRESS   ; Lhs = invty ; Rhs = x ; Het ; Hc3 $
CREATE    ; u = e / (1 - Qfr(x,xxi)) $
MATRIX    ; v = Bhhh(x,u) - 1/n * x'u * u'x
          ; v = {(n-1)/n} * xxi * v * xxi
          ; Stat(b,v,x) $
```

Uncorrected

INVTY	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-2.34517	3.40086	-.69	.4956	-9.01072	4.32039
SALES	1.94812***	.10404	18.72	.0000	1.74420	2.15203
SALES2	-.00182***	.00057	-3.22	.0030	-.00293	-.00071

White

INVTY	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-2.34517	4.08316	-.57	.5699	-10.34800	5.65767
SALES	1.94812***	.13420	14.52	.0000	1.68509	2.21115
SALES2	-.00182**	.00081	-2.26	.0310	-.00340	-.00024

White HC1

INVTY	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-2.34517	4.27617	-.55	.5873	-10.72630	6.03596
SALES	1.94812***	.14054	13.86	.0000	1.67266	2.22358
SALES2	-.00182**	.00084	-2.16	.0388	-.00348	-.00017

White HC2

INVTY	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-2.34517	4.46150	-.53	.6029	-11.08955	6.39922
SALES	1.94812***	.14877	13.10	.0000	1.65654	2.23970
SALES2	-.00182*	.00091	-2.00	.0544	-.00361	-.00004

White HC3

INVTY	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-2.34517	4.89932	-.48	.6355	-11.94766	7.25733
SALES	1.94812***	.16593	11.74	.0000	1.62290	2.27333
SALES2	-.00182*	.00104	-1.76	.0885	-.00385	.00021

MacKinnon and White

Matrix	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	-2.34517	4.82659	-.49	.6270	-11.80510	7.11477
SALES	1.94812***	.16346	11.92	.0000	1.62775	2.26849
SALES2	-.00182*	.00102	-1.79	.0742	-.00382	.00018

E10.3 Estimating Models with Heteroscedasticity

There are many procedures for estimating heteroscedastic regression models. We consider two that use weighted least squares here, then more elaborate models in the next two sections.

E10.3.1 Weighted Least Squares

For the model in which ω_i is either known or has been estimated already, the weighted least squares estimator is requested with

REGRESS ; Lhs = ... ; Rhs = ... ; Wts = weighting variable $

In computing weighted estimators, we use the formulas:

n	=	the current sample size, after skipping any missing observations,
w_i	=	$(n/\Sigma_i W_i)W_i \quad = \quad Scale \times W_i$ (note that $\Sigma_i w_i = n$),
$\mathbf{b}_w$	=	$[\Sigma_i w_i \mathbf{x}_i \mathbf{x}_i']^{-1}[\Sigma_i w_i \mathbf{x}_i y_i]$,
s_w^2	=	$\Sigma_i w_i(y_i - \mathbf{x}_i'\mathbf{b}_w)^2$,
Est.Var.$[\mathbf{b}_w]$	=	$[s_w^2/(n\text{-}K)][\Sigma_i w_i \mathbf{x}_i \mathbf{x}_i']^{-1}$,

where W_i is your weighting variable. Your original weighting variable is not modified (scaled) during this computation. The scale factor is computed separately and carried through the computations.

NOTE: Apart from the scaling, your weighting variable is the reciprocal of the *individual specific variance*, not the standard deviation, and not the reciprocal of the standard deviation. This construction is used to maintain consistency with the other models in *LIMDEP*.

For example, consider the common case, $\text{Var}[\varepsilon_i] = \sigma^2 z_i^2$. For this case, you would use

```
CREATE      ; wt = 1 / z ^ 2 $
REGRESS     ; Lhs = ... ; Rhs = ... ; Wts = wt $
```

Weighted least squares is requested on the Main page of the regression command builder as shown in Figure E10.1.

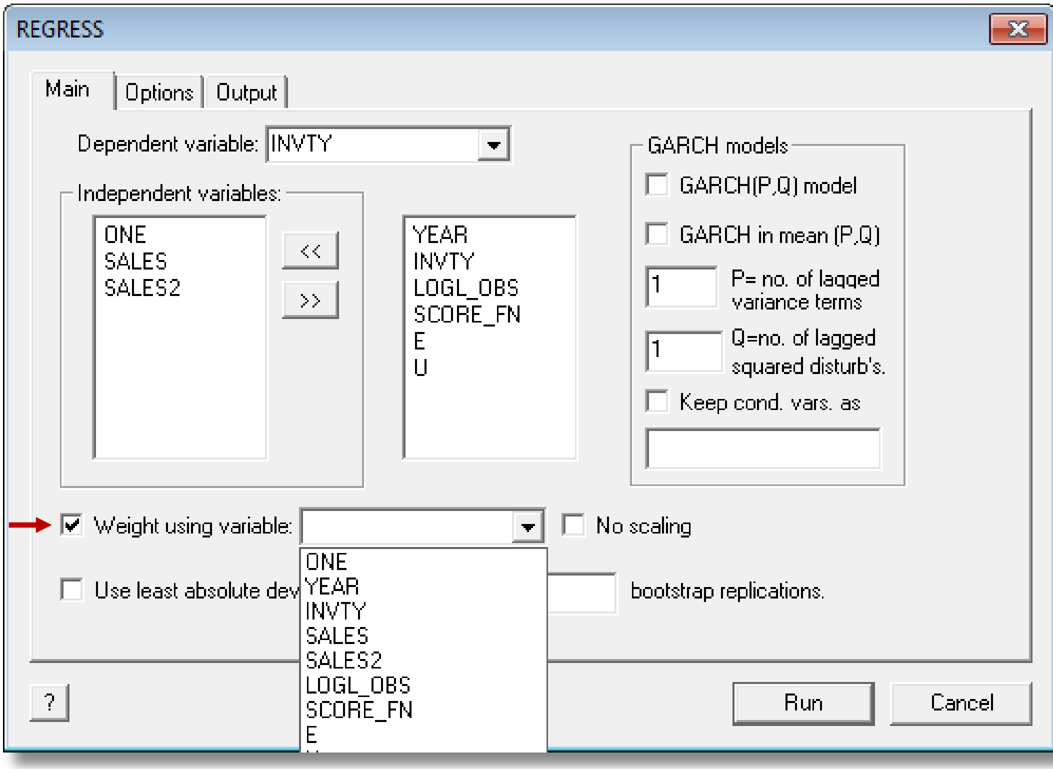


Figure E10.1 Command Builder for Weighted Least Squares

E10.3.2 Variance Proportional to the Square of the Mean

The case in which the variance is proportional to the square of the mean is simple to handle with linear regression. The model is

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \ \mathrm{Var}[\varepsilon_i] = \sigma^2[\boldsymbol{\beta}'\mathbf{x}_i]^2.$$

This can be estimated iteratively as a weighted regression as follows:

```
REGRESS     ; Lhs = y ; Rhs = x ; Keep = bx $
CREATE      ; w = 1 / bx ^ 2 $
REGRESS     ; Lhs = y ; Rhs = x ; Wts = w ; Keep = bx $
```

The second and third steps can be repeated until satisfactory convergence is achieved. A convenient approach would be to put the three lines in a procedure, then use

```
EXECUTE     ; Query $
```

E10.3.3 Testing for Heteroscedasticity

There are several different tests for heteroscedasticity that one might use. The tests due to Glejser (1965) involve regressions of squares, logs of squares or absolute values, or the absolute values themselves, of least squares residuals on a set of regressors. For present purposes, this calls for no new techniques. The special aspect of the procedure concerns using the appropriate covariance matrix to calculate the statistic. Consider some examples:

1. variance linear in **z** $\quad \text{Var}[\varepsilon_i] = \sigma^2[1 + \boldsymbol{\alpha}'\mathbf{z}_i]$,
2. standard deviation linear in **z** $\quad \text{Var}[\varepsilon_i] = \sigma^2[1 + \boldsymbol{\alpha}'\mathbf{z}_i]^2$,
3. log of variance linear in **z** $\quad \text{Var}[\varepsilon_i] = \sigma^2\exp[\boldsymbol{\alpha}'\mathbf{z}_i]$.

For these three cases, we would carry out the test by regression of the squares, absolute values, and logs of absolute values of the residuals on $\mathbf{z}_i$. A joint test of the significance of the coefficients constitutes a test of homoscedasticity. The second step regression is necessarily heteroscedastic, so we use the White estimator to compute the asymptotic covariance matrix. The following can be used: The first step is to obtain the coefficients that are in the variance functions.

```
NAMELIST   ; x = ...
           ; z = one, ... $
REGRESS    ; Lhs = y ; Rhs = x ; Res = e $
```

Use **f = e^2**, **f = abs(e)** and **f = log(e^2)** for the three functions above.

```
CREATE     ; f = ... the appropriate function of e $
REGRESS    ; Lhs = f ; Rhs = z,one ; Het $
```

Now, in each case, carry out a test of the joint hypothesis that the coefficient vector not including the constant term is zero.

```
CALC       ; m = Col(z) $
MATRIX     ; a = b(1:m) ; va = varb(1:m,1:m) ; wald = a’ <va>a $
CALC       ; Ctb(wald,m) ; 1 - Chi(wald,m) $
```

The Goldfeld-Quandt test is simple to carry out when the heteroscedasticity can be identified as a monotonic function of a single variable, z. If

$$\text{Var}[\varepsilon_i] = \sigma^2[1 + f(z_i)],$$

then, the test statistic is

$$F[\, n_1\text{-}K \,,\, n_2\text{-}K \,] = [\mathbf{e}_1'\mathbf{e}_1/(n_1\text{-}K)] \;/\; [\mathbf{e}_2'\mathbf{e}_2/(n_2\text{-}K)],$$

where group '1' is associated with high values of z_i and group '2' is associated with low values of z_i. The optimal way to split the sample, including how many observations to discard in the middle is unclear. (See Greene (2011).) With normally distributed disturbances, the statistic has an F distribution with n_1-K and n_2-K degrees of freedom. If the calculated value comes out less than one, for convenience in using the tables, we can take the reciprocal and reverse the degrees of freedom.

To carry out the test, one need only decide what values to use for the cutoffs for z_i, then

```
NAMELIST   ; x = ... $
REGRESS    ; Lhs = ... ; Rhs = x ; Res = e $
CREATE     ; d1 = zi < lower value ; e1 = d1*e
           ; d2 = zi > upper value ; e2 = d2*e $
CALC       ; df1 = d1'd1-kreg ; dfn = df1
           ; df2 = d2'd2-kreg ; dfd = df2
           ; f = (e1'e1/df1) / (e2'e2/df2)
           ; If[f < 1] ; f = 1/f ; dfn = df2 ; dfd = df1 $
CALC       ; List ; f ; 1 - Fds(f,dfn,dfd)
           ; Ftb(.95,df1,df2) $
```

For the data used in the earlier example, we used *sales* as *z* and split the sample at *sales* = 50. (The upper and lower values are 50 in the preceding routine.) The results are shown below:

```
CREATE     ; zi = sales $
REGRESS    ; Lhs = inventor ; Rhs = x ; Res = e $
...
CALC       ; List ; f ; 1 - Fds(f,dfn,dfd)
           ; Ftb(.95,df1,df2) $
```

```
F       =  .23324659539519980D+01
Result  =  .59576135098001640D-01
Result  =  .25331099831399990D+01
```

Since the sample statistic, 2.33, is less than the critical value, 2.53, the hypothesis of homoscedasticity based on the high and low values of *sales* is not rejected.

The Breusch and Pagan (1980) Lagrange multiplier test is also a simple calculation. The model is assumed to be of the form:

$$\text{Var}[\varepsilon_i] = \sigma^2 h(1 + \boldsymbol{\alpha}'\mathbf{z}_i).$$

Different normalizations involving the explicit σ^2 parameter produce the identical result. The extra parameter is actually superfluous since if $\boldsymbol{\alpha} = \mathbf{0}$, then $h(1)$ is the constant variance. As stated above, we would be assuming that $h(1) = 1$. The LM statistic is then simply one half the explained sum of squares in the regression of

$$u_i = e_i /(\mathbf{e}'\mathbf{e}/n) - 1$$

on $\mathbf{z}_i$. This statistic is always reported for the *x*s in the regression when you use the **; Het** option on the **REGRESS** command. It is also reported for the specified *z* vector when you use the **HREG** command described below. The second regression in Section E10.2 above produces

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=INVTY    Mean                    =     111.74412
             Standard deviation      =      78.44813
             Number of observs.      =            34
Model size   Parameters              =             3
             Degrees of freedom      =            31
Residuals    Sum of squares          =       1080.51
             Standard error of e     =       5.90383
Fit          R-squared               =        .99468
             Adjusted R-squared      =        .99434
Model test   F[  2,    31] (prob) =  2897.8(.0000)
White heteroscedasticity robust covariance matrix.
Br./Pagan LM Chi-sq [  2]  (prob) =  15.30 (.0005)
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
   INVTY|  Coefficient     Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -2.34517      4.08316     -.57  .5699   -10.34800   5.65767
   SALES|    1.94812***    .13420    14.52  .0000     1.68509   2.21115
  SALES2|    -.00182**     .00081    -2.26  .0310     -.00340   -.00024
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

For the more general case, the built in procedure is

REGRESS ; … ; BPT = list of variables $

For example, if we replace **; Het** with **; BPT** = sales in the preceding regression, we obtain

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=INVTY    Mean                    =     111.74412
             Standard deviation      =      78.44813
             No. of observations     =            34  Degrees of freedom
Regression   Sum of Squares          =       202005.             2
Residual     Sum of Squares          =       1080.51            31
Total        Sum of Squares          =       203086.            33
             Standard error of e     =       5.90383
Fit          R-squared               =        .99468  R-bar squared =   .99434
Model test   F[  2,    31]           =    2897.77082  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =    -107.04403  Akaike I.C.   =  3.63530
             Restricted (b=0)        =    -196.05929  Bayes  I.C.   =  3.76998
             Chi squared [  2]       =     178.03052  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]       =       2.86393  Prob C2 > C2* =   .09059
--------+--------------------------------------------------------------------
```

(The regression coefficients are the same. The standard errors differ because this second command uses the original estimator, $s^2(\mathbf{X}'\mathbf{X})^{-1}$ while the first one uses the White estimator. Using **; BPT** disables the **; Het** option.)

You can replicate the computations for the Breusch and Pagan test using the programming language. For the general test, use

```
NAMELIST    ; z = ... definition -- include one $
REGRESS     ; Lhs = y ; Rhs = ... ; Res = e $
CREATE      ; u = e*e / (sumsqdev/n) - 1 $
CALC        ; List ; lmbp =  .5 * Xss(z,u)     ? LM statistic
            ; 1 - Chi(lmbp, (Col(z)))          ? p value
            ; Ctb(.95, (Col(z)))               $ critical value from table
```

A variant due to Koenker and Bassett (1982) which allows for nonnormality is even simpler. Their 'studentized' version produces the LM statistic as nR^2 in the regression of the squared residuals on z and a constant term. Thus, this statistic is obtained with

```
REGRESS     ; Lhs  = y ; Rhs = ... ; Res = e $
CREATE      ; u = e*e  $
CALC        ; List ; lmkb =  n * Rsq(z,one,u)
            ; 1 - Chi(lmbp, (Col(z)))
            ; Ctb(.95, (Col(z))) $
```

E10.4 Multiplicative Heteroscedasticity

In this section, we modify the regression model for a specific type of heteroscedasticity,

$$\text{Var}[\varepsilon_i] = \exp(\gamma_0 + \boldsymbol{\gamma}_1'\mathbf{z}_{i1}) = \exp(\boldsymbol{\gamma}'\mathbf{z}_i).$$

(This model is developed in Harvey (1976).) We assume that the set of variables specified as $\mathbf{z}_i$ contains a constant term, so that the variance can be written in the first form when necessary. This is just for convenience. The implication is that

$$\text{Var}[\varepsilon_i] = \sigma^2\exp(\boldsymbol{\gamma}_1'\mathbf{w}_{i1}) \quad \text{and} \quad \gamma_0 = \log\sigma^2.$$

This is a general model which accommodates several kinds of heteroscedasticity. For example, the model

$$\text{Var}[\varepsilon_i] = \sigma^2 z_i^{\gamma}$$

is obtained by defining $\mathbf{z}_i$ to be [1,logz_i]. A model of groupwise heteroscedasticity for a panel of data with G groups, $\text{Var}[\varepsilon_{ig}] = \sigma_i^2$, can be produced by defining $\mathbf{z}_i$ to be a constant term and a set (minus the last one) of group specific dummy variables. By this definition,

$$\sigma_G^2 = \exp(\gamma_0),\ \sigma_i^2 = \exp(\gamma_g),\ g = 1,...,G\text{-}1.$$

The command for estimating a linear regression with this form of heteroscedasticity is

```
HREG        ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Rh2 = variables in z – do not include one in this list  $
```

It is not necessary to include a constant term in *z* since one is included automatically. Results which are saved for later use are:

Matrices:	*b*	=	estimate of $\boldsymbol{\beta}$
	varb	=	estimate of the asymptotic covariance matrix
	gamma	=	estimate of $\boldsymbol{\gamma} = [\sigma,\boldsymbol{\gamma}_1]$, (note, σ, not γ_0; $\sigma = \exp(\gamma_0/2)$)
Scalars:	*ssqrd*	=	$\exp(\gamma_0)$ = estimate of σ^2
	s	=	estimate of σ
	ybar	=	mean of Lhs variable
	sy	=	standard deviation of Lhs variable
	kreg	=	number of *x*s
	nreg	=	number of observations used
	log	=	log likelihood function

Last Model: *b_x...*
b_x... c_sigma c_z... if the command includes **; Par**

Last Function: Conditional mean = $\mathbf{b}'\mathbf{x}$

The full asymptotic covariance matrix for the estimate of $[\boldsymbol{\beta},\boldsymbol{\gamma}]$ is given below. This is a $(K+M)\times(K+M)$ block diagonal matrix. If you include

; Parameters

in the **HREG** command, *b* and *varb* include both parts of the full parameter vector.

NOTE: The first element of the '**C**' part of the saved parameter vector is $\hat{\sigma}$, not $\hat{\gamma}_0$.

The predictions for this model are the same as those for the linear regression model. But, in order to construct the confidence interval for the prediction, *LIMDEP* uses the sample mean of the **z**s, rather than the individual values, when it computes the 'σ^2' part of the forecast variance, which is $\sigma^2 + \mathbf{x}_i'\mathbf{VARB}\mathbf{x}_i$.

The method of scoring is always used for the iterations. This model does not allow you to supply starting values or to control the convergence rules for the iterations. The starting values used are **b** (OLS) for $\boldsymbol{\beta}$, $\exp(1.2704)\times s^2$ for γ_0, and **0** for $\boldsymbol{\gamma}_1$. You can specify the number of iterations with

; Maxit = maximum

NOTE: There is no need to use **; Maxit = 0** to carry out an LM test of $\boldsymbol{\gamma}_1 = \mathbf{0}$. The LM statistic is presented with the standard output in the OLS results.

The convergence rule used is given below with the technical details.

WARNING: This estimator can become unstable particularly with badly scaled data. For example, it blows up in our example below if the full sample is used. It will abort if this occurs.

E10.4.1 Results

HREG produces two sets of estimates. Starting values for the slopes are obtained by ordinary least squares. Consistent starting values for the variance parameters are obtained by regressing the logs of the squares of the least squares residuals on the variables in *z*. A consistent estimate of the unconditional variance is obtained multiplying the constant term in this regression by 1.2704. These OLS estimates of $\boldsymbol{\beta}$ and the starting values for γ_0 and $\boldsymbol{\gamma}_1$ are presented with appropriate covariance matrices. Two statistics for testing the hypothesis of homoscedasticity, Lagrange multiplier and Wald, are also presented with the initial results. Finally, the full set of maximum likelihood results is presented in the standard format.

Further details on this set of computations are given in Section E10.4.5. Two applications are presented below.

E10.4.2 Application 1 – Heteroscedastic Regression

To continue our example, we specify the multiplicative model for our inventory-sales data. One model is

$$\text{Var}[\varepsilon_i] = \sigma^2 Sales^{\gamma}$$

so $$\mathbf{z} = [one, \log(Sales)].$$

As noted earlier, the estimator diverges when the full sample is used. This may be because the variance of the disturbance appears to grow dramatically at the end of the sample.

HREG ; Lhs = invty ; Rhs = one,sales,sales2 ; Rh2 = Log(sales) $

```
HREG:Estimates diverging. Variances vanishing or exploding.
```

With this failure, we respecify the variance function as

$$\text{Var}[\varepsilon_i] = \sigma^2 \exp(\gamma\, Sales)$$

The maximum likelihood estimates are given below after the initial OLS estimates.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=INVTY    Mean                    =     111.74412
             Standard deviation      =      78.44813
             No. of observations     =            34  Degrees of freedom
Regression   Sum of Squares          =       202005.           2
Residual     Sum of Squares          =       1080.51          31
Total        Sum of Squares          =       203086.          33
             Standard error of e     =       5.90383
Fit          R-squared               =        .99468  R-bar squared =   .99434
Model test   F[  2,    31]           =    2897.77082  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =    -107.04403  Akaike I.C.   =  3.63530
             Restricted (b=0)        =    -196.05929  Bayes  I.C.   =  3.76998
             Chi squared [  2]       =     178.03052  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]       =       2.86393  Prob C2 > C2* =   .09059
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   INVTY|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -2.34517        3.45969     -.68  .4979    -9.12603   4.43570
   SALES|    1.94812***      .13287    14.66  .0000     1.68769   2.20855
  SALES2|    -.00182**       .00088    -2.08  .0378     -.00354   -.00010
   Sigma|    1.86504***      .59924     3.11  .0019      .69056   3.03952
   SALES|     .02443***      .00801     3.05  .0023      .00872    .04013
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Multiplicative Heteroscedastic Regr. Model
Dependent variable                   INVTY
Log likelihood function         -104.50980
Restricted log likelihood       -107.04403
Chi squared [   1 d.f.]            5.06847
Significance level                  .02436
McFadden Pseudo R-squared         .0236747
Estimation based on N =     34, K =   5
Inf.Cr.AIC  =  219.020 AIC/N =    6.442
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   INVTY|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Regression (mean) function
Constant|    -.74529        2.85562     -.26  .7941    -6.34220   4.85161
   SALES|    1.89012***      .09627    19.63  .0000     1.70142   2.07881
  SALES2|    -.00149***      .00057    -2.63  .0086     -.00260   -.00038
        |Variance function (log-linear)
   Sigma|    3.84664***      .78681     4.89  .0000     2.30451   5.38876
   SALES|     .00953*        .00510     1.87  .0618     -.00047    .01953
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E10.4.3 Application 2 – Groupwise Heteroscedasticity

The Grunfeld data used in several earlier examples lend themselves well to this estimator. (These are Table F10.4 in Greene (2011).The model is

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it},\ \ \varepsilon_{it} \sim \text{N}[0, \sigma^2\exp(\gamma_i)],\ \gamma_1 = 0.$$

The variance function is constructed by defining four firm specific dummy variables for the last four firms – the first is omitted. Then, the command for the model is

```
IMPORT      ; File = "C:/…/Grunfeld.dat" $
CREATE      ; Expand(firm) = d1,d2,d3,d4,d5,d6,d7,d8,d9,d10 $
NAMELIST    ; firms = d2,d3,d4,d5,d6,d7,d8,d9,d10 $
HREG        ; Lhs = i ; Rhs = one,f,c ; Rh2 = firms $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                 =      145.95825
             Standard deviation   =      216.87530
             No. of observations  =            200  Degrees of freedom
Regression   Sum of Squares       =    .760409E+07            2
Residual     Sum of Squares       =    .175585E+07          197
Total        Sum of Squares       =    .935994E+07          199
             Standard error of e  =       94.40840
Fit          R-squared            =         .81241  R-bar squared =   .81050
Model test   F[  2,    197]       =      426.57573  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    -1191.80236  Akaike I.C.   =  9.11015
             Restricted (b=0)     =    -1359.15096  Bayes  I.C.   =  9.15962
             Chi squared [  2]    =      334.69719  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       I|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|   -42.7144***     8.37643    -5.10  .0000    -59.1319   -26.2969
       F|     .11556***      .00740    15.61  .0000      .10105     .13007
       C|     .23068***      .03370     6.85  .0000      .16463     .29673
(Initial estimates of variance parameters omitted)
-----------------------------------------------------------------------------
Multiplicative Heteroscedastic Regr. Model
Dependent variable                      I
Log likelihood function      -956.68911
Restricted log likelihood   -1191.80235
Chi squared [   9 d.f.]       470.22648
Significance level               .00000
McFadden Pseudo R-squared      .1972754
Estimation based on N =    200, K =  13
Inf.Cr.AIC  = 1939.378 AIC/N =    9.697
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       I|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Regression (mean) function
Constant|   -5.33753***      .52060   -10.25  .0000    -6.35789   -4.31718
       F|     .11113***      .00415    26.77  .0000      .10300     .11927
       C|     .09772***      .00555    17.60  .0000      .08683     .10860
        |Variance function (log-linear)
   Sigma|    202.972***    32.09269     6.32  .0000     140.071    265.873
      D2|    -.06361         .44721     -.14  .8869     -.94014     .81291
      D3|    -.55857         .44721    -1.25  .2117    -1.43509     .31795
      D4|   -4.12473***      .44721    -9.22  .0000    -5.00125   -3.24820
      D5|   -4.85631***      .44721   -10.86  .0000    -5.73283   -3.97979
      D6|   -6.20559***      .44721   -13.88  .0000    -7.08211   -5.32906
      D7|   -5.86779***      .44721   -13.12  .0000    -6.74432   -4.99127
      D8|   -3.34115***      .44721    -7.47  .0000    -4.21767   -2.46463
      D9|   -4.53059***      .44721   -10.13  .0000    -5.40712   -3.65407
     D10|   -9.42288***      .44721   -21.07  .0000   -10.29940   -8.54635
--------+--------------------------------------------------------------------
```

The chi squared test in the final results rejects the hypothesis of homoscedasticity. We can extract the firm specific variances from the saved results. Recall, the retained estimates are

Matrix: *gamma* = $[\sigma = \exp(\gamma_0),\ \gamma_1,\ldots,\gamma_M]$

Scalar: *s* = σ

Therefore, to extract the parts, we can use the following:

```
CALC      ; v1 = s*s $
MATRIX    ; gamma1 = gamma(2:10)  ? You would use different subscripts
          ; v = v1*Expn(gamma1)
          ; List ;  v = [v1/v]         $ Stacks variances in a vector
```

```
       V|           1
--------+--------------
       1|      41197.6
       2|      38658.6
       3|      23566.1
       4|      666.080
       5|      320.482
       6|      83.1420
       7|      116.553
       8|      1458.24
       9|      443.875
      10|      3.33096
```

E10.4.4 Restrictions

The **HREG** estimation program does not use *LIMDEP*'s built-in function optimization routines. Therefore, restrictions on the regression parameters ($\boldsymbol{\beta}$) can be tested using the **; Test:...** specification as usual, but not imposed, by using the **; CML:...** specification as with other maximum likelihood estimators. You can estimate this model subject to linear restrictions, however, as follows: For either group of parameters, linear restrictions can be built directly into the index function. For example, to impose $\beta_2+\beta_3=1$, or $\beta_3 = 1 - \beta_2$, create $y - x_3$ to use as the Lhs variable and replace (x_2,x_3) with x_2-x_3 (one variable) on the Rhs. Similar constructions could be used to constrain the elements of $\boldsymbol{\gamma}_1$. (You should not restrict σ^2.) It is also possible to impose fixed value restrictions on $\boldsymbol{\gamma}_1$, but the method of doing so is a bit indirect. Suppose that the variance can be written as

$$\text{Var}[\varepsilon_i] = \sigma^2 \exp(c_0 f_i + \boldsymbol{\gamma}_1{}'\mathbf{w}_i) = \sigma^2 \exp(c_0 f_i) \exp(\boldsymbol{\gamma}_1{}'\mathbf{w}_i)$$

where c_0 is a fixed, known coefficient and f_i is a variable. You can fit the model in this form by treating $1/\exp(c_0 f_i)$ as if it were a weight and this were just a problem in weighted least squares. That is,

```
CREATE    ; wt  =  1 / Exp(c0 * fi)  $
HREG      ; Lhs = ... ; Rhs = ... ; Rh2 = ... ; Wts = wt $
```

The model estimation routine will sort this out internally and treat the variables in the two parts of the variance properly. (Note that this would allow you to impose nonhomogeneous restrictions on the variance parameters. For example, to impose $\gamma_2+\gamma_3=1$, the variance becomes

$$\text{Var}[\varepsilon_i] = \sigma^2 e^{\gamma_2 w_2 + (1-\gamma_2) w_3} = \sigma^2 e^{w_3} e^{\gamma_2 (w_2 - w_3)}$$

so, you would use **wt** = $1/\exp(w_3)$ in the procedure described above.)

E10.4.5 Technical Details on Computation of the HREG Model

The computations for this model are derived in Harvey's (1976) paper. The interested reader is referred to that source for background. Additional analysis appears in Greene (2011). We will sketch the computations here.

Starting values for the slopes are obtained by OLS,

$$\mathbf{b}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Since Var[ε_i] equals exp($\boldsymbol{\gamma}'\mathbf{z}_i$), an estimate of $\boldsymbol{\gamma}$ is obtained by regressing the logs of the squared residuals on $\mathbf{Z}$:

$$\mathbf{c}_0 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}$$

where $u_i = \log(e_i^2)$.

The constant term in this regression is inconsistent. To make it consistent, it is necessary to add 1.2704. These results are presented in the initial output of least squares results. The corrected covariance estimator,

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

is used for the OLS slopes. This is computed using the consistent second round estimates of the variance parameters. The estimate of $\boldsymbol{\gamma}$ is asymptotically uncorrelated with $\mathbf{b}_0$, and its asymptotic covariance matrix is

$$\mathbf{Q} = 4.9348(\mathbf{Z}'\mathbf{Z})^{-1}.$$

This is also presented with the OLS results.

We enter iteration k with $\mathbf{c}_{k-1}$ and $\mathbf{b}_{k-1}$ in hand. Compute weights and residuals

$$v_i = \exp(-\mathbf{c}_{k-1}'\mathbf{z}_i)$$

$$e_i = y_i - \mathbf{b}_{k-1}'\mathbf{x}_i.$$

Then, we regress $f_i = (e_i^2 v_i - 1)$ on $\mathbf{Z}$ (v_i is the estimate of ω_i)

to obtain $\mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{f}$.

Convergence is based on $\mathbf{d}'\mathbf{d}$. If $\mathbf{d}'\mathbf{d} < 10^{-9}$, exit the iteration. If not,

$$\mathbf{b}_k = [\Sigma_i v_i \mathbf{x}_i \mathbf{x}_i']^{-1}[\Sigma_i v_i \mathbf{x}_i y_i] \quad \text{(this is GLS)}$$

and $\mathbf{c}_k = \mathbf{c}_{k-1} + \mathbf{d}$.

At convergence, the asymptotic covariance matrix of $\mathbf{b}_k$ is the inverse matrix above while the asymptotic covariance matrix of $\mathbf{c}_k$ is

$$\text{Asy.Var.}[\mathbf{c}_k] = 2(\mathbf{Z}'\mathbf{Z})^{-1}.$$

The asymptotic covariance matrix of $\mathbf{c}_k$ and $\mathbf{b}_k$ is zero, so the full matrix is block diagonal.

A Program for the Multiplicative Heteroscedasticity Model

The preceding presents a straightforward application of the method of scoring. For the interested reader, we present a *LIMDEP* program which does the same iterations. The routine is specialized for the application above, but only the definitions of **x**, *y*, and **z** need be changed for a different application.

```
SAMPLE        ; 1-100                        $ Use only the first five firms
NAMELIST      ; x = one,f,c                  $ Variables in regression
NAMELIST      ; z = one,d2,d3,d4,d5          $ Variables in variance
CREATE        ; y = i                        $ Generic name for Lhs
REGRESS       ; Lhs = y
              ; Rhs = x
              ; Res = e                      $ Starting value for beta
MATRIX        ; beta= b                      $ Retrieve estimate of beta
CREATE        ; logesq = Log(e^2)            $ Log of squared residuals
REGRESS       ; Lhs = logesq
              ; Rhs = z                      $ Starting value for gamma
CALC          ; fix = b(1) + 1.2704          $ Constant term in gamma
MATRIX        ; cg = b ; cg(1) = fix         $ Correct bias in gamma(1)
CALC          ; delta = 1                    $ Initialize for iterations
PROCEDURE                                    ? This is just iterated FGLS
CREATE        ; vi = 1/Exp(cg'z)             ? 1 / variance
              ; w = (y - x'beta)^2 * vi - 1  $ Derivatives wrt variance
MATRIX        ; dc = <z'z> * z'w             $ b from this regression is update
CALC          ; List ; delta = dc'dc         $ Use to assess convergence
MATRIX        ; beta = <x'[vi]x>*x'[vi]y     $ GLS
MATRIX        ; cg = cg + dc                 $ Update variance parameters
ENDPROCEDURE
EXECUTE       ; while delta > .00000001      $ Convergence criterion
MATRIX        ; vc = 2*<z'z>
              ; vb = <x'[vi]x>
              ; Stat(beta,vb,x)
              ; Stat(cg,vc,z) $
```

Execution of the procedure with the Grunfeld data produces the results below.

```
DELTA   =  .36328260374394480D+01
DELTA   =  .15749871040670130D+01
DELTA   =  .56041869652277020D+00
DELTA   =  .11606002902370090D+01
DELTA   =  .77354953496737570D-01
(iterations 6 – 60 omitted)
DELTA   =  .29262536724845500D-07
DELTA   =  .10624133192229550D-07
DELTA   =  .16080737781443860D-07
DELTA   =  .58387642785196380D-08

DELTA>.00000001
```

```
-----------------------------------------------------------------------------
Number of observations in current sample =     100
Number of parameters computed here       =       3
Number of degrees of freedom             =      97
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  Matrix|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    16.0973***     5.41199    2.97  .0029      5.4900    26.7046
       F|      .10303***     .00809   12.74  .0000      .08718     .11888
       C|      .04354***     .00968    4.50  .0000      .02457     .06250
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Number of observations in current sample =     100
Number of parameters computed here       =       5
Number of degrees of freedom             =      95
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  Matrix|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    11.0692***      .31623   35.00  .0000     10.4494    11.6890
      D2|    -.39363         .44721    -.88  .3788    -1.27015     .48289
      D3|   -1.19920***      .44721   -2.68  .0073    -2.07572   -.32268
      D4|   -4.15204***      .44721   -9.28  .0000    -5.02856  -3.27551
      D5|   -6.57499***      .44721  -14.70  .0000    -7.45151  -5.69846
--------+--------------------------------------------------------------------
```

E10.5 ARCH(m) and GARCH(m) Models

Engle's (1982) original model of autoregressive conditional heteroscedasticity, ARCH(1),

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t$$

$$\varepsilon_t = v_t[\sigma_0 + \alpha_1\varepsilon_{t-1}^2]^{1/2}$$

$$v_t \sim N[0,1],$$

has provided a foundation in the literature on volatility in financial markets. The model has since been generalized in many directions. The most straightforward extension is the ARCH(q) model,

$$\text{Var}[\varepsilon_t \mid \varepsilon_{t-1}, ..., \varepsilon_{t-q}] = \sigma_0^2 + \sum_{q=1}^{Q} \alpha_q \varepsilon_{t-q}^2 .$$

An important variation on the ARCH theme is the generalized ARCH model, or GARCH(p,q) model,

$$\text{Var}[\varepsilon_t \mid \varepsilon_{t-1}, ..., \varepsilon_{t-q}] = \sigma_t^2 = \sigma_0^2 + \sum_{s=1}^{q} \alpha_s \varepsilon_{t-s}^2 + \sum_{r=1}^{p} \delta_r \sigma_{t-r}^2$$

The command for estimating a linear regression with ARCH or GARCH disturbances is

```
REGRESS     ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Model = GARCH(p,q)  $
```

Estimation of these models can be done by two step (or iterated) weighted least squares or by (approximate) maximum likelihood. (With this variance specification, the model could not be based on the normal distribution save for some special cases, so the analysis is viewed as approximate. See Bollerslev (1986) and Greene (2011) for discussion.) Maximum likelihood has become the standard approach in recent years. (See Fiorentini, Calzolari, and Panattoni (1996) and McCullough and Renfro (1999) for discussion and analysis.) Finally, an interesting innovation by Engle, Lilien, and Robins (1987) is the ARCH(*m*), which we generalize here to the 'GARCH in mean,' or GARCH(*m*) model in which the variance appears in the conditional mean function:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \lambda\sigma_t^2 + \varepsilon_t$$

$$\mathrm{Var}[\varepsilon_t \mid \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}] = \sigma_t^2 = \sigma_0^2 + \sum_{s=1}^{q} \alpha_s \varepsilon_{t-s}^2 + \sum_{r=1}^{p} \delta_r \sigma_{t-r}^2$$

The ARCH and GARCH models are discussed in Sections E10.5.1 and E10.5.2. The GARCH(*m*) model is discussed in Section E10.5.3. Finally, technical details on estimation are presented in Section E10.5.4.

The GARCH(*p*,*q*) model may be fit by maximum likelihood for any *p* and *q* by using

REGRESS	**; Lhs = ... dependent variable**	
	; Rhs = ... independent variables	(May be just the constant term, *one*)
	; Model = GARCH (p,q) $	(You provide *p* and *q*)

The ARCH model is specified by providing a value of 0 for *p*.

The command builder for this model appears on the Main page for the linear regression model: Model:Linear Models/Regression.

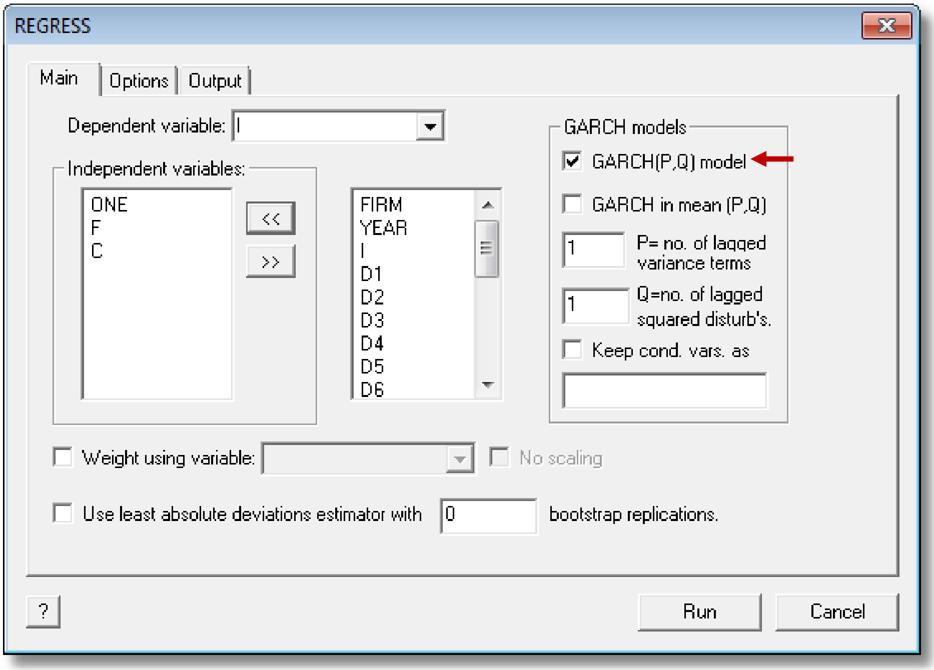


Figure E10.2 Command Builder for ARCH and GARCH Models

Output for this estimator consists of the ordinary least squares regression used to obtain the starting values for the slope parameters followed by the maximum likelihood estimators of the models parameters. Two additional statistics are included in the results:

Equilibrium variance $= \bar{\sigma}^2 = \sigma_0^2 / (1 - \alpha_1 - ... - \alpha_q - \delta_1 - ... - \delta_p)$

Wald statistic for the hypothesis of GARCH(0,0) $= \mathbf{c}'\{\text{Est.Asy.Var}[\mathbf{c}]\}^{-1}\mathbf{c}$
where **c** is the vector of estimates of the GARCH parameters.

Standard results available for later analysis include:

Matrices:

b	=	estimates of $\boldsymbol{\beta}$
varb	=	estimated asymptotic covariance matrix (includes the variance parameters if you include **; Parameters** in the command)
variance	=	estimates, in order, σ_0^2, δ_1, ... δ_P, α_1, ..., α_Q, $\bar{\sigma}^2$

Scalars:

ssqrd	=	$\mathbf{e}'\mathbf{e}/T$
rsqrd	=	$1 - \mathbf{e}'\mathbf{e} / \sum_{t=1}^{T}(y_t - \bar{y})^2$ (do not use this!)
s	=	$\sqrt{ssqrd}$
rho	=	0.0
degfrdm	=	T - number of parameters in $\boldsymbol{\beta}$
sy	=	standard deviation of *y*
ybar	=	mean of *y*
kreg	=	number of *x* variables
nreg	=	total number of observations
logl	=	log likelihood
exitcode	=	the usual

Last Function: None

Fitted values and residuals are based on the regression part of the model, not the GARCH part. As such, the confidence limits listed are conditional, and are based only on the equilibrium variance. They will be somewhat narrower than would be strictly appropriate if a full accounting of all estimated parameters were included. Thus, use

; Keep = variable name to retain the regression values as predictions
; Res = variable name to retain residuals, as usual.

You may also obtain estimates of the conditional variances,

$$\sigma_t^2 = \sigma_0^2 + \sum_{q=1}^{Q}\alpha_q\varepsilon_{t-q}^2 + \sum_{p=1}^{P}\delta_p\sigma_{t-p}^2$$

Use

; Cvar = variable name to retain estimates of conditional variances.

E10.5.1 Example: ARCH(0,1) Model for Expected Inflation

We examine a model for expected inflation of the form

$$P_t^e = P_{t-1}^e + \lambda_1(P_t - P_{t-1}^e) + \lambda_2(P_{t-1} - P_{t-2}^e) + \varepsilon_t.$$

We examine the model in the context of the ARCH models. The data are from the UK, so this more or less coincides with Engle's analysis. To simplify matters, we compute the lagged values initially, discard the incomplete observations, and treat the remainder as the full sample.

READ ; Nobs = 54 ; Nvar = 2 ; Names = pa,pe ; By Variables $

```
.99 1.62 1.87 2.89 2.23 1.69 3.67 5.20 6.59 11.94 7.78 7.19 8.98 8.80 6.91
4.20 5.04 4.92 5.33 5.94 8.11 7.88 6.63 2.76  2.70 3.14 2.53 2.55 2.76 4.55
7.11 5.50 5.78 7.42 4.32 2.14 1.25 4.93 2.61  2.19 3.24 3.35 1.55 1.79 1.59
2.49 1.88 1.28 1.83 3.46 1.69 1.36 1.95 3.11
0.79 1.94 2.97 3.37 3.65 1.62 3.02 4.43 4.70  8.13 10.6 7.48 7.28 7.63 6.26
6.76 5.86 6.09 6.23 6.94 7.86 8.73 7.04 6.16  4.02 3.89 3.69 4.14 3.95 4.82
5.96 6.39 5.73 6.78 5.74 3.47 2.24 2.04 3.44  3.37 3.55 4.10 2.70 2.10 1.58
2.14 2.56 1.62 1.78 3.33 2.91 1.76 2.26 2.70
```

CREATE ; pe1 = pe[-1] ; pe2 = pe[-2] ; pa1 = pa[-1] ; pa2 = pa[-2]
; y = pe-pe1 ; x1 = pa-pe1 ; x2 = pa1-pe2 $
NAMELIST ; x = x1,x2 $

For this exercise, 'all observations' is 3 to 54.

SAMPLE ; 3-54 $
REGRESS ; Lhs = y ; Rhs = x ; Model = GARCH(0,1) $

The results of running this program follow. They provide little evidence that the ARCH model is appropriate for these data.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                    =         .01462
             Standard deviation      =        1.19653
             No. of observations     =             52  Degrees of freedom
Regression   Sum of Squares          =        46.6732           1
Residual     Sum of Squares          =        26.3433          50
Total        Sum of Squares          =        73.0165          51
             Standard error of e     =         .72586
Fit          R-squared               =         .63921  R-bar squared =   .63200
Model test   F[  1,    50]           =       88.58650  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =      -56.10402  Akaike I.C.   =  -.60311
             Restricted (b=0)        =      -82.61029  Bayes  I.C.   =  -.52806
             Chi squared [  1]       =       53.01253  Prob C2 > C2* =   .00000
White heteroscedasticity consistent Asy.Cov matrix
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|     .41206***      .15797     2.61  .0091      .10244     .72167
      X2|     .15189         .20488      .74  .4585     -.24967     .55344
--------+--------------------------------------------------------------------
Normal exit:   8 iterations. Status=0, F=    55.61610
```

```
-----------------------------------------------------------------------------
GARCH MODEL
Dependent variable                       Y
Log likelihood function        -55.61610
Restricted log likelihood      -56.10402
Chi squared [   1 d.f.]           .97584
Significance level                .32323
McFadden Pseudo R-squared       .0086967
Estimation based on N =     52, K =   4
Inf.Cr.AIC  =  119.232 AIC/N =    2.293
GARCH Model, P = 0, Q = 1
Wald statistic for GARCH =          .517
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Regression parameters
      X1|     .40228***      .06270     6.42  .0000      .27939    .52517
      X2|     .14229*        .08108     1.75  .0793     -.01662    .30120
        |Unconditional Variance
Alpha(0)|     .42588***      .11224     3.79  .0001      .20590    .64585
        |Lagged Squared Disturbance Terms
Alpha(1)|     .15219         .21167      .72  .4721     -.26267    .56704
        |Equilibrium variance, a0/[1-D(1)-A(1)]
EquilVar|     .50232**       .24966     2.01  .0442      .01300    .99165
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E10.5.2 A Benchmark GARCH(1,1) Model for Exchange Rates

The Bollerslev and Ghysels (1986) model for the daily percentage nominal returns for the Deutschemark/Pound exchange rate (BG data) have become a de facto benchmark for calibrating software for estimating GARCH models They analyzed 1974 observations, and fit a GARCH (1,1) model,

$$y_t = \mu + \varepsilon_t,\ \mathrm{Var}[\varepsilon_t] = \sigma_t^2 = \sigma_0^2 + \alpha_1\varepsilon_{t-1}^2 + \delta_1\sigma_{t-1}^2.$$

McCullough and Renfro (1999) provide the following benchmark values:

$$\mu = -0.00619041$$
$$\sigma_0^2 = 0.0107613$$
$$\alpha_1 = 0.153134$$
$$\delta_1 = 0.805974$$

They also discuss algorithms and the computation of asymptotic standard errors, which we turn to in the next section.

LIMDEP's estimates based on the BG data are as follows:

REGRESS ; Lhs = y ; Rhs = one ; Model = GARCH(1,1) $

The benchmarks are matched to all reported digits for the (1,1) model. The literature provides virtually no guidance on model formulation. For better or worse, specification searches appear to be largely ad hoc. In fact, the log likelihood function for the GARCH model is extremely complicated, and iterations will often break down. The second set of results given below show an example, where we attempt to fit a GARCH(2,2) model to the same BG data.

```
-----------------------------------------------------------------------------
OLS Starting Values for GARCH Model...............
Ordinary     least squares regression ............
LHS=Y        Mean                 =      -.01643
             Standard deviation   =       .47024
             Number of observs.   =         1974
Model size   Parameters           =            1
             Degrees of freedom   =         1973
Residuals    Sum of squares       =      436.289
             Standard error of e  =       .47024
Fit          R-squared            =       .00000
             Adjusted R-squared   =       .00000
Model test   F[  1,  1973] (prob) =     .0(*****)
Diagnostic   Log likelihood       =   -1311.09644
             Restricted(b=0)      =   -1311.09644
             Chi-sq [  1]  (prob) =   .0( 1.0000)
Info criter. Akaike Info. Criter. =     -1.50850
White heteroscedasticity consistent Asy.Cov matrix
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.01643        .05212     -.32  .7527     -.11859    .08574
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  14 iterations. Status=0, F=    1106.608

-----------------------------------------------------------------------------
GARCH MODEL
Dependent variable                     Y
Log likelihood function     -1106.60788
Restricted log likelihood   -1311.09637
Chi squared [   2 d.f.]       408.97699
Significance level               .00000
McFadden Pseudo R-squared      .1559676
Estimation based on N =   1974, K =   4
Inf.Cr.AIC  = 2221.216 AIC/N =    1.125
GARCH Model, P = 1, Q = 1
Wald statistic for GARCH =     3727.503
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Regression parameters
Constant|    -.00619        .00873     -.71  .4783     -.02330    .01092
        |Unconditional Variance
Alpha(0)|     .01076***     .00312     3.45  .0006      .00464    .01688
        |Lagged Variance Terms
Delta(1)|     .80597***     .03015    26.73  .0000      .74688    .86507
        |Lagged Squared Disturbance Terms
Alpha(1)|     .15313***     .02732     5.60  .0000      .09958    .20668
        |Equilibrium variance, a0/[1-D(1)-A(1)]
EquilVar|     .26316        .59402      .44  .6577     -.90108   1.42741
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Evidently, the GARCH(2,2) model is overparameterized. The iterations have terminated abnormally. The log likelihood function for the (2,2) model is only slightly larger than for the (1,1) model in spite of the fact that the additional parameters are a fairly substantial expansion of the model.

```
Warning   141: Iterations:current or start estimate of sigma is nonpositive
Warning   141: Iterations:current or start estimate of sigma is nonpositive
Warning   141: Iterations:current or start estimate of sigma is nonpositive
Warning   141: Iterations:current or start estimate of sigma is nonpositive
Line search at iteration 27 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
GARCH MODEL
Dependent variable                       Y
Log likelihood function      -1104.17574
Restricted log likelihood    -1311.09637
Chi squared [   4 d.f.]        413.84126
Significance level                .00000
McFadden Pseudo R-squared       .1578226
Estimation based on N =   1974, K =   6
Inf.Cr.AIC  = 2220.351 AIC/N =    1.125
GARCH Model, P = 2, Q = 2
Wald statistic for GARCH =     6133.099
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Regression parameters
Constant|-.14942D-04       .00894      .00  .9987 -.17533D-01  .17504D-01
        |Unconditional Variance
Alpha(0)|     .00709**     .00326     2.18  .0293      .00071    .01347
        |Lagged Variance Terms
Delta(1)|     .50193       .36318     1.38  .1670     -.20990   1.21375
Delta(2)|     .34624       .31458     1.10  .2710     -.27032    .96280
        |Lagged Squared Disturbance Terms
Alpha(1)|     .17699***    .05464     3.24  .0012      .06989    .28409
Alpha(2)|    -.05154       .06747     -.76  .4450     -.18379    .08071
        |Equilibrium variance, a0/[1-D(1)-A(1)]
EquilVar|     .26899     10.17952      .03  .9789   -19.68251   20.22048
--------+--------------------------------------------------------------------
```

E10.5.3 The GARCH in Mean Model

Engle, et al. (1987) found that it would be useful to relax the independence of the mean and the variance in the GARCH model. The ‘GARCH in mean’ model, or GARCH(m) model is

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \lambda\sigma_t^2 + \varepsilon_t$$

$$\text{Var}[\varepsilon_t \,|\, \varepsilon_{t-1}, ..., \varepsilon_{t-q}] = \sigma_t^2 = \sigma_0^2 + \sum\nolimits_{s=1}^{q}\alpha_s\varepsilon_{t-s}^2 + \sum\nolimits_{r=1}^{p}\delta_r\sigma_{t-r}^2$$

$$\varepsilon_t \,|\, \Psi_t \sim N[0,\sigma_t^2].$$

This model is requested with

REGRESS ; Lhs = ... ; Rhs = ...
; Model = GARCH(p,q,1)
; ... any other optional specifications $

In particular, the addition of the '1' to the GARCH(p,q) specification triggers this specification. (It is possible to extend this model to additional lags in the variance, and to nonlinear functions in the variance term in the regression. Whether or not this provides substantial value added to the specification remains to be verified. *LIMDEP* is limited to this simple specification, however.) Save for this change in the specification, the model is otherwise the same as the GARCH model.

The results below show the outcome of extending the BG model to a GARCH(1,1,1) model. The predicted values shown for the (1,1) and (1,1,1) models suggest that a somewhat better fit is obtained with the extension. The GARCH(1,1) model is shown above. The GARCH(m) (1,1,1) model is shown below. The iterations for the GARCH(m) model did not actually reach a 'clean' convergence. The likelihood function has become flat at the point reported, and no further improvement could be produced. Nonetheless, it does improve somewhat on the GARCH(1,1) model; the log likelihood function has improved. The prediction from the GARCH model is simply the mean, as there are no covariates. So, for the GARCH model with no mean term, the predicted value equals the estimated mean, -.00619041. The mean value of the Lhs variable is -.016427. For the GARCH in mean model, we use the estimated unconditional variance to form the forecasted value, so for this model,

$$
\begin{aligned}
\hat{y}_t &= \hat{\mu} + \hat{\lambda}\hat{\bar{\sigma}}^2 \\
&= .0057666 - .077164\left(\frac{.0109532}{1-.803872-.1545375}\right) \\
&= -.00145552
\end{aligned}
$$

which is a bit of an improvement. The log likelihood function has improved somewhat as well.

```
Line search at iteration 13 does not improve fn. Exiting optimization.
-----------------------------------------------------------------------------
GARCH IN MEAN MODEL
Dependent variable                    Y
Log likelihood function     -1106.05926
Restricted log likelihood   -1311.09637
Chi squared [   3 d.f.]       410.07422
Significance level               .00000
McFadden Pseudo R-squared      .1563860
Estimation based on N =   1974, K =   5
Inf.Cr.AIC  = 2222.119 AIC/N =    1.126
GARCH in mean model, P = 1, Q = 1
Wald statistic for GARCH =    13229.183
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Regression parameters
Constant|     .00577         .01441       .40  .6890     -.02247    .03401
        | GARCH in Mean Term
GRCHMean|    -.07716         .07169     -1.08  .2818     -.21768    .06336
        | Unconditional Variance
Alpha(0)|     .01095***      .00317      3.46  .0005      .00475    .01716
        | Lagged Variance Terms
Delta(1)|     .80387***      .03109     25.85  .0000      .74293    .86481
        | Lagged Squared Disturbance Terms
Alpha(1)|     .15454***      .02813      5.49  .0000      .09941    .20967
        | Equilibrium variance, a0/[1-D(1)-A(1)]
EquilVar|     .26336         .60122       .44  .6614     -.91501   1.44173
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E10.5.4 Technical Details on Estimation of the GARCH(*m*) Model

With normally distributed disturbances, the log likelihood function for the GARCH(*m*) model is

$$\log L = \frac{-1}{2}\left[\sum_{t=1}^{T}\left(\log\sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}\right)\right] - \frac{T\log(2\pi)}{2}$$

where $\varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t - \lambda\sigma_t^2$

and $\sigma_t^2 = \sigma_0^2 + \sum_{s=1}^{q}\alpha_s\varepsilon_{t-s}^2 + \sum_{r=1}^{p}\delta_r\sigma_{t-r}^2$.

To maximize the function, it is necessary to minimize the term in square brackets. We do this with *LIMDEP*'s general optimization package. Two aspects of this optimization make it more complicated than usual. First, the variances must be computed recursively, and, since they are a difference equation, must be initialized at values that will affect the ultimate solution. Second, as a consequence of the first factor, derivatives must be computed recursively as well. We turn to these considerations first, then discuss how standard errors are computed for the estimates.

The *t*th term in the function to be maximized is

$$\log L_t = \log\sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}$$

The involved part of the computation is the computation of the variances, σ_t^2, as each depends on the previous ones. *LIMDEP* does its initializations as follows: The computation involves p+q lagged values, as the pth lag of σ_t^2 itself involves q lags of ε_t. Take first the case in which there is no 'in mean' term; $\lambda = 0$. For this case, as McCullough and Renfro (1999) note, there are various approaches to initialization. We initialize all presample values of both ε_t^2 and σ_t^2 at an estimate of the unconditional variance using the then current estimate of $\boldsymbol{\beta}$;

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t\right)^2 .$$

This begins the recursion and enables us to compute the function and its derivatives. The derivatives must be computed recursively as well. Extensive detail on the procedure may be found in Fiorentini, Calzolari, and Panattoni (FCP, 1996) and in Greene (2011). An issue arises with respect to the derivatives of the initial values. *LIMDEP* accounts completely for these as well – once again, details appear in FCP. (*LIMDEP* uses analytic derivatives, not numerical approximations, for all computations in the GARCH(*m*) models.)

The GARCH(*m*) model presents a substantial complication in this set of computations. In order to compute the initial estimates of the variances, we need the estimates of the disturbances. But, the disturbances involve the variances. In order to complete the loop, we extend the assumption used in the simpler case. If the variances in all sample periods had stabilized at the same value, then, assuming that the presample disturbances take their conditional means, that variance would be

$$\bar{\sigma}^2 = \sigma_0^2 / (1 - \delta_1 - ... - \delta_p - \alpha_1 - ... - \alpha_q).$$

(Of course, there seems to be a bit of an inconsistency in assuming that the initial disturbances are zero and their squares equal the initial variances – we are setting each value to its expectation, not to a forecasted value.) Therefore, for the GARCH in mean model, we initialize the variances and squared disturbances at the revised estimate

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t - \hat{\lambda}\hat{\bar{\sigma}}^2\right)^2 .$$

where
$$\hat{\bar{\sigma}}^2 = \frac{\hat{\sigma}_0^2}{1-\hat{\delta}_1-\ldots-\hat{\delta}_p-\hat{\alpha}_1-\ldots-\hat{\alpha}_q}$$

One complication remains. If the denominator of $\hat{\bar{\sigma}}^2$ is nonpositive, the function will be nonsensical, and problems will emerge in the function evaluation. If this occurs, we revert to an alternative estimator; we recompute $\hat{\bar{\sigma}}^2$ without the α terms. If the denominator of this revised estimate is also nonpositive, the log likelihood function will be noncomputable. Before reaching this point, the trial value of the parameter vector that produced this situation would have been rejected, and we would have reentered the iteration with another set of estimates. Note, this is the situation which produces the diagnostic

```
Warning   141: Iterations:current or start estimate of sigma is nonpositive
```

which appears before our estimates of the GARCH(2,2) model earlier.

Asymptotic standard errors for the coefficient estimators are obtained via a hybrid form of the 'sandwich' (robust) covariance matrix estimator. We compute the BHHH estimator first,

$$\mathbf{B} = \sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t'$$

where $\mathbf{g}_t$ is the vector of derivatives of $\log L_t$ with respect to the full vector of parameters. We then compute

$$\mathbf{H} = -\mathrm{E}\left[\sum_{t=1}^{t}\mathbf{H}_t\right]$$

where $\mathbf{H}_t$ is the second derivatives matrix of $\log L_t$ with respect to the full vector of parameters. The actual derivatives are extremely complicated. However, the expectations have an extremely simple form. (See Bollerslev (1986).) Finally, the estimated asymptotic covariance matrix is computed as

$$\text{Est.Asy.Var}[\ .] = \mathbf{H}^{-1}\,\mathbf{B}\,\mathbf{H}^{-1}.$$

McCullough and Renfro label this the Bollerslev and Wooldridge (1992) estimator, and provide the following benchmark values for the standard errors for the BG GARCH(1,1) model: (.00873092, .00312364, .0273219, .0301509). *LIMDEP*'s values for these are (.0087309247, .0031236375, .027321934, .030150886), which agree with all reported digits.

After estimation of the structural parameters,

$$\bar{\sigma}^2 = \sigma_0^2 / (1 - \alpha_1 - \ldots - \alpha_Q - \delta_1 - \ldots - \delta_P)$$

is estimated. The standard error is estimated using the delta method.

E11: Autocorrelation in the Linear Model

E11.1 Introduction

This chapter will detail estimation of linear regression models with autocorrelated disturbances. The models presented here are of autoregressive disturbances:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t,$$

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \ldots + \rho_p\varepsilon_{t-p} + u_t\,.$$

Moving average disturbances,

$$\varepsilon_t = u_t + \lambda_1 u_{t-1} + \lambda_2 u_{t-2} + \ldots + \lambda_q u_{t-q}\,,$$

are presented in Chapter E12 under the subject of ARIMA and ARMAX models. *LIMDEP* has a somewhat limited facility for handling mixed disturbances. For the most part, mixed models can be handled by prior modification of the data (to set up the autoregressive part) and/or use of the ARIMA/ARMAX procedure.

Autocorrelation in the linear regression is modeled with different forms of

REGRESS ; Lhs = dependent variable
; Rhs = independent variables
; AR1 and/or other specifications $

E11.2 Correcting the OLS Covariance Matrix

Section E10.2 describes how to obtain a consistent covariance matrix for the OLS estimates in the presence of heteroscedasticity (the 'White estimator'). Newey and West's (1987) counterpart is consistent in the presence of generally unspecified autocorrelation. The specification for requesting the Newey-West estimator as part of a regression is

; Pds = L

where '*L*' is the number of periods for which lags are to be computed. You can also access this on the Options page of the command builder by selecting Model:Linear Models/Regression.

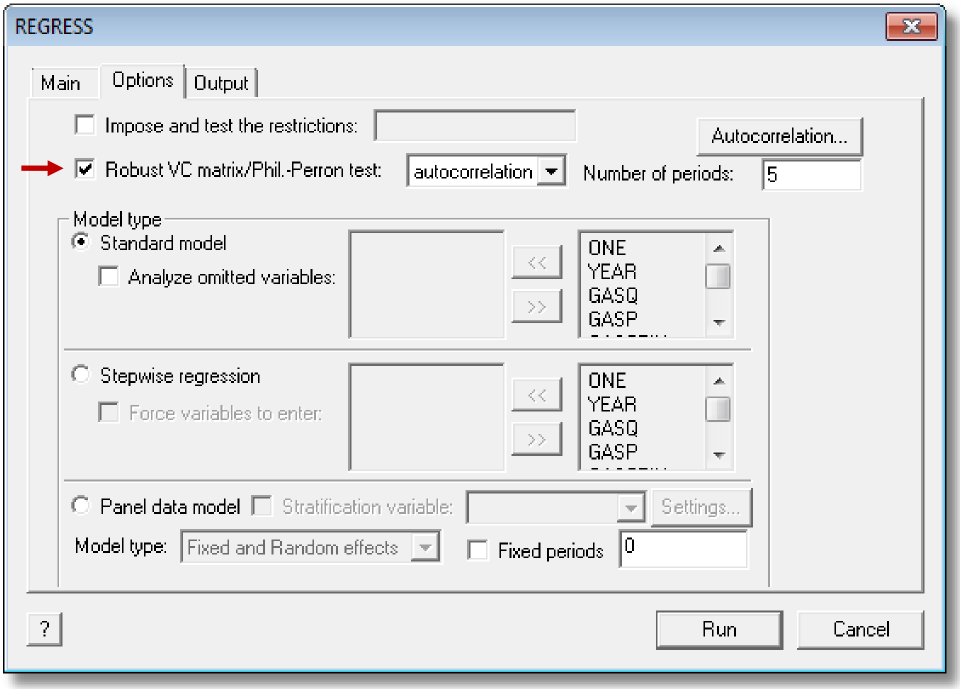


Figure E11.1 Command Builder for Linear Models with Autocorrelation

The computations are listed as follows, where **V** is the final result:

$$\mathbf{V} = \text{White estimator} + \mathbf{SRS},$$

$$\mathbf{S} = (\mathbf{X}'\mathbf{X})^{-1},$$

$$\mathbf{R} = \sum_{i=1}^{i=L} w(i,l)\mathbf{R}(i),$$

$$\mathbf{R}(i) = \sum_{t=i+1}^{t=T} [\mathbf{x}_t\mathbf{x}_{t-i}' + \mathbf{x}_{t-i}\mathbf{x}_t']e_t e_{t-i},$$

$$w(i,L) = 1 - i/(L+1) \text{ is a scalar weight,}$$

and e_t = the OLS residual for period t, t=1,...,T.

The value 'L' is the number of periods used in computing **R**. There is little theoretical guidance on the best choice of L. If the model were a moving average, L is the maximum lag. Of course, if it were known that an MA(L) model applied, this would be the wrong procedure to use in the first place. For autoregressions and mixed processes, the picture is far less clear. Readers are referred to Newey and West (1987), and for some background material, White (1981).

We will base this example and several to follow on the gasoline market data set used earlier in Chapters E7 and E8. Results of fitting a multiple regression with an autocorrelation robust covariance matrix estimator with lags of five periods are shown below. The accompanying residual plot strongly suggests the presence of autocorrelation. The rather large increase in the estimated standard errors that occurs when the Newey-West correction is applied is to be expected.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGG     Mean                 =        -.25713
             Standard deviation   =         .23849
             Number of observs.   =             52
Model size   Parameters           =              3
             Degrees of freedom   =             49
Residuals    Sum of squares       =        .176898
             Standard error of e  =         .06008
Fit          R-squared            =         .93902
             Adjusted R-squared   =         .93653
Model test   F[  2,    49] (prob) =   377.3(.0000)
Robust VC    Newey-West, Periods  =              5
Model was estimated on May 17, 2011 at 10:58:54 AM
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    LOGG|  Coefficient     Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.99007***    1.25617    -7.16  .0000   -11.45213  -6.52802
   LOGPG|    -.17124**      .07992    -2.14  .0371     -.32789   -.01459
  LOGINC|     .96865***     .15883     6.10  .0000      .65735   1.27994
--------+--------------------------------------------------------------------
Uncorrected
--------+--------------------------------------------------------------------
Constant|   -8.99007***     .58201   -15.45  .0000   -10.13078  -7.84936
   LOGPG|    -.17124***     .03789    -4.52  .0000     -.24550   -.09698
  LOGINC|     .96865***     .07376    13.13  .0000      .82408   1.11322
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

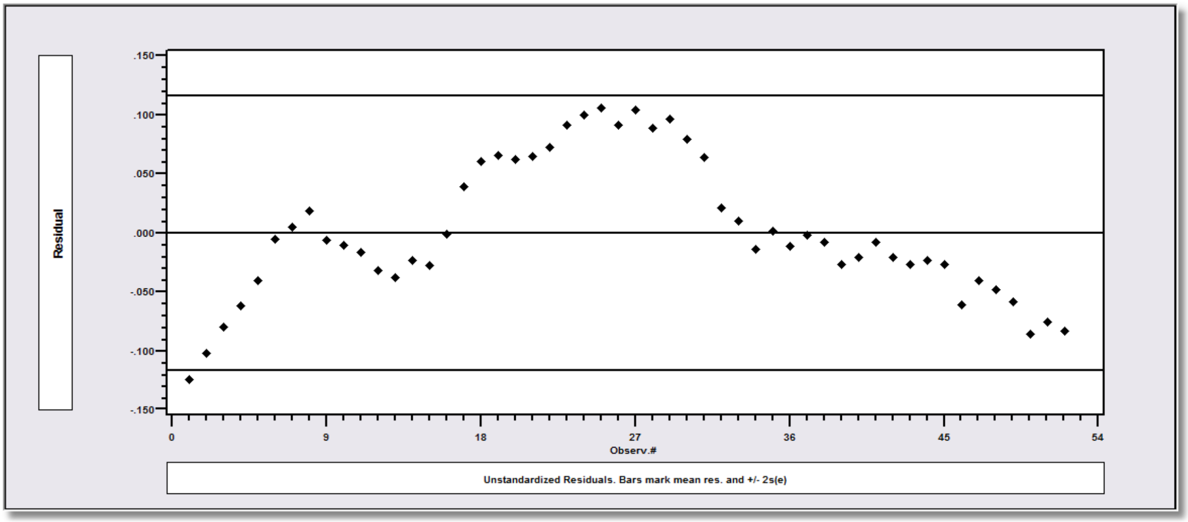


Figure E11.2 Autocorrelated Residuals

E11.3 Correcting for First Order Autocorrelation

There are numerous procedures for estimating a linear regression with first order autoregressive disturbances,

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t,$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t\,.$$

The simplest form of the command is

REGRESS ; Lhs = ... ; Rhs = ... ; AR1

The command builder for the linear regression model may also be used. Note in Figure E11.1 at the upper right of the dialog box, there is a button for Autocorrelation. This dialog box is shown in Figure E11.3:

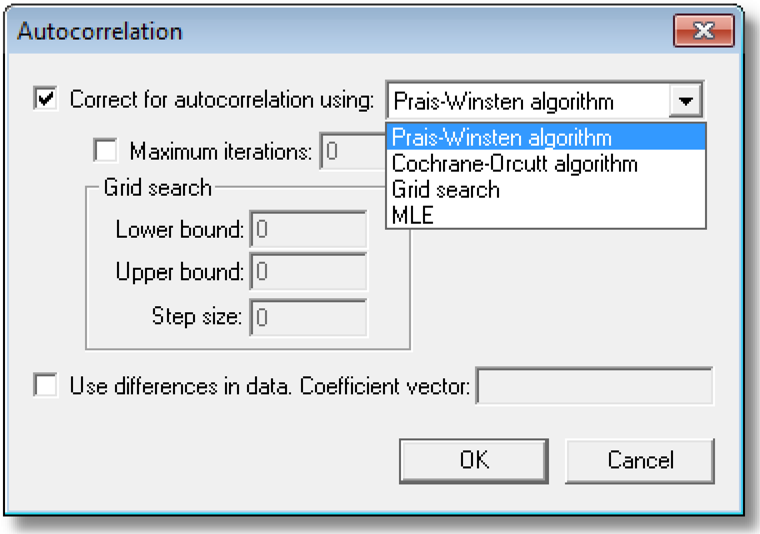


Figure E11.3 Command Builder for Specifying Autocorrelation Estimator

The default estimator is the iterative Prais-Winsten algorithm. That is, the first observation is *not* discarded; the full GLS transformation is used. This is a repeated two step estimator:

Step 1. OLS regression of **y** on **X**. Then, estimate ρ with

$$r = 1 - ½ \times \text{Durbin-Watson statistic.}$$

Step 2. OLS regression of

$$y_1{*} = (1 - r^2)^{1/2}\, y_1$$

$$y_t{*} = y_t - ry_{t-1},\ t = 2,\ldots,T$$

on the same transformation of $\mathbf{x}_t$.

After Step 2, r is recomputed based on the GLS estimator, and the regression is repeated. This iteration continues until the change in r from one iteration to the next is less than 0.0001. The covariance matrix for the slope estimators is the usual OLS estimator, $s^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$ based on the transformed data. The asymptotic variance for r is estimated by $(1 - r^2)/(T-1)$.

Results and diagnostics are presented for both transformed and untransformed models. The example below shows the specific results given.

> **NOTE:** If no other specification is given, the estimator is allowed to iterate to convergence, which usually occurs after a small number of iterations. The updated value of r at each iteration is computed from the Durbin-Watson statistic based on the most recent GLS coefficients estimates. Iterating these estimators to convergence does not produce a maximum likelihood estimator.

The ordinary least squares regression results are the same as in the previous section and are omitted.

```
-----------------------------------------------------------------------------
AR(1) Model:     e(t) = rho * e(t-1) + u(t)
Initial value of rho        =         .95624
Maximum iterations          =            100
Method = Prais - Winsten
Iter=  1, SS=        .014, Log-L=    138.050
Iter=  2, SS=        .014, Log-L=    138.346
Iter=  3, SS=        .014, Log-L=    138.276
Iter=  4, SS=        .014, Log-L=    138.227
Final value of Rho     =             .986786
Iter=  4, SS=        .014, Log-L=    138.227
Durbin-Watson:   e(t) =              .026428
Std. Deviation:  e(t) =              .104145
Std. Deviation:  u(t) =              .016875
Durbin-Watson:   u(t) =             1.317196
Autocorrelation: u(t) =              .341402
N[0,1] used for significance levels
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGG|   Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -8.25481***      .75044   -11.00  .0000    -9.72565   -6.78398
   LOGPG|    -.12339***      .02281    -5.41  .0000     -.16809    -.07868
  LOGINC|     .86578***      .07892    10.97  .0000      .71110    1.02046
     RHO|     .98679***      .02269    43.49  .0000      .94232    1.03125
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Other estimation procedures are requested by adding them to the **; AR1** request:

; AR1 ; Alg = Corc

requests the iterative Cochrane-Orcutt estimator. The first observation is skipped, and the pseudo-difference defined above is applied to the remaining observations. (We do not recommend this estimator, as it needlessly discards the information contained in the first observation, with no accompanying gain in speed, efficiency, or any statistical properties.) Alternatively,

; AR1 ; Alg = MLE

requests the maximum likelihood estimator of Beach and MacKinnon (1978). In this model, the MLE is not GLS because in addition to the generalized sum of squares, the log likelihood function contains an extra term, the Jacobian for the first observation, ½log(1 - ρ^2). This term becomes de minimis as $T\to\infty$, so in a large sample, the MLE and the other GLS estimators should not differ substantially.

TECHNICAL NOTE: The maximum likelihood estimator uses the Beach and MacKinnon method. The iteration is as shown at the beginning of this section. However, the recomputation of *r* is done differently, as follows: Let $D = (T\text{-}1)\sum_{t=2}^{T-1} e_t^2$. Then,

$$a = -\frac{(T-2)\sum_{t=2}^{T} e_t e_{t-1}}{D};\ b = \frac{(T-1)e_1^2 - T\sum_{t=1}^{T-1} e_t^2 - \sum_{t=2}^{T} e_t^2}{D};\ c = \frac{T\sum_{t=2}^{T} e_t e_{t-1}}{D}$$

Now, $p = b - a^2/3$, $q = c - ab/3 + 2a^3/27$, and $\phi = \arccos\left[\left(q\sqrt{27}\right)/\left(2p\sqrt{-p}\right)\right]$. Finally,

$\hat{\rho} = -2\sqrt{-p/3}\cos(\phi/3+\pi/3) - a/3$. Iteration of the feasible GLS procedure with this formula for *r* at each step produces the maximum likelihood estimates.

To use a grid search for the autocorrelation coefficient, use

; AR1 ; Alg = grid(lower, upper, step)

This requests a simple grid search over the indicated range with a stepsize as given. The method used for the grid search is the default Prais-Winsten estimator. To request the Cochrane-Orcutt estimator, instead, use

; AR1 ; Alg = grid(lower, upper, step, 1)

(As before, the Cochrane-Orcutt estimator is inferior to the MLE or Prais-Winsten estimator.) You can request a particular value for ρ by a simple request:

; AR1 ; Rho = specific value

When you use this form of the model command, the output will still contain an estimated standard error for the estimate of ρ, as if it had been estimated. The number of iterations allowed for the first three estimators can be controlled with the specification

; Maxit = maximum

The results saved by this estimator are the same as for the model without autocorrelation. The estimate of ρ is saved, as before, in the scalar, *rho*. Matrices *b* and *varb* contain the FGLS estimates for **β**. The **; Parameters** switch has no effect here. Residuals, predictions, and the confidence interval are the same as in the model without autocorrelation; the only adjustment is to use the GLS estimates of the residual variance and the covariance matrix of the slopes. *The set of fitted values does not contain predictions of the residuals.* Thus, if you use **; Fill** to extrapolate beyond the sample data, we do not use the BLU forecast,

$$\hat{y}_{T+1} = \mathbf{b}_{gls}'\mathbf{x}_{T+1} + \mathrm{r}e_T.$$

This can easily be constructed, if desired, with the **CREATE** command.

E11.4 Autocorrelation with a Lagged Dependent Variable

Hatanaka (1974) has derived an efficient estimator for this model which is asymptotically equivalent to maximum likelihood. The procedure is as follows: The model is

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma y_{t-1} + \varepsilon_t ,$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t .$$

Step 1. Use instrumental variables to estimate $[\boldsymbol{\beta},\gamma]$. Any consistent estimator will do. A suitable instrumental variable for the lagged value of y_t might be the lagged value of the prediction of y_t from a regression on $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$.

Step 2. Using the consistent estimates in Step 1, estimate ρ consistently by the autocorrelation of the residuals, $e_t = y_t - \mathbf{b}_{iv}'\mathbf{x}_t - c_{iv}y_{t-1}$. That is, compute the residuals using actual values, not predictions.

Step 3. Now, use the Cochrane-Orcutt transformation to do GLS based on the original data, but add an additional regressor to the model, e_{t-1}. (The transformation is not applied to the lagged residual.)

Step 4. The efficient estimate of ρ is the original estimate plus the slope on the lagged residual in the regression at Step 3. The asymptotic covariance for this estimate is that provided for the slope in Step 3. I.e., the GLS regression in Step 3 provides the full set of covariances.

This procedure uses the **2SLS** command, not **REGRESS**. The command is

```
2SLS        ; Lhs = y  ;  Rhs = ...
            ; Inst = full set of instruments ; AR1 ;  Hatanaka $
```

Note that the set of instruments includes:

1. all exogenous variables in x on the Rhs,
2. *one* if it is included in the Rhs,
3. additional instrumental variables.

For example

```
DATE        ; 1953$
PERIOD      ; 1953-2004$
CREATE      ; glag = g[-1] $
PERIOD      ; 1954-2004 $
2SLS        ; Lhs  = g
            ; Rhs  = one,gasp,pcinc,glag,pnc,puc,ppt
            ; Inst  = one,gasp,pcinc,pnc,puc,ppt,pd,pn,ps,pop
            ; AR1 ; Hatanaka $
```

```
----------------------------------------------------------------------------
Two stage    least squares regression ............
LHS=G        Mean                    =          .80034
             Standard deviation      =          .16477
             Number of observs.      =              51
Model size   Parameters              =               7
             Degrees of freedom      =              44
Residuals    Sum of squares          =     .842308E-02
             Standard error of e     =          .01384
Fit          R-squared               =          .99281
             Adjusted R-squared      =          .99183
Model test   F[  6,    44] (prob) =  1012.2(.0000)
Diagnostic   Log likelihood          =       149.70356
             Restricted(b=0)         =        20.10366
             Chi-sq [  6]  (prob) = 259.2(  .0000)
Info criter. Akaike Info. Criter. =        -8.43410
Autocorrel   Durbin-Watson Stat.     =         1.80882
             Rho = cor[e,e(-1)]      =          .09559
Not using OLS or no constant. Rsqrd & F may be < 0
Instrumental Variables:
ONE       GASP      PCINC     PNC       PUC       PPT
PD        PN        PS        POP
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       G|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|      .03643          .02380     1.53  .1258     -.01021    .08306
    GASP|     -.00097***       .00022    -4.33  .0000     -.00141   -.00053
   PCINC| .10512D-04*      .5502D-05     1.91  .0560 -.27102D-06  .21296D-04
    GLAG|      .85475***       .08541    10.01  .0000      .68735   1.02216
     PNC|     -.00090          .00060    -1.52  .1293     -.00207    .00026
     PUC|      .00103***       .00040     2.61  .0091      .00026    .00181
     PPT|     -.00045          .00031    -1.46  .1434     -.00106    .00015
--------+-------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
AR(1) Model:     e(t) = rho * e(t-1) + u(t)
Initial value of rho       =          .09559
Maximum iterations         =             100
Method = Prais - Winsten
Hatanaka 2 step estimator
Iter=  1, SS=       .008, Log-L=    151.546
Final value of Rho     =             .682554
Iter=  1, SS=       .008, Log-L=    151.546
Durbin-Watson:   e(t) =            1.796876
Std. Deviation:  e(t) =             .017963
Std. Deviation:  u(t) =             .013128
Durbin-Watson:   u(t) =            1.274928
Autocorrelation: u(t) =             .362536
N[0,1] used for significance levels
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       G|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .10495**       .04792     2.19  .0285       .01103    .19887
    GASP|    -.00114***      .00030    -3.80  .0001     -.00173   -.00055
   PCINC| .12255D-04      .7642D-05     1.60  .1088 -.27238D-05  .27234D-04
    GLAG|     .72615***      .14447     5.03  .0000       .44299   1.00930
     PNC|    -.00096         .00087    -1.11  .2664     -.00266    .00073
     PUC|     .00106**       .00050     2.11  .0349       .00008    .00205
     PPT|    -.00026         .00049     -.54  .5876     -.00122    .00069
     RHO|     .68255***      .11367     6.00  .0000       .45976    .90535
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E11.5 Differencing and Higher Order Autocorrelation

LIMDEP does not have built-in estimators for other models of autoregressive disturbances. But, the Cochrane-Orcutt method is easily generalized. Suppose the desired model is

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \dots \rho_p\varepsilon_{t-p} + u_t.$$

If consistent estimates of ρ_1,... are in hand, the counterpart to the Cochrane-Orcutt estimator will use least squares regression of

$$y_t - r_1 y_{t-1} - r_2 y_{t-2} - \dots r_p y_{t-p}$$

on the same transformation of the **x** vector. The initial p observations are lost.

LIMDEP does have a simple procedure for using lagged values in this fashion. The feature is available generally, but is likely to be most useful for estimating this model. Suppose you wish to regress

$$y_t^* = y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} - \dots - \rho_p y_{t-p}$$

on the same transformation of **x**. It is not necessary to compute the transformed variables. Use

REGRESS ; (as usual) ; Dfr = r1, r2, ..., rp $

The differences are used during estimation, but not retained during or after estimation. As always, you may specify the set of values any way you like, i.e., as a matrix, set of values, etc. *One* is not differenced by this procedure, so if you specify a constant term in the regression, it will remain after the differences. Of course, you might want to drop it at the outset, as

$$y_t = \alpha + \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t$$

implies

$$y_t - y_{t-1} = \boldsymbol{\beta}'(\mathbf{x}_t - \mathbf{x}_{t-1}) + \varepsilon_t - \varepsilon_{t-1}$$

without a constant term.

For example, if you wish to do the Cochrane-Orcutt transformation for an AR(1) model yourself, you could use the following:

```
SAMPLE      ; 1 - ... $
REGRESS     ; ... (as usual) ... $   (Saves rho)
SAMPLE      ; 2 - ... $
REGRESS     ; ... (same as above) ... ; Dfr = rho $
```

Any of the coefficients in the **; Dfr** list may be 1.0. As such, you can use this procedure to estimate equations in differences. For examples, to regress y_t - y_{t-1} on $\mathbf{x}_t$ - $\mathbf{x}_{t-1}$, use

```
            ; Dfr = 1
```

To use second differences, $\Delta^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ include **; Dfr = 2,-1** in the command.

With this in hand, the following is a procedure you can use for a *p*th order autoregression model. For our example, we use a fourth order model.

```
REGRESS     ; Lhs  = ... ; Rhs = ...
            ; Res  = e $            (Get OLS residuals)
SAMPLE      ; 5 - ... $
REGRESS     ; Lhs  = e
            ; Rhs  = e[-1],e[-2],e[-3],e[-4],one $
MATRIX      ; r = b (1 : 4) $      (Strip off constant term)
REGRESS     ; ... as above...
            ; Dfr  = r $            (Does GLS instead of OLS)
```

This method could also be used to estimate a model of fourth order autocorrelation for quarterly data. That is, one for which

$$\varepsilon_t = \rho\varepsilon_{t-4} + u_t.$$

Suitable commands would be

```
REGRESS     ; Lhs  = ...
            ; Rhs  = ...
            ; Res  = e $            (Get OLS residuals)
SAMPLE      ; 5 - ... $
REGRESS     ; Lhs  = e
            ; Rhs  = e[-4], one $
REGRESS     ; ...as above...
            ; Dfr  = 0,0,0,b(1) $
```

E11.6 Testing for Autocorrelation

For testing against the hypothesis of autocorrelated disturbances, the Durbin-Watson statistic produced with the initial output provides a rough and ready test for a fairly small class of models. Godfrey (1978) has devised a more general test for the case of *p*th order autoregressive or moving average (or a mixture). The formalities of the procedure can be found in the article. In practical terms, the test statistic can be computed by regressing the least squares residuals on the original set of regressors and *p* lagged values of the residuals. (See Greene (2011).) To do this with *LIMDEP*, use the **; Res = name** option to retain the residuals from the regression. Then use **REGRESS** to compute the least squares regression, including the original regressors and the lagged residuals on the Rhs and the current residual on the Lhs *without resetting the sample*. Then,

$$\text{chi squared } (p) = TR^2$$

is a chi squared statistic with *p* degrees of freedom. The **CALC** command can be used to obtain the significance level for the test statistic. The commands would be

```
NAMELIST    ; x = one,... $
CREATE      ; e1 = 0 ;  e2 = 0 ; ... ep = 0 $
REGRESS     ; Lhs  = ... ; Rhs = x ; Res = e $
SAMPLE      ; Set sample to p+1 to T $
CREATE      ; e1 = e[-1] ; e2 = e[-2] ... ; ep = e[...] $
SAMPLE      ; All $
REGRESS     ; Lhs  = e
            ; Rhs  = x, e1,... $
CALC        ; List
            ; lmtest = n * rsqrd
            ; List ; 1 - Chi(lmtest, (kreg-1)) $
```

The LM test requires that the initial values of the lagged residuals be filled with zeros and that these observations be included in the regression. (An alternative computation is obtained by dropping the initial observations. Authors differ on this – Godfrey, himself changes his prescription in a later article.) The commands above which create the lagged residuals set them to zero initially because *LIMDEP* would otherwise fill the lagged values with the missing value code, -999.

Finally, the Box-Pierce *Q*-statistic may be used in a similar fashion to test for higher order autocorrelation. To obtain this statistic for the OLS residuals, keep the residuals as described above, then use

```
IDENTIFY    ; Rhs  = ... residuals ; Pds = L $
```

to obtain the autocorrelations and the associated statistics. You must supply the value for *L*.

E12: ARIMA, ARMAX and Distributed Lag Models

E12.1 Introduction

This chapter will detail some of *LIMDEP*'s time series capabilities. Although *LIMDEP* is primarily oriented to cross section and panel data analysis, many common applications in time series analysis, including autocorrelation, identification, spectral analysis, unit root tests and some distributed lag models can be handled as well. (It would be possible, with some effort, to work with other time series techniques, such as unit roots, VARs, and cointegration tests. Some of these, such as basic ADF and Phillips-Perron tests are presented in Chapter E5.) This chapter will describe two estimation programs and some procedures constructed from the matrix and regression commands. Note as well that some other time series topics have been covered in earlier chapters, in particular, spectral analysis, time series identification and unit root tests in Chapter E5, GARCH models in Chapter E10 and autocorrelation in the linear regression model in Chapter E11.

E12.2 Box-Jenkins ARIMA and ARMAX Models

The models estimated by this procedure are

$$y_t = \mu + \boldsymbol{\beta}'\mathbf{x}_t + \phi_1 y_{t-1} \ldots \phi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_q\varepsilon_{t-q}, \text{ where } y_t = (1 - L)^d Y_t$$

and L is the lag operator. I.e., y_t is the original series differenced 'd' times, if this is desired. More compactly

$$\boldsymbol{\Phi}(L) \times [\delta(L)^d\, Y_t] = \mu + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\Theta}(L)\varepsilon_t,$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ are polynomials in the lag operator and $\delta(L) = (1\text{-}L)$. The nonstochastic part, $\boldsymbol{\beta}'\mathbf{x}_t$ is optional. Without it, a pure ARIMA model results. You may also specify no differences (d = 0), which will be most of the time. If you leave out $\boldsymbol{\Phi}(L)$ as well, (p = 0), a pure moving average regression results. Do note that this estimator requires that q, the order of the moving average part, be greater than zero. If q = 0, you can just estimate the equation by least squares, so there is no need for the ARMAX estimator. In this case, a diagnostic is issued and estimation is halted.

E12.2.1 Model Command

The command for this model is

ARMAX **; Lhs = dependent variable**
; Rhs = variables (optional)
; Model = p,d,q $ (in exactly that order)

If you want to include a constant term in the model (μ), *include one in the Rhs list*. Note that if d is nonzero, a nonzero constant implies a nonstationary series. You must provide all three values for p,d,q, even if they are zero. Other options are

; Res = name	to retain residuals
; Keep = name	to retain fitted values
; List	to display predicted values and residuals
; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**

You can use the model to predict beyond the end of the sample by adding

; Pds = number of periods to forecast

If there are regressors (Rhs variables) in the model, you must have valid data available for the forecasts in the rows that immediately follow the last observation in the estimating sample. If the model is a pure ARIMA model (no Rhs variables), you can forecast as many post sample periods as you like. You may also plot the fitted and actual values, including any post sample predictions by including

; Plot

in the command.

You can check the adequacy of the model by using **IDENTIFY** and/or **PLOT** to examine the residuals. To do so, be sure to include **; Res = name** to retain the residuals after estimation.

For the identification step, (i.e., for determining the orders of the lag structures in the model) here is a practical hint: If you are estimating a pure ARIMA model, use **IDENTIFY** on y_t to determine the appropriate p and q. If you are estimating a pure MA model, you need first to obtain a 'clean' set of residuals for the identification step. So, use ordinary least squares to obtain an estimate of [μ, $\boldsymbol{\beta}$], and use **IDENTIFY** on the residuals from this least squares regression. Assuming that an MA model is appropriate, the PACF from this step will reveal its order. If you are estimating an ARMAX model, in order to obtain a consistent set of residuals, you can do the following: Use instrumental variables to estimate [μ,$\boldsymbol{\beta}$], and the ϕs simultaneously. The independent variables will be *one*, the *x*s, then *y*[**-**1],...,*y*[**-***p*]. The instruments will be *one*, the *x*s, then *y***[-***q***-**1**]**,...,*y***[-***q*-*p***]**. For example, to estimate an ARMAX(1,0,1) (no differencing), your IV estimator is a 2SLS regression of *y* on *one*, the *x*s, and *y*[**-**1] with instruments *one*, the *x*s, and *y*[**-**2].

The starting values for the iterations are generated internally by this estimator. (See below.) The controls for nonlinear estimation,

; Start = list
; Tlj (j=b, f, g) [=value]

are ignored by this program. Convergence is described in the technical details in Section E12.2.4. You can set the maximum number of iterations with

; Maxit = value

if you wish. The default is 25.

HINT: This estimator normally converges in a small number of iterations. If a large number of iterations is required, there is probably a problem with the model specification. One problem that can impede convergence is overfitting, that is, making q too large.

Finally, restrictions on coefficients can *only* be imposed on the elements of $\boldsymbol{\beta}$ by building them into the model. The specification

; Rst = list

is ignored. You can test restrictions on the coefficients by two methods. First,

; Test: ...

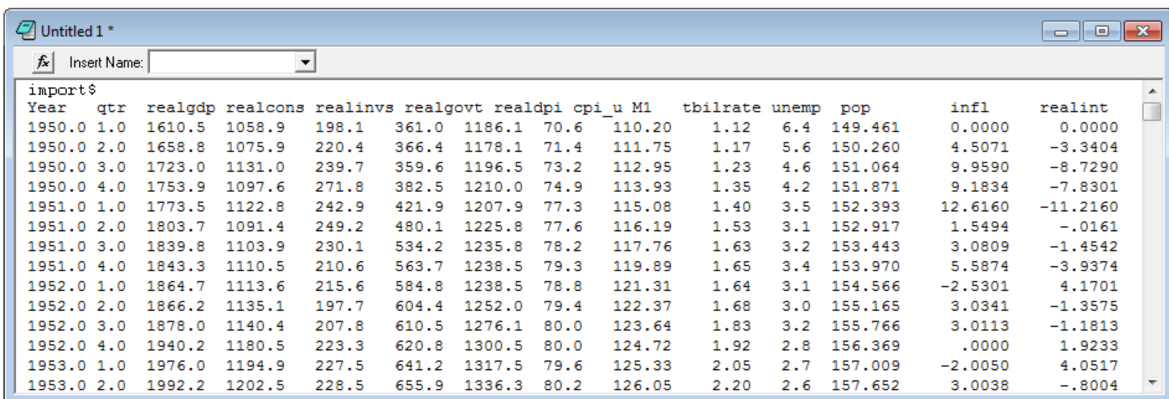
can be used for Wald tests. The coefficient vector is, in order (μ's position will correspond to *one*, in the Rhs list)

$$\gamma \ = \ \phi_1,...,\phi_p,\mu,\beta_1,...,\beta_K,\theta_1,...,\theta_q.$$

Some of these may not be present in your model. For purposes of testing hypotheses, these are b(1),...,b(M). Since this is likely to be a bit cumbersome, using the **WALD** command with the *Last Model* results is likely to be much simpler (see Section R14.4). The list of labels to use is given in the next section.

E12.2.2 Model Output

The data listed below are a some of the quarterly macroeconomic data in Greene (2011, Table F5.2). There are 204 quarterly observations, from 1950I to 2000IV. The data are quarterly observations on a number of familiar variables including real GDP, real consumption, real investment, real government spending, real disposable income, the money stock, short term interest rate, unemployment rate, population, the rate of inflation and a real interest rate. We will use these data in several examples below.


Untitled 1 *

Insert Name:

import$

Year	qtr	realgdp	realcons	realinvs	realgovt	realdpi	cpi_u	M1	tbilrate	unemp	pop	infl	realint
1950.0	1.0	1610.5	1058.9	198.1	361.0	1186.1	70.6	110.20	1.12	6.4	149.461	0.0000	0.0000
1950.0	2.0	1658.8	1075.9	220.4	366.4	1178.1	71.4	111.75	1.17	5.6	150.260	4.5071	-3.3404
1950.0	3.0	1723.0	1131.0	239.7	359.6	1196.5	73.2	112.95	1.23	4.6	151.064	9.9590	-8.7290
1950.0	4.0	1753.9	1097.6	271.8	382.5	1210.0	74.9	113.93	1.35	4.2	151.871	9.1834	-7.8301
1951.0	1.0	1773.5	1122.8	242.9	421.9	1207.9	77.3	115.08	1.40	3.5	152.393	12.6160	-11.2160
1951.0	2.0	1803.7	1091.4	249.2	480.1	1225.8	77.6	116.19	1.53	3.1	152.917	1.5494	-.0161
1951.0	3.0	1839.8	1103.9	230.1	534.2	1235.8	78.2	117.76	1.63	3.2	153.443	3.0809	-1.4542
1951.0	4.0	1843.3	1110.5	210.6	563.7	1238.5	79.3	119.89	1.65	3.4	153.970	5.5874	-3.9374
1952.0	1.0	1864.7	1113.6	215.6	584.8	1238.5	78.8	121.31	1.64	3.1	154.566	-2.5301	4.1701
1952.0	2.0	1866.2	1135.1	197.7	604.4	1252.0	79.4	122.37	1.68	3.0	155.165	3.0341	-1.3575
1952.0	3.0	1878.0	1140.4	207.8	610.5	1276.1	80.0	123.64	1.83	3.2	155.766	3.0113	-1.1813
1952.0	4.0	1940.2	1180.5	223.3	620.8	1300.5	80.0	124.72	1.92	2.8	156.369	.0000	1.9233
1953.0	1.0	1976.0	1194.9	227.5	641.2	1317.5	79.6	125.33	2.05	2.7	157.009	-2.0050	4.0517
1953.0	2.0	1992.2	1202.5	228.5	655.9	1336.3	80.2	126.05	2.20	2.6	157.652	3.0038	-.8004

Figure E12.1 Quarterly Macroeconomic Data

The ARMAX model is an extension of the linear regression model that is completely specified by p, d, and q. *LIMDEP* will estimate the full set of parameters by nonlinear least squares procedures. Results will resemble the following:

ARMAX ; Lhs = realinvs ; Rhs = one,tbilrate
; Model = 1,0,1 $

```
-----------------------------------------------------------------------------
Model:y(t) = mu + bx + phi(1)y(t-1)...phi(p)y(t-p))
          + e(t) + theta(1)e(t-1)...theta(q)e(t-q))
      y(t) = [(1-L)^d]Y(t)  (differences))
Dependent variable                             REALINVS
Raw data were differenced d =  0 times.
Sum of squares at best estimates:         157812.826207
Estimated standard deviation of e(t):         27.881949
For diagnostic checking, use IDENTIFY with residuals.
Number of iterations completed                        6
Number of observations in the sample                204
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
REALINVS|  Coefficient     Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
  Phi(1)|    1.01824***     .00506   201.09  .0000     1.00832   1.02816
      Mu|    3.39819       4.29008      .79  .4283    -5.01021  11.80659
TBILRATE|   -1.41447**      .68390    -2.07  .0386    -2.75489   -.07406
Theta(1)|     .08009        .07060     1.13  .2566     -.05828    .21847
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The parameter vector, *b* and covariance matrix, *varb* stored for later use correspond to

$$b = [\phi_1,\ldots,\phi_p,\mu,\beta_1,\ldots,\beta_K,\theta_1,\ldots,\theta_q.].$$

NOTE: If you provide *one* as one of your regressors, and it is not first in the list, the order of [μ,$\boldsymbol{\beta}$] will correspond to your command, not the preceding.

The retrievable results are:

Scalars: *sumsqdev* = $\Sigma_t\, \hat{\varepsilon}_t^2$ (the first p+q residuals are zero)

ssqrd $= \hat{\sigma}_\varepsilon^2 = \dfrac{\sum_{t=p+q+1}^{T}\hat{\varepsilon}_t^2}{T-p-q} - \left[\dfrac{\left(\sum_{t=p+q+1}^{T}\hat{\varepsilon}_t\right)}{T-p-q}\right]^2$

kreg $= K$

nreg $= T - p - q$

Last Model: *phi1,...phip,mu,b_variables...,theta1,...thetaq*, labels in order

The predictions are computed as

$$\hat{y}_t = m + f_1 y_{t-1} + \ \ldots\ + f_p y_{t-p} + \mathbf{b}'\mathbf{x}_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}\,.$$

I.e., these are static forecasts which assume the current disturbance is zero. If the series has been differenced, the forecasts are integrated back to produce the forecasts of Y_t, not y_t.

E12.2.3 Examples

We will fit two models to these data, one in raw form and one in logs. In each case, we used a simple [1,0,1] specification. For the first,

```
DATES     ; 1950.1 $
PERIOD    ; 1950.1 - 1980.4 $
ARMAX     ; Lhs = realgdp ; Rhs = one,m1 ; Res = e ; Model = 1,0,1$
IDENTIFY  ; Rhs = e ; Pds = 10 $
PLOT      ; Rhs = e ; Bars = 0  $  (Note, time series plot. No Lhs.)
```

```
-----------------------------------------------------------------------------
Model:y(t) = mu + bx + phi(1)y(t-1)...phi(p)y(t-p))
          + e(t) + theta(1)e(t-1)...theta(q)e(t-q))
      y(t) = [(1-L)^d]Y(t)  (differences))
Dependent variable                            REALGDP
Raw data were differenced d =  0 times.
Sum of squares at best estimates:        149310.903414
Estimated standard deviation of e(t):        34.841208
For diagnostic checking, use IDENTIFY with residuals.
Number of iterations completed                      14
Number of observations in the sample               124
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 REALGDP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  Phi(1)|     .99960***      .00839   119.17  .0000      .98316   1.01604
      Mu|    18.3515**      8.74490     2.10  .0359      1.2118   35.4912
      M1|     .05043         .10312      .49  .6248     -.15168    .25254
Theta(1)|     .28164***      .08811     3.20  .0014      .10894    .45434
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Time series identification for E
Box-Pierce Statistic =     7.1764      Box-Ljung Statistic  =     7.6471
Degrees of freedom   =         10      Degrees of freedom   =         10
Significance level   =      .7087      Significance level   =      .6633
* => |coefficient| > 2/sqrt(N) or > 95% significant.
PACF is computed using Yule-Walker equations.
```

```
----+-------------------------------+-------+-------------------------------+
Lag |  Autocorrelation Function     |Box/Prc|    Partial Autocorrelations   |
----+-------------------------------+-------+-------------------------------+
  1 |-.014 |           *|           |   .02 |-.014 |           * |           |
  2 | .126 |            |*          |  2.01 | .129 |             |*          |
  3 |-.032 |           *|           |  2.13 |-.033 |           * |           |
  4 | .028 |            |*          |  2.23 | .005 |             |*          |
  5 |-.062 |           *|           |  2.70 |-.063 |           * |           |
  6 |-.063 |           *|           |  3.19 |-.086 |           * |           |
  7 |-.100 |           *|           |  4.42 |-.102 |           * |           |
  8 |-.146 |          **|           |  7.06 |-.168 |          ** |           |
  9 |-.016 |           *|           |  7.09 | .002 |             |*          |
 10 | .026 |            |*          |  7.18 | .075 |             |*          |
----+-------------------------------+-------+-------------------------------+
```

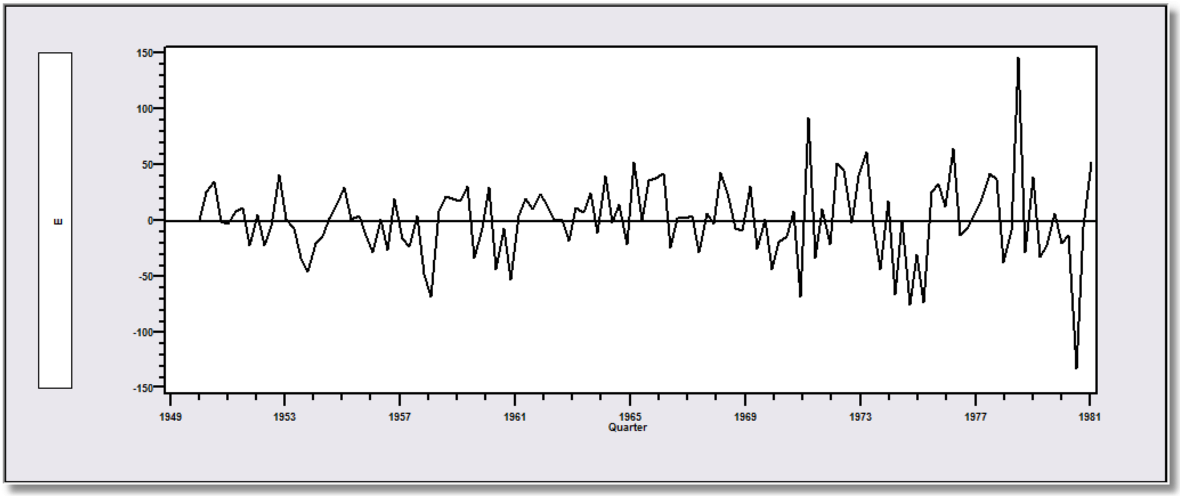


Figure E12.2 Residual Plot

PERIOD ; 1950.1 - 1997.4 $
ARMAX ; Lhs = unemp ; Rhs = one,realgdp
; Model = 1,0,1 ; Pds = 12 ; Plot ; List $

```
Maximum iterations. Exit status for parameter search = 2.
  Error   806: Maximum iterations. Exit status for parameter search = 2.

-----------------------------------------------------------------------------
Model:y(t) = mu + bx + phi(1)y(t-1)...phi(p)y(t-p))
          + e(t) + theta(1)e(t-1)...theta(q)e(t-q))
      y(t) = [(1-L)^d]Y(t)  (differences))
Dependent variable                              UNEMP
Raw data were differenced d =  0 times.
Sum of squares at best estimates:              20.110865
Estimated standard deviation of e(t):            .324488
For diagnostic checking, use IDENTIFY with residuals.
Number of iterations completed                      100
Number of observations in the sample                192
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
   UNEMP|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
  Phi(1)|      .91365***      .00686   133.19  .0000      .90021     .92710
      Mu|      .39221***      .03470    11.30  .0000      .32419     .46022
 REALGDP| .22347D-04***   .5073D-05     4.40  .0000 .12404D-04  .32290D-04
Theta(1)|      .66658***      .05201    12.82  .0000      .56464     .76852
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.

Predicted Values          (* => observation was not in estimating sample.)
Observation          Observed Y   Predicted Y   Residual     95% Forecast Interval
  1950.1             6.4000000    6.4000000      .000000      .000000      .000000
  1950.2             5.6000000    6.2766500    -.6766500      .000000      .000000
  1950.3             4.6000000    5.0961218    -.4961218      .000000      .000000
  1950.4             4.2000000    4.3034963    -.1034963      .000000      .000000
  1951.1             3.5000000    4.2001894    -.7001894      .000000      .000000
  1951.2             3.1000000    3.1635643    -.0635643      .000000      .000000
  1951.3             3.2000000    3.2232713    -.0232713      .000000      .000000
  1951.4             3.4000000    3.3415732     .0584268      .000000      .000000
  (observations omitted)
* 1999.1            No data        5.3288      No data
* 1999.2            No data        5.4569      No data
* 1999.3            No data        5.5761      No data
* 1999.4            No data        5.6891      No data
* 2000.1            No data        5.7935      No data
* 2000.2            No data        5.8917      No data
* 2000.3            No data        5.9821      No data
* 2000.4            No data        6.0657      No data
```

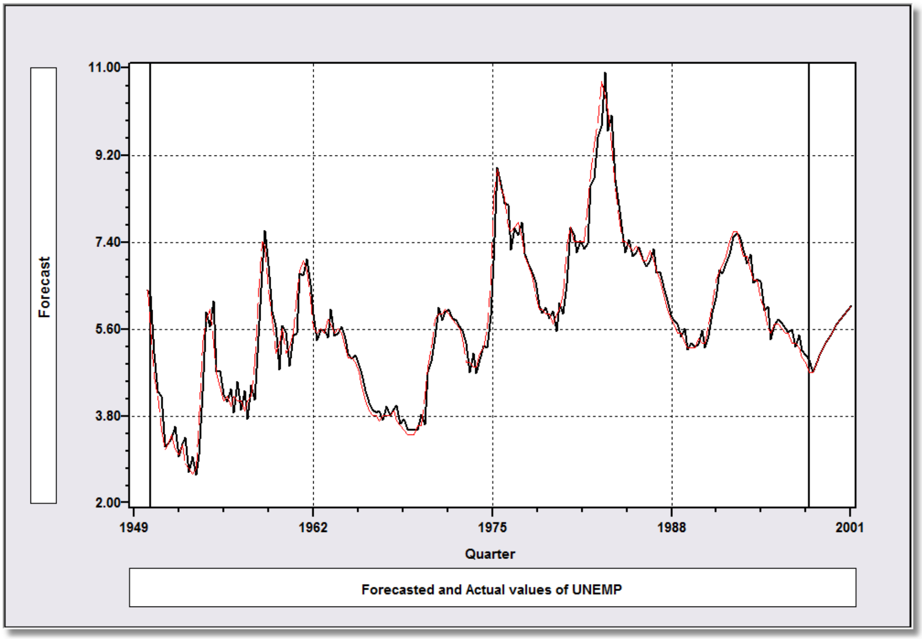


Figure E12.3 Plot of Forecasted and Actual Values of Unemployment

E12.2.4 Technical Details

Estimates are computed by nonlinear least squares. (References are Harvey (1988), Box and Jenkins (1984) and Greene (2011).) The sum of squares is

$$S(\mu,\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\theta}) \;=\; S(\gamma) \;=\; \Sigma_t\varepsilon_t^{\,2}.$$

The iterative process is as follows: Given estimates $\boldsymbol{\gamma}^k$ in hand at entry to the kth iteration,

$$\boldsymbol{\gamma}^{k+1} \;=\; \boldsymbol{\gamma}^k \;+\; [\mathbf{G}_k{'}\mathbf{G}_k]^{-1}\mathbf{G}_k\mathbf{e}_k \;=\; \boldsymbol{\gamma}^k \;+\; \boldsymbol{\delta}^k.$$

The $(T\text{-}p\text{-}q)\times M$ matrix of derivatives $\mathbf{G}$ is the time series of vectors of cross partials whose tth row is

$$\mathbf{g}_t{'} \;=\; \partial\varepsilon_t/\partial\boldsymbol{\gamma}',$$

and $\mathbf{e}_k$ is the vector of estimates of ε_t based on $\boldsymbol{\gamma}^k$. We exit the iterations normally when $\boldsymbol{\delta}'\boldsymbol{\delta}$ is less than 10^{-8} or abnormally either at maximum iterations, which you may set with

; Maxit = the maximum

or if the estimates (generally $\boldsymbol{\theta}$) diverge. The sum of squares is not unimodal, so on the way to convergence, a local minimum may be encountered. *LIMDEP* uses the parameter vector associated with the minimum sum of squares computed during the iterations.

Starting values are computed in two steps as follows:

Step 1. $[\mu,\boldsymbol{\beta},\boldsymbol{\phi}]$ are estimated by the instrumental variable method described in Section E12.2.1. The instruments for lagged ys are just deeper lagged ys, outside the range of the moving average part of the disturbance.

Step 2. $\boldsymbol{\theta}$ is initially estimated by a method suggested by Box and Jenkins. Beginning with $\boldsymbol{\theta} = \mathbf{0}$, we compute a vector of autocovariances, c_0 and $[c_1,...,c_q]$, for the residuals from the instrumental variable estimator above. Then, the iteration is:

a. $s^2 \;=\; c_0 \,/\, (1 + \boldsymbol{\theta}'\boldsymbol{\theta})$,

b. For $i = q,...,1$ (counting backwards) $\theta_i \;=\; c_i/s^2 \;-\; \sum_{j=1}^{q-i} \theta_j\theta_{j+\mathrm{i}}$,

c. Check for convergence based on change from the last iteration. Exit if more than 20 iterations have been taken or if $\boldsymbol{\theta}$ has exploded. Otherwise, return to Step a.

NOTE: This method of estimating the moving average parameters is not guaranteed to be stable. The estimates can diverge. For example, the following will produce the outcome:

```
SAMPLE   ; 1-125 $
CREATE   ; w = Rnn(0,1) ; v = Rnn(0,1) $
ARMAX    ; Lhs = w ; Rhs = one,v ; Model = 1,0,1$
```

```
ARMAX: Moving average terms are explosive. Exit iterations.
```

The derivatives are computed as follows:

$$\partial\varepsilon_t/\partial\mu = 1 + \theta_1\partial\varepsilon_{t-1}/\partial\mu + ... + \theta_q\partial\varepsilon_{t-q}/\partial\mu,$$

$$\partial\varepsilon_t/\partial\beta_k = x_{tk} + \theta_1\partial\varepsilon_{t-1}/\partial\beta_k + ... + \theta_q\partial\varepsilon_{t-q}/\partial\beta_k\,,$$

$$\partial\varepsilon_t/\partial\phi_r = y_{t-r} + \theta_1\partial\varepsilon_{t-1}/\partial\phi_r + ... + \theta_q\partial\varepsilon_{t-q}/\partial\phi_r\,,$$

$$\partial\varepsilon_t/\partial\theta_s = \varepsilon_{t-s} + \theta_1\partial\varepsilon_{t-1}/\partial\theta_s + ... + \theta_q\partial\varepsilon_{t-q}/\partial\theta_s\,.$$

These are difference equations which we initialize at zero for q periods. Collecting each set of derivatives in a matrix, we obtain the following convenient representation:

$$\mathbf{G}_\mu = [\partial\varepsilon_{t-s}/\partial\mu] \quad (1\times q),$$

$$\mathbf{G}_\beta = [\partial\varepsilon_{t-s}/\partial\beta_k] \quad (k\times q),$$

$$\mathbf{G}_\phi = [\partial\varepsilon_{t-s}/\partial\phi_r] \quad (p\times q),$$

$$\mathbf{G}_\theta = [\partial\varepsilon_{t-s}/\partial\theta_s] \quad (q\times q).$$

Then,

$$\partial\varepsilon_t/\partial\mu = 1 + \mathbf{G}_\mu\boldsymbol{\theta},$$

$$\partial\varepsilon_t/\partial\beta = \mathbf{x}_t + \mathbf{G}_\beta\boldsymbol{\theta},$$

$$\partial\varepsilon_t/\partial\phi = \mathbf{y}_{lags} + \mathbf{G}_\phi\boldsymbol{\theta},$$

$$\partial\varepsilon_t/\partial\theta = \boldsymbol{\varepsilon}_{lags} + \mathbf{G}_\theta\boldsymbol{\theta}.$$

As noted, the derivatives are initialized at zero for the first q observations. Thereafter, the difference equation is evaluated in seriatim, simply by right shifting the columns of the matrices and inserting the current value in the vacant first column in preparation for the next observation.

At exit from the iterations, the variance estimator $\hat{\sigma}^2$ is the mean square of the estimated residuals minus the squared mean (since they do not have mean zero),

$$\hat{\varepsilon}_t = y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t - \hat{\boldsymbol{\phi}}'\mathbf{y}_{lags} - \hat{\theta}_1\hat{\varepsilon}_{t-1} - ... - \hat{\theta}_q\hat{\varepsilon}_{t-q}$$

with the series begun with q initial values of zero for the disturbances.

E12.3 The Geometric Lag Model

In the geometric distributed lag model, the moving average form is

$$y_t \ = \ \alpha + \beta(1\text{-}\lambda)[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \ldots\ ] + \varepsilon_t.$$

A nonlinear least squares estimator can be computed based on the regression

$$y_t \ \approx \ \alpha + \beta(1\text{-}\lambda)x_t{*} + \lambda^t\mu^0$$

$$y_t \ \approx \ \alpha + \beta(1\text{-}\lambda)x_t{*} + \mu^0 z_t$$

where μ^0 is the 'truncation remainder' and

$$x_t{*} \ = \ x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \ldots + \lambda^{t-1}x_1.$$

Thus, we regress y_t on a constant, $x_t{*}$ and $z_t = \lambda^t$. (See, for example, Greene (2011).) This is a nonlinear regression because of λ. With a known value of λ, the regression is linear in parameters α, $\beta(1\text{-}\lambda)$ and μ^0. We obtain the nonlinear least squares estimates by computing the linear model for different values of λ. The one associated with the smallest sum of squares gives the nonlinear least squares estimator.

The procedure involves just setting up the artificial regressors for the value of λ, then computing the linear regression. Each regression estimates α, $k = \beta(1\text{-}\lambda)$, and μ^0 given the value of λ. We collect the values of λ and the associated sums of squares in a vector, plot them, then go back and examine in detail the one which minimizes the sum of squares. For computational purposes, the two artificial regressors can be obtained as follows:

$$x_1{*} \ = \ x_1 \qquad \text{and } z_1 \ = \ \lambda$$

then $$x_t{*} \ = \ x_t + \lambda x_{t-1} \qquad \text{and } z_t \ = \ \lambda z_{t-1},\ t = 2,\ldots,T.$$

One might compute the regression for values of λ from .01 to .99 in steps of .01. When the best value is located in this grid, we could repeat the exercise over a finer grid if more precision is desired. The following lists a basic routine for computing the regression coefficients. We then turn to computation of the asymptotic covariance matrix.

First set up initial value of the counter and estimation criterion. Then, define the procedure to estimate the regression and keep the results. We also retain the optimal value of *lambda* by using two **CALC** commands. The procedure is defined generally so that you can pass any two variables you wish to it.

The commands are:

```
PROCEDURE = GeoLag(y,x) $
MATRIX      ; ee = Init(99,1,0) ; l = ee $
CALC        ; i =  1
            ; eemin = 9999999
            ; best = 0
            ; lambda = .01 $
DO WHILE    ; 10 ; i  <  100  $
CREATE      ; If(_obsno = 1) | z = lambda ; xstar = x $
CREATE      ; If(_obsno > 1) | z = lambda * z[-1]
            ; xstar = x + lambda * xstar[-1] $
REGRESS     ; Lhs = y ; Rhs = one,z,xstar ; Quietly $
MATRIX      ; ee(i) = sumsqdev ; l(i) = lambda
            ; If[sumsqdev < eemin] best = lambda
            ; eemin = sumsqdev $
CALC        ; i = i + 1
            ; lambda = lambda + .01 $
ENDDO       ; 10 $
ENDPROCEDURE
```

Execute the procedure and display the plot to show the results.

```
EXECUTE     ; Proc = GeoLag(dep. var…, indep. var...) $
MPLOT       ; Lhs = l ; Rhs = ee
            ; Fill ; Endpoints = 0,1 $
```

Notice the use of the scalars in the **CREATE** commands and as subscripts for the matrices.

We will apply this to the GDP and M1 data. For this example, we use the first differences of the logs of GDP and M1, so an observation is lost.

```
PERIOD      ; 1950.1 - 2000.4 $
CREATE      ; loggdp = Log(realgdp) ; logm1 = Log(m1) $
CREATE      ; dlgdp = loggdp - loggdp[-1]
            ; dlm1 = logm1 - logm1[-1] $
PERIOD      ; 1950.2 - 2000.4 $
EXECUTE     ; Proc = GeoLag(dlgdp,dlm1) $
MPLOT       ; Lhs = l ; Rhs = ee
            ; Fill ; Endpoints = 0,1
            ; Grid
            ; Title = Sum of Squares for Values of Lambda $
```

The result is contained in the figure below:

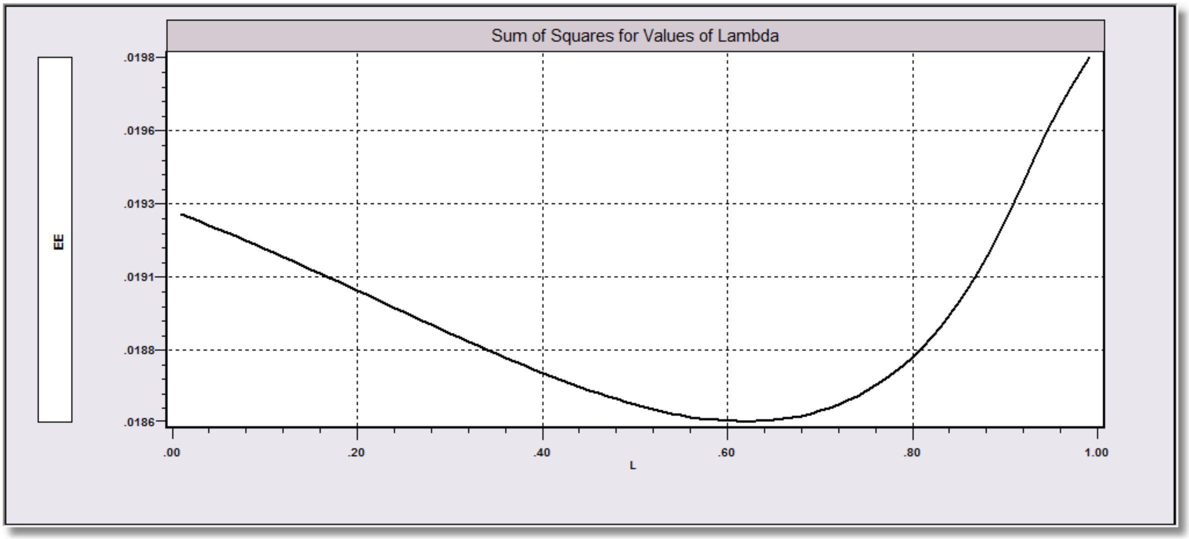


Figure E12.4 Plot of Sum of Squares

Assuming that we have found the optimal value of λ, we can now compute the appropriate asymptotic covariance matrix for the estimates. Let s^2 be the mean squared residual,

$$s^2 = (1/T)\Sigma_t e_t^2.$$

Let $\mathbf{W}(\lambda)$ be the $T\times 3$ regressor matrix used in the regression. Each row is $\mathbf{w}_t(\lambda) = [1, x_t^*, \lambda^t]$. Denoting the nonlinear least squares estimates by a, k, m^0, and l, we require the matrix

$$\text{Est.Var}[a, k, m^0, \ell] = s^2 \begin{bmatrix} \mathbf{W'W} & \mathbf{W'd} \\ \mathbf{d'W} & \mathbf{d'd} \end{bmatrix}^{-1}$$

where $\mathbf{d}$ is an extra column which accounts for the estimate of λ. Each element in $\mathbf{d}$ is

$$\begin{aligned} d_t &= \partial[a + kx_t^* + \lambda^t m^0]/\partial\lambda \\ &= t\lambda^{t-1}m^0 + k\partial x_t^*/\partial\lambda \\ &= t\lambda^{t-1}m^0 + k\delta_t. \end{aligned}$$

We can compute δ_t by using the recursion

$$\begin{aligned} \delta_1 &= 0, \\ \delta_2 &= x_1, \\ \delta_t &= x_{t-1}^* + \lambda\delta_{t-1} \text{ for } t > 2. \end{aligned}$$

(Because k is a function of β, a free parameter, we need not account for the presence of λ in k when we differentiate.) The procedure could be modified as follows:

```
PROCEDURE = GeoLag(y,x) $
```

Set up the iterations using two column vectors, a counter, and the parameter.

```
MATRIX      ; ee = Init(99,1,0) ; l = ee $
CALC        ; i = 1 ; eemin = 9999999 ; best = 0 ; lambda = .01 $
```

Execute the grid search.

```
DO WHILE    ; 10 ; i  <  100  $
CREATE      ; If(_obsno = 1) | z = lambda ; xstar = x $
CREATE      ; If(_obsno > 1) | z = lambda * z[-1]
            ; xstar = x + lambda * xstar[-1] $
REGRESS     ; Lhs = y ; Rhs = one,z,xstar ; Quietly $
MATRIX      ; ee(i) = sumsqdev ; l(i) = lambda
            ; If[sumsqdev < eemin] best = lambda
            ; eemin = sumsqdev $
CALC        ; i = i + 1 ; lambda = lambda+.01 $
ENDDO       ; 10 $
```

The grid search kept the optimal value. Now, compute the parts, run the regression, and get the covariance matrix.

```
CREATE      ; If(_obsno = 1) | z = best  ; xstar = x $
CREATE      ; If(_obsno > 1) | z = best * z[-1]
            ; xstar = x + best * xstar[-1] $
REGRESS     ; Lhs = y ; Rhs = one,z,xstar $
CREATE      ; delta = 0 $
CREATE      ; If( _obsno = 1) delta = 0 $
CREATE      ; If( _obsno = 2) delta = x $
CREATE      ; If( _obsno > 2) delta = xstar[-1] + best * delta[-1]
            ; t = Trn(1,1)
            ; d = b(2) * delta + t * b(3) * lambda^(t-1) $
NAMELIST    ; xs = one,xstar,z,d $
```

Display the full set of results.

```
MATRIX      ; var = ssqrd * <xs’xs>
            ; beta = [b/best] ;  Stat(beta,var,x) $
ENDPROC
```

We applied the preceding to the data given earlier, using the changes in the logarithms;

```
PERIOD      ; 1950.1 - 1980.4 $
CREATE      ; dlgnp = loggnp - loggnp[-1]
            ; dlm1  = logm1  - logm1[-1] $
PERIOD      ; 1950.2 - 1980.4 $
EXECUTE     ; Proc = GeoLag(dlgnp,dlm1) $
MPLOT       ; Lhs = l ; Rhs = ee ; Fill ; Endpoints = 0,1 $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=DLGDP    Mean                 =        .00911
             Standard deviation   =        .01145
             No. of observations  =           123  Degrees of freedom
Regression   Sum of Squares       =    .231095E-02            2
Residual     Sum of Squares       =    .136724E-01          120
Total        Sum of Squares       =    .159833E-01          122
             Standard error of e  =        .01067
Fit          R-squared            =        .14459  R-bar squared =   .13033
Model test   F[  2,   120]        =      10.14140  Prob F > F*   =   .00009
Diagnostic   Log likelihood       =     385.40111  Akaike I.C.   = -9.05578
             Restricted (b=0)     =     375.79674  Bayes  I.C.   = -8.98719
             Chi squared [  2]    =      19.20875  Prob C2 > C2* =   .00007
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
   DLGDP|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|     .00315         .00193     1.63  .1060     -.00064    .00694
       Z|     .06594***      .02031     3.25  .0015      .02613    .10574
   XSTAR|     .27584***      .08367     3.30  .0013      .11185    .43983
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Number of observations in current sample =     123
Number of parameters computed here       =       4
Number of degrees of freedom             =     119
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  Matrix|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  BETA_1|     .00315         .00213     1.48  .1391     -.00102    .00732
  BETA_2|     .06594         .11218      .59  .5567     -.15394    .28581
  BETA_3|     .27584***      .02124    12.99  .0000      .23421    .31747
  BETA_4|     .47000***      .00014  3447.46  .0000      .46973    .47027
--------+--------------------------------------------------------------------
```

HINT: In even a moderate sample, λ^t and δ_t may degenerate to sequences of zeros, and *LIMDEP* will return diagnostics about over- or underflows. The diagnostics can be ignored.

E12.4 Roots of Dynamic Equations

For the dynamic equation,

$$y_t = \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \ldots + \gamma_p y_{t-p} + \text{any other terms},$$

the characteristic equation is

$$1 - \gamma_1 z - \gamma_2 z^2 - \ldots - \gamma_p z^p = 0.$$

The difference equation is stable if all of the roots of this polynomial are outside the unit circle. (They may be complex.) The roots of the equation are the reciprocals of the characteristic roots of the matrix

$$\mathbf{A} = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} & \gamma_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

If the root is a complex pair ($a \pm bi$), the reciprocal is ($a/M \mp b/Mi$), where $M = (a^2 + b^2)^{½}$, the modulus. In the **MATRIX** command

MATRIX ; Root(A) $

if **A** is a row or column vector, *LIMDEP* assumes that **A** is the set of lag coefficients of a difference equation. It then sets up the preceding matrix, computes the roots, and reports the reciprocals in the form of the complex pair. You can then compute the modulus of the smallest one and resolve the stability question. If **A** is not a vector, then *LIMDEP* assumes **A** is a symmetric matrix and reports the (real) characteristic roots.

For example: Is the equation $y_t = .7y_{t-1} - .5y_{t-2} + .3y_{t-3} + \varepsilon$ stable?

MATRIX ; List ; A = [.7, -.5, .3] ; Root(A) $

LIMDEP reports a 3×2 matrix,

```
       A|           1           2           3
--------+------------------------------------
       1|     .700000    -.500000     .300000
  Result|           1           2
--------+------------------------
       1|    .0586812     1.46563
       2|    .0586812    -1.46563
       3|     1.54930     .000000
```

The smallest root is greater than one, so the answer is yes.

A related computation is the stability of a dynamic system of linear equations. Greene (2011) discusses this at length. The computational aspect can be reduced to the following: The dynamic equation system, of any lag length, can be reconstructed in the form

$$\mathbf{y}_t = \mathbf{q}_t + \mathbf{R}\mathbf{y}_{t-1}.$$

Stability of the system depends on the characteristic roots of **R** being less than 1.0 in absolute value. To obtain the characteristic roots of a nonsymmetric matrix, then check the modulus of the dominant root, use

```
MATRIX    ; z  = Cxrt(r) $
CALC      ; List
          ; bigroot = z(1,1)^2  +  z(1,2)^2 $
```

The function Cxrt computes complex roots for nonsymmetric matrices. The result is a $K\times 2$ matrix whose first column is the real parts and the second column is the imaginary parts of the roots. K is the number of rows in the source matrix.

WARNING: The eigenvalue problem for nonsymmetric matrices must be solved iteratively, if it can be solved at all. It may, in fact, be impossible, to find the roots. If so, an error message is sent, instead of an answer.

Example

The data in Figure E12.5 are the classic 'Klein I' data used to build and test simultaneous equations estimators. (See Greene (2011, Chapter 10).)

```
IMPORT $
Year  C     P    Wp    I     K1    X    Wg   G    T
1920 39.8 12.7 28.8  2.7 180.1 44.9 2.2  2.4  3.4
1921 41.9 12.4 25.5 -0.2 182.8 45.6 2.7  3.9  7.7
1922 45.0 16.9 29.3  1.9 182.6 50.1 2.9  3.2  3.9
1923 49.2 18.4 34.1  5.2 184.5 57.2 2.9  2.8  4.7
1924 50.6 19.4 33.9  3.0 189.7 57.1 3.1  3.5  3.8
1925 52.6 20.1 35.4  5.1 192.7 61.0 3.2  3.3  5.5
1926 55.1 19.6 37.4  5.6 197.8 64.0 3.3  3.3  7.0
1927 56.2 19.8 37.9  4.2 203.4 64.4 3.6  4.0  6.7
1928 57.3 21.1 39.2  3.0 207.6 64.5 3.7  4.2  4.2
1929 57.8 21.7 41.3  5.1 210.6 67.0 4.0  4.1  4.0
1930 55.0 15.6 37.9  1.0 215.7 61.2 4.2  5.2  7.7
1931 50.9 11.4 34.5 -3.4 216.7 53.4 4.8  5.9  7.5
1932 45.6  7.0 29.0 -6.2 213.3 44.3 5.3  4.9  8.3
1933 46.5 11.2 28.5 -5.1 207.1 45.1 5.6  3.7  5.4
1934 48.7 12.3 30.6 -3.0 202.0 49.7 6.0  4.0  6.8
1935 51.3 14.0 33.2 -1.3 199.0 54.4 6.1  4.4  7.2
1936 57.7 17.6 36.8  2.1 197.7 62.7 7.4  2.9  8.3
1937 58.7 17.3 41.0  2.0 199.8 65.0 6.7  4.3  6.7
1938 57.5 15.3 38.2 -1.9 201.8 60.9 7.7  5.3  7.4
1939 61.6 19.0 41.6  1.3 199.9 69.5 7.8  6.6  8.9
1940 65.0 21.1 45.0  3.3 201.2 75.7 8.0  7.4  9.6
1941 69.7 23.5 53.3  4.9 204.5 88.4 8.5 13.8 11.6
```

Figure E12.5 Klein I Data

For Klein's Model I as estimated by two stage least squares, the relevant **R** matrix is

$$\mathbf{R} = \begin{matrix} 0.172 & -0.051 & -0.008 \\ 1.511 & 0.848 & 0.743 \\ -0.287 & -0.161 & 0.818 \end{matrix}$$

The command

MATRIX ; R = [.172, -.051, -.008 / 1.511, .848, .743 / -.287, -.161, .818]
; List ; Cxrt(r) $

produces

```
  Result|             1             2
--------+----------------------------
       1|       .769242      -.349385
       2|       .769242       .349385
       3|       .299516       .000000
```

which is the widely cited result. The modulus of the first root is .8448, so the system is, as expected, stable.

E13: The Box-Cox Regression Model

E13.1 Introduction

The Box-Cox transformation is $q^{(\gamma)} = (q^{\gamma} - 1)/\gamma$ or $\log(q)$ if $\gamma = 0$. The Box-Cox regression model is:

$$y^{(\theta)} = \boldsymbol{\beta}'\mathbf{x}^{(\lambda)} + \boldsymbol{\alpha}'\mathbf{z} + \varepsilon.$$

This model allows different transformations for the Rhs and Lhs variables. The vector **z** contains any variables to which the transformation should not be applied, for example dummy variables, etc. Four forms of the model may be estimated:

Model 1: transformation (λ) applied only to the Lhs variable, y,
Model 2: transformation (λ) applied only to the Rhs variables, x,
Model 3: same transformation (λ) for the Lhs, y, and Rhs, x, variables,
Model 4: different transformations (θ) for the Lhs and (λ) for the Rhs variables.

The estimator also allows heteroscedasticity:

$$\text{Var}(\varepsilon) = \sigma^2[w^2]^{(\lambda)}$$

where w is any variable. The same transformation that is applied to the right hand side is also applied to the weights.

The estimator is maximum likelihood. If only a single value of λ and/or θ are specified, the estimator is least squares conditioned on that (those) values. Otherwise, you may specify a grid search over values of λ with a fixed θ or a full algorithmic search over λ and θ for the fully general model.

E13.2 Model Commands

The essential command for the Box-Cox model is

BOXCOX ; Lhs = dependent variable
; Rhs = independent variables that are to be transformed
; Lambda = the value of λ $

Specifications **; Lhs**, **; Rhs**, and **; Lambda** are mandatory. This basic form requests Model 1, transformation of the Lhs variable by the value of λ that you specify. The standard errors of the estimates are estimated as if λ had been estimated by maximum likelihood.

Since the standard errors cannot be computed unless all transformed variables are strictly positive, the data are checked for this. Also, if heteroscedasticity is specified, the variable '*w*' must always be strictly greater than one, once again to prevent computing logs of negative numbers when computing standard errors and the log likelihood function.

E13.2.1 Specification of the Model

The four different variations are requested as follows:

Model 1: As above. **; Lambda = the specification of λ**

Model 1 specifies that only the Lhs variable is to be transformed. For any of the other three models, the Rhs variables are to be transformed, so the following will be very important. Add

; Rh2 = list of any variables which are not transformed

Of course, the constant term, *one*, is not transformed. You may include *one* in either Rhs or Rh2, or neither if you prefer. Note, as well that for Model 1, although you provide an Rhs list, all variables are actually of type Rh2. This is taken care of internally, and you need not worry about the distinction. For the other three forms of the model, you will use:

Model 2: **; Lambda = specification ; Model = 2**
Model 3: **; Lambda = specification ; Model = 3**
Model 4: **; Lambda = specification ; Theta = value ; Model = 4**

In Models 1-3, there are various specifications of **; Lambda**. But, in Model 4, you always provide only a single value of θ.

E13.2.2 Specification of the Estimation Method

You may specify that the model is to be estimated conditioned on the specific value(s) you specify for λ (and θ), or that a search for the optimal value(s) be undertaken. For searching, you may choose a grid search or a full algorithmic function optimization procedure. To compute estimates at a specific value for λ (and θ if Model 4), use:

; Lambda = the value [; Theta = the value for Model 4]

To specify a grid search over the range λ = lower to λ = upper, estimating the parameters at N values of λ including the endpoints use:

; Lambda = lower,upper ; Pts = number of points [; Theta = value]

θ is still fixed at the single value. The estimates are those in the specified range associated with the highest value of the log likelihood function. To do a full maximum likelihood estimation procedure for λ (and θ if Model 4), use:

; Lambda = value [; Theta = value is optional]
; MLE ; Model = 1,2,3 or 4

Note the inclusion of **; MLE**. In this case, you are providing the starting values for the transformation parameters. The starting values for the other parameters are obtained by ordinary least squares involving transformed variables. This can be used for any of the four models.

If your command specifies the MLE (**; MLE**), then you are using the general optimization program and the iteration control options are also available as listed below. You may still provide a grid of values for λ. If you do, and if the MLE happens to be in the grid you provide, **; MLE** will just fine tune the estimates. If the MLE is not in the grid, the grid search may have been a waste of time, but might still have improved the start value you provided.

E13.2.3 Starting Values

The starting values for ML iterations are obtained by ordinary least squares. If you have specified values for λ and/or θ, these are used for the transformation parameters. Either the grid search or estimation at a specific value will provide new estimates of the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ parameters. The transformation parameters, themselves, that you provide are the starting values for MLE. In any event, the **; Start = list of starting values** specification is not used by this estimator.

HINT: For some values of λ, the iterations will terminate with a message about a nonpositive variance. The problem is that for very high or very low powers, a variable can become just a column of zeros or simply too large. Rescaling the data may help.

E13.2.4 The Asymptotic Covariance Matrix

As noted earlier, we use the analytic second derivatives matrix to compute the estimated asymptotic covariance matrix of the estimated parameters. In the unusual case in which this is not positive definite, the Berndt, Hall, Hall, and Hausman estimator is used, instead. Although it is inadvisable, you can obtain (perhaps for comparison purposes) an estimated covariance matrix which takes λ and/or θ as fixed value(s) rather than an estimated parameter(s). The estimator is simply the conventional estimator from the least squares procedure. Request this estimator with

; Fixed

The literature varies on the method of computation of the asymptotic covariance matrix. Use of the BHHH estimator is common. Spitzer (1984) argues (incorrectly, in fact, as he neglects to account for the variation of the variance estimator) that one should not use the BHHH estimator, and that the Hessian is the appropriate choice. We use the Hessian. You may, if you wish, dictate that the variation in the transformation parameter(s) be ignored, in which case the covariance matrix of the estimates is simply what is produced by least squares.

E13.2.5 Model Specifications

This is the full list of general specifications for this nonlinear estimation program. Specific elements of the model command are detailed in Section E13.3.

Controlling Output from Model Commands

; Par keeps ancillary parameters in main results vector *b*.
; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Display Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **Test: spec**.
; Maxit = 0 ; Start = the restricted values specifies the Lagrange multiplier test.

E13.3 Model Components

There are two additional specifications which modify the basic model.

E13.3.1 Heteroscedasticity

To request the heteroscedastic form of the disturbance variance, use:

; Wts = name of variable w

This model is different from other ones in *LIMDEP* in that you *do not* provide the reciprocals of the variances. The variable you give must be untransformed, since it will be transformed by this estimator. In addition, this variable must be strictly greater than 1.0 for all observations.

E13.3.2 Restrictions on Parameters

The preceding discussion shows, by implication, how to restrict λ and/or θ to specific values if you wish. Restrictions on the slope parameters $[\boldsymbol{\beta},\boldsymbol{\alpha}]$ cannot be imposed directly, except, of course, by building them into the model. For example, to force $\beta_j = \beta_k$, the *j*th and *k*th regressors need only to appear in the form of their sum to impose the restriction.

HINT: The scaling of the parameters in this model depends crucially on the transformation parameters. Thus, any nonhomogeneous restriction on the model parameters, e.g., $\beta_j+\beta_k = 1$, will be extremely problematic.

You can test restrictions, however, by the usual two methods. The

; Test: ...

specification can be used to test linear restrictions on the parameters. The parameter vector is exactly $[\boldsymbol{\beta},\boldsymbol{\alpha}]$ in the order of your **; Rhs,** then **; Rh2** lists. The other method is to use the *Last Model* formulation with the **WALD** command. The labels to use for this approach are given below with the listing of the saved results.

E13.4 Output and Saved Results

After the display of intermediate output from the minimization routine (if you use **; MLE**), this program displays the usual output for a regression model. This includes the initial table of fit measures and diagnostic statistics and the coefficient estimates, standard errors, t ratios, etc. The predicted values for this model are computed using:

$$y_i = \{\theta\,[\boldsymbol{\beta}'\mathbf{x}_i^{(\lambda)} + \boldsymbol{\alpha}'\mathbf{z}_i] + 1\}^{1/\theta}$$

Residuals are simply the difference between the actual value and the prediction. The additional variables, '*var*1' and '*var*2' usually shown in the listing are not computed for this model.

You may also display marginal effects for this model with

; Partial Effects

(In previous versions of *LIMDEP* and *NLOGIT*, the command was **; Marginal Effects**. This form is still supported, and has the same meaning in the current versions of *LIMDEP* and *NLOGIT*.)

For the Box-Cox model, the elasticities are

$$\partial \log y/\partial \log x = \beta(x^{\lambda})/(y^{\theta}),$$

so the marginal effects are

$$\partial y/\partial x = (y/x)(\partial \log y/\partial \log x) = \beta(x^{\lambda-1})/(y^{\theta-1}).$$

Values kept for later use are:

Matrices: *b* = slope coefficients
varb = estimated asymptotic covariance matrix

If you include **; Par** in the command, the additional parameters $[\lambda,\theta,\sigma^2]$ are included in *b* and *varb*.

epsilon = elasticities, $\partial \log y/\partial \log x_k$. This additional matrix contains elasticities for all of the Rhs and Rh2 variables in the model. These are computed at the sample means of all the exogenous variables according to the result above.

Scalars:

ssqrd	=	$\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/n$
s	=	$\sqrt{ssqrd}$
rsqrd	=	$1 - (\mathbf{e}'\mathbf{e}/n)/s_y^2$ (not necessarily positive)
sumsqdev	=	$\mathbf{e}'\mathbf{e}$
rho	=	0
degfrdm	=	*n*
nreg	=	*n*
sy	=	sample standard deviation of Lhs
ybar	=	sample mean of Lhs
logl	=	log likelihood function at best estimates
lmda	=	λ
theta	=	θ or 1 if Model 2 or λ if Model 1 or 3

Last Model: *b_variables* **; Rhs** and **; Rh2** are combined.

Last Function: None

Note, no function is stored for the **PARTIALS** and **SIMULATE** programs. However, partial effects and predictions are provided with the model command. You can also use **; Function = …** and provide your own specification of the Box-Cox model in either **PARTIALS** or **SIMULATE**.

E13.5 Application

To illustrate the Box-Cox model, we will use the macroeconomic data used in the application in Chapter E12. (See Section E12.2.4 for discussion.) We will fit several different forms of the Box-Cox model. We emphasize, for these data, the calculations are purely illustrative, and are not intended to provide any evidence about the interest elasticity of the demand for money. We precede estimation with:

```
DATES      ; 1950.1 $
PERIOD     ; 1950.1 - 2000.4 $
CREATE     ; lm = Log(m1) ; loggdp = Log(realgdp) $
NAMELIST   ; x = one,tbilrate,loggdp $
```

The first model specifies the log of *money* on the Lhs and transformations on the Rhs only.

```
BOXCOX     ; Lhs = lm ; Rhs = x
           ; Model = 2
           ; Lambda = -2,2
           ; Pts = 50
           ; List
           ; Partial Effects $
```

The next model specifies transformation of both sides of the equation.

```
BOXCOX     ; Lhs = m1 ; Rhs = one,tbilrate,realgdp
           ; Model = 3
           ; Lambda = -1,1
           ; Pts = 100 $
```

We now allow the estimator to find the MLE. The initial part of the search encounters some difficulties in optimizing the function, but after several iterations, the interior maximizer is found, actually near where the grid search located it in the previous command.

```
BOXCOX     ; Lhs = m1 ; Rhs = one,tbilrate,realgdp
           ; Model = 3
           ; Lambda = -1,1
           ; Pts = 100
           ; MLE $
```

Finally, we attempt a full ML estimation of Model 4. This is extremely sensitive to the starting values. The following terminates after 500 iterations. The values are similar to the estimates after 100 iterations.

```
BOXCOX     ; Lhs = m1 ; Rhs = one,tbilrate,realgdp
           ; Model = 4
           ; Lambda = 0.0
           ; Theta = .35
           ; MLE
           ; Maxit = 500 $
```

```
-----------------------------------------------------------------------------
Box-Cox Nonlinear Regression Model................
Maximum likelihood estimator, Het.:W(i) = ONE
LHS=LM       Mean                 =       5.79786
             Standard deviation   =        .80557
             Number of observs.   =           204
Model size   Parameters           =             3
             Degrees of freedom   =           201
Residuals    Sum of squares       =       4.65159
             Standard error of e  =        .15100
Fit          R-squared            =        .96486
             Adjusted R-squared   =        .96503
Model test   F[  2,   201] (prob) =  2759.7(.0000)
Diagnostic   Log likelihood       =      96.18952
             Restricted(b=0)      =    -244.85556
             Chi-sq [  2]  (prob) = 682.1(  .0000)
Info criter. Akaike Info. Criter. =      -3.75150
Not using OLS or no constant. Rsqrd & F may be < 0
BxCx transformations: RHS= Lambda   , LHS= ONE
Elasticities have been kept in matrix EPSILON
Log-L acctg. for LHS transformation =    96.18851
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LM|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variables transformed by LAMBDA =    -.44898
TBILRATE|    -.59771***      .10567    -5.66  .0000     -.80483   -.39060
  LOGGDP|    40.4062***    11.20460     3.61  .0003     18.4456   62.3669
        |Variables that were not transformed
Constant|   -48.7572***     8.87602    -5.49  .0000    -66.1539  -31.3606
        |Variance and transformation parameters
  Lambda|    -.44898***      .12986    -3.46  .0005     -.70350   -.19446
Sigma-sq|     .02280***      .00226    10.10  .0000      .01838    .02723
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+---------------------------------------------------------+
| Marginal Effects for Box-Cox                            |
+----------+----------+----------+----------+----------+
| Variable | Mean     | Coeff.   | Slope    | Elast.   |
+----------+----------+----------+----------+----------+
| TBILRATE |  5.22941 |  -.59771 |  -.05438 |  -.04934 |
| LOGGDP   |  8.31231 | 40.40624 |  1.87841 |  2.70879 |
+----------+----------+----------+----------+----------+
Predicted Values          (* => observation was not in estimating sample.)
Observation         Observed Y   Predicted Y   Residual     95% Forecast Interval
  1950.1            4.7022969    4.4975736    .2047233      .000000      .000000
  1950.2            4.7162642    4.5387079    .1775564      .000000      .000000
  1950.3            4.7269452    4.5950364    .1319088      .000000      .000000
  1950.4            4.7355842    4.5844282    .1511560      .000000      .000000
 (Observations omitted)
```

```
-----------------------------------------------------------------------------
Box-Cox Nonlinear Regression Model...............
Maximum likelihood estimator, Het.:W(i) = ONE
LHS=M1       Mean                 =      453.92147
             Standard deviation   =      359.72633
             Number of observs.   =            204
Model size   Parameters           =              3
             Degrees of freedom   =            201
Residuals    Sum of squares       =    .539951E-01
             Standard error of e  =         .01627
Fit          R-squared            =        1.00000
             Adjusted R-squared   =        1.00000
Model test   F[  2,    201] (prob) =********(.0000)
Diagnostic   Log likelihood       =      550.70874
             Restricted(b=0)      =    -1489.57232
             Chi-sq [  2]  (prob) =4080.6(  .0000)
Info criter. Akaike Info. Criter. =       -8.20757
Not using OLS or no constant. Rsqrd & F may be < 0
BxCx transformations: RHS= Lambda  , LHS= Lambda
Elasticities have been kept in matrix EPSILON
Log-L acctg. for LHS transformation =  -1074.09734
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      M1|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variables transformed by LAMBDA =    -.37374
TBILRATE|    -.04533**       .02210    -2.05  .0402     -.08864   -.00203
 REALGDP|    4.52905***     1.13338     4.00  .0001     2.30767   6.75044
        |Variables that were not transformed
Constant|   -9.16115***      .72330   -12.67  .0000   -10.57879  -7.74350
        |Variance and transformation parameters
  Lambda|    -.37374***      .10128    -3.69  .0002     -.57225   -.17522
Sigma-sq|     .00026         .00031      .85  .3965     -.00035    .00088
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Warning   141: Iterations:current or start estimate of sigma is nonpositive
```

(This warning is repeated 10 times)

```
Normal exit:  27 iterations. Status=0, F=    1074.095

-----------------------------------------------------------------------------
Box-Cox Nonlinear Regression Model...............
Maximum likelihood estimator, Het.:W(i) = ONE
LHS=M1       Mean                 =      453.92147
             Standard deviation   =      359.72633
             Number of observs.   =            204
Model size   Parameters           =              3
             Degrees of freedom   =            201
Residuals    Sum of squares       =    .497649E-01
             Standard error of e  =         .01562
Fit          R-squared            =        1.00000
             Adjusted R-squared   =        1.00000
Model test   F[  2,    201] (prob) =********(.0000)
Diagnostic   Log likelihood       =      559.03023
             Restricted(b=0)      =    -1489.57232
             Chi-sq [  2]  (prob) =4097.2(  .0000)
Info criter. Akaike Info. Criter. =       -8.28915
```

```
Not using OLS or no constant. Rsqrd & F may be < 0
BxCx transformations: RHS= Lambda  , LHS= Lambda
Elasticities have been kept in matrix EPSILON
Log-L acctg. for LHS transformation =  -1074.09493
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      M1|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variables transformed by LAMBDA =    -.38077
TBILRATE|    -.04385**       .02140    -2.05  .0404     -.08579   -.00191
 REALGDP|    4.60829***     1.15434     3.99  .0001     2.34582   6.87075
        |Variables that were not transformed
Constant|   -9.21038***      .74449   -12.37  .0000   -10.66955  -7.75120
        |Variance and transformation parameters
  Lambda|    -.38077***      .10135    -3.76  .0002     -.57941   -.18214
Sigma-sq|     .00024         .00029      .85  .3965     -.00032    .00081
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit: 280 iterations. Status=0, F=    1101.852

-----------------------------------------------------------------------------
Box-Cox Nonlinear Regression Model................
Maximum likelihood estimator, Het.:W(i) = ONE
LHS=M1       Mean                 =      453.92147
             Standard deviation   =      359.72633
             Number of observs.   =            204
Model size   Parameters           =              3
             Degrees of freedom   =            201
Residuals    Sum of squares       =    .494601E-01
             Standard error of e  =         .01557
Fit          R-squared            =        1.00000
             Adjusted R-squared   =        1.00000
Model test   F[  2,   201] (prob) =********(.0000)
Diagnostic   Log likelihood       =      559.65693
             Restricted(b=0)      =    -1489.57232
             Chi-sq [  2]  (prob) =4098.5(  .0000)
Info criter. Akaike Info. Criter. =       -8.29530
Not using OLS or no constant. Rsqrd & F may be < 0
BxCx transformations: RHS= Lambda  , LHS= Theta
Elasticities have been kept in matrix EPSILON
Log-L acctg. for LHS transformation =  -1101.78924
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      M1|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variables transformed by LAMBDA =    -.68116
TBILRATE|    -.07275         .06217    -1.17  .2419     -.19461    .04910
 REALGDP|    48.9134       40.03205     1.22  .2218    -29.5480  127.3748
        |Variables that were not transformed
Constant|   -69.2590       52.61522    -1.32  .1881   -172.3829   33.8650
        |Variance and transformation parameters
  Lambda|    -.68116***      .11088    -6.14  .0000     -.89848   -.46384
   Theta|    -.40472***      .13817    -2.93  .0034     -.67552   -.13391
Sigma-sq|     .00024         .00040      .60  .5487     -.00055    .00103
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Lastly, we do a grid search and find the value graphically.

```
MATRIX        ; loglik = Init(51,1,0) ; lamda = loglik $
CALC          ; i = 0 $
PROCEDURE
CALC          ; i = i + 1 $
BOXCOX        ; Quietly ; Lhs  =  lm ; Rhs = x ; Model = 2
              ; Lambda = value $
MATRIX        ; loglik(i) = logl ; lamda(i) = value $
ENDPROCEDURE
EXECUTE       ; Silent ; value = -1,1,.04 $  (51 points, .04 apart)
MPLOT         ; Lhs = lamda ; Rhs = loglik
              ; Fill ; Endpoints = -1,1
              ; Grid
              ; Title = Log Likelihood for Box-Cox Model $
```

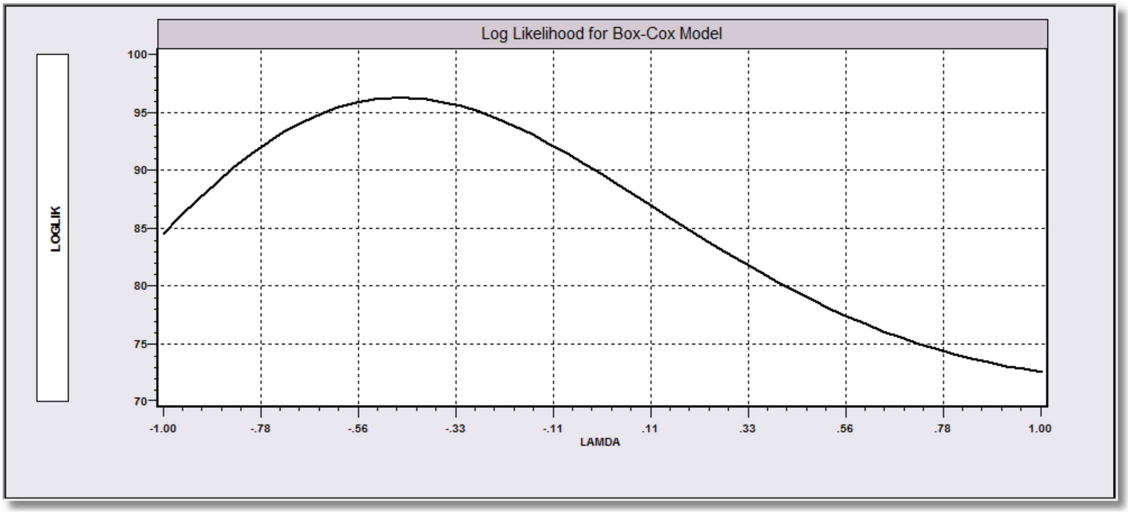


Figure E13.1 Plot of Log Likelihood for Box-Cox Model

E13.6 Technical Details

Estimation of the Box-Cox model is done in one of two ways. In the grid search procedure, the estimator is ordinary or weighted least squares. The following is needed for computation of the asymptotic covariance matrix. The maximum likelihood method is applied simply by treating the problem as an ordinary optimization problem.

The Box-Cox transformation of a variable x, for nonzero λ is:

$$x^{(\lambda)} = (x^{\lambda} - 1) / \lambda.$$

This transformation obeys the following differential equation for i = 1,...

$$d^i x^{(\lambda)}/d\lambda^i = [x^{\lambda} (\log x)^i - i(d^{i-1}x^{(\lambda)}/d\lambda^{i-1})] / \lambda.$$

The first term in the sequence is $x^{(\lambda)}$ when i=0. If λ equals 0, the preceding are replaced by:

$$x^{(0)} = \log x$$

and $$d^i x^{(\lambda)}/d\lambda^i{}_{|\lambda=0} = (\log x)^{i+1} / (i+1).$$

For purposes of the discussion to follow, it is convenient to define a notation for the function and its first and second derivatives. Thus, let

$$x_\lambda = dx^{(\lambda)}/d\lambda \quad \text{and} \quad x_{\lambda\lambda} = d^2x^{(\lambda)}/d\lambda^2.$$

The model is $$y^{(\theta)} = \Sigma_k\beta_k x_k^{(\lambda)} + \Sigma_m\alpha_m z_m + \varepsilon.$$

The xs are the Rhs variables subject to the transformation, and the zs are the Rh2 variables that are not transformed. The variance of ε is

$$\text{Var}[\varepsilon] = f = \sigma^2[w^2]^{(\lambda)}.$$

There are various restrictions on the general model which lead to the model estimated. The case of homoscedasticity is imposed by deleting the 'w' term from the model, not by a simple parametric restriction. (Setting w = 1 is insufficient, since $1^{(\lambda)} = 0$, not 1.) Other specifications are imposed by

Model 1: $\theta = \lambda$ and all regressors classified as Rh2.
Model 2: $\theta = 1$.
Model 3: $\theta = \lambda$.

The log likelihood for the Box-Cox model is

$$\text{Log}L = (\theta\text{-}1)\Sigma_i\log y_i - ½\Sigma_i[\log 2\pi + \log f_i + \varepsilon_i^2/f_i].$$

The first derivatives of the log likelihood are obtained as follows:

Let $$f_i = \sigma^2 \text{ or } \sigma^2[w\,]^{(\lambda)} \text{ whichever is appropriate.}$$

For a vector, $\mathbf{x}_i$, let $$\mathbf{x}_i^{(\lambda)} = \text{the vector of transformed variables, and}$$

$$\mathbf{x}_{i\lambda} = [x_{i1\lambda}, x_{i2\lambda},\ldots,x_{iK\lambda}]'.$$

Thus, $$\varepsilon_i = y_i^{(\theta)} - \boldsymbol{\beta}'\mathbf{x}_i^{(\lambda)} - \boldsymbol{\alpha}'\mathbf{z}_i.$$

So, $$\partial\log L/\partial\boldsymbol{\beta} = \Sigma_i[\varepsilon_i/f_i]\mathbf{x}_i^{(\lambda)},$$

$$\partial\log L/\partial\boldsymbol{\alpha} = \Sigma_i[\varepsilon_i/f_i]\mathbf{z}_i,$$

and $$\partial\log L/\partial\lambda = \Sigma_i[\varepsilon_i/f_i]\boldsymbol{\beta}'\mathbf{x}_{i\lambda}.$$

If the disturbance is heteroscedastic, add

$$\Sigma_i \{\tfrac{1}{2}(\varepsilon_i^2/f_i - 1)[w_i^2]_\lambda \,/\, [w_i^2]^{(\lambda)}\} \text{ to } \partial\text{logL}/\partial\lambda.$$

If the model is Model 1, 3, or 4,

$$\partial\text{log}L/\partial\theta \;=\; \Sigma_i[-\varepsilon_i/f_i]y_{i\theta} + \text{log}y_i.$$

If the model is Model 3, so that $\theta = \lambda$, ∂logL/$\partial\theta$ is added to ∂log$L$/$\partial\lambda$. It is omitted for Model 2. Finally,

$$\partial\text{log}L/\partial\sigma^2 \;=\; \Sigma_i \tfrac{1}{2}(\varepsilon_i^2/f_i - 1)/\sigma^2.$$

The BHHH estimator of the asymptotic covariance matrix of the estimator is obtained by summing the outer products of the individual terms listed in the summations above. The Hessian is obtained as follows: Let

$$\begin{aligned}
\boldsymbol{\delta}_i &= \boldsymbol{\beta}'\mathbf{x}_{i\lambda},\\
\gamma_i &= \boldsymbol{\beta}'\mathbf{x}_{i\lambda\lambda}, \text{ extending the vector notation defined above,}\\
w_{i\lambda}^{\;2} &= (\text{d}[w_i^2\,]^{(\lambda)}/\text{d}\lambda) \,/\, [w_i^2]^{(\lambda)},\\
w_{i\lambda\lambda}^{\;2} &= (\text{d}^2[w_i^2]^{(\lambda)}/\text{d}\lambda^2) - (w_{i\lambda}^{\;2})^2.
\end{aligned}$$

The last two terms are zero if the disturbance is homoscedastic. Denote second derivatives with subscripts;

$$\partial^2\text{log}L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' \;=\; \text{H}_{\boldsymbol{\beta\beta}}.$$

For convenience, combine $\mathbf{x}^{(\lambda)}$ and $\mathbf{z}$ in vectors $\mathbf{v}_i = [\mathbf{x}_i^{(\lambda)}\,,\ \mathbf{z}_i]$, $\mathbf{v}_{i\lambda} = [\mathbf{x}_{i\lambda}\,,\ \mathbf{0}]$, $\mathbf{v}_{i\lambda\lambda} = [\mathbf{x}_{i\lambda\lambda}\,,\ \mathbf{0}]$. Derivatives with respect to $\boldsymbol{\beta}$ below include the vector $\boldsymbol{\alpha}$ defined above. Then, the Hessian is

$$\begin{aligned}
\text{H}_{\boldsymbol{\beta\beta}} &= \Sigma_i -(1/f_i)\mathbf{v}_i\mathbf{v}_i',\\
\text{H}_{\boldsymbol{\beta}\lambda} &= \Sigma_i\; (\varepsilon_i/f_i)[\mathbf{v}_{i\lambda} - w_{i\lambda}^{\;2}\mathbf{v}_i] \;-\; (\delta_i/f_i)\mathbf{v}_i,\\
\text{H}_{\lambda\lambda} &= \Sigma_i\; (\varepsilon_i/f_i)(_{\text{i}} - (\delta_i/f_i)(\delta_i{+}2\varepsilon_i w_{i\lambda}^{\;2}) + \;\tfrac{1}{2}[w_{i\lambda\lambda}^{\;2}(\varepsilon_i^2/f_i - 1) - (\varepsilon_i/f_i)(w_{i\lambda}^{\;2})^2],\\
\text{H}_{\boldsymbol{\beta}\theta} &= \Sigma_i\; (y_{i\theta}/f_i)\mathbf{v}_i,\\
\text{H}_{\lambda\theta} &= \Sigma_i\; (y_{i\theta}/f_i)[\delta_i + w_{i\lambda}^{\;2}\varepsilon_i],\\
\text{H}_{\theta\theta} &= \Sigma_i -(1/f_i)[y_{i\theta\theta}\varepsilon_i + (y_{i\theta})^2],\\
\text{H}_{\sigma\boldsymbol{\beta}} &= \Sigma_i -(\varepsilon_i/f_i)\mathbf{v}_i/\sigma^2,\\
\text{H}_{\sigma\lambda} &= \Sigma_i -(\varepsilon_i/f_i)\delta_i/\sigma^2,\\
\text{H}_{\sigma\theta} &= \Sigma_i\; (\varepsilon_i/f_i)\text{y}_{i\theta}/\sigma^2),\\
\text{H}_{\sigma\sigma} &= \Sigma_i\; (\tfrac{1}{2} \;-\; \varepsilon_i^2/f_i)/\sigma^4).
\end{aligned}$$

The foregoing applies to Model 4. If the model is Model 3, then terms involving θ are simply added to terms involving λ. If the model is Model 2, terms involving θ are dropped. The first of these can be accomplished as follows: When $\theta = \lambda$,

$$H_{\beta\lambda} \text{ (new)} = [H_{\beta\lambda} + H_{\beta\theta}] \text{ (old)},$$

$$H_{\sigma\lambda} \text{ (new)} = [H_{\sigma\lambda} + H_{\sigma\theta}] \text{ (old)},$$

and

$$H_{\lambda\lambda} \text{ (new)} = [H_{\lambda\lambda} + H_{\theta\theta} + 2H_{\lambda\theta}] \text{ (old)}.$$

E14: Nonlinear Least Squares

E14.1 The Nonlinear Regression Model

This chapter details nonlinear least squares estimation of the general nonlinear regression model

$$y_i = f(\mathbf{x}_i,\boldsymbol{\beta}) + \varepsilon_i.$$

The function f(•,•) may be any function that is continuous in the parameters. Four estimation methods may be used:

- nonlinear ordinary least squares estimation,
- nonlinear weighted least squares estimation,
- nonlinear two stage least squares (instrumental variables - IV) estimation,
- GMM estimation.

The first two are described here. The IV estimation techniques are presented in Chapter E21. (Weights may be used with the latter two estimators as well.) GMM estimation is presented in Chapters E21 and E23.

The essential command fitting nonlinear regression models with nonlinear least squares is

NLSQ ; Lhs = dependent variable
; Fcn = the definition of the nonlinear regression function
; Labels = symbols to use for the parameters to be estimated
; Start = the starting values for the iterations $

The basic command specifies nonlinear ordinary least squares. That is, you would instruct *LIMDEP* to choose the **β** to

$$\text{Minimize wrt } \boldsymbol{\beta} \;\; \tfrac{1}{2}\Sigma_i\, [y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2 = \tfrac{1}{2}\Sigma_i\, \varepsilon_i^2.$$

If the function you specify is linear, this will produce the ordinary least squares results. To request nonlinear weighted least squares, you will specify

; Wts = weighting variable

as usual. The estimation criterion is then,

$$\text{Minimize wrt } \boldsymbol{\beta} \;\; \tfrac{1}{2}\Sigma_i\, w_i[y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2 = \tfrac{1}{2}\Sigma_i\, w_i\, \varepsilon_i^2.$$

where w_i is the weighting variable. If this is a correction for heteroscedasticity, the weighting variable should be the reciprocal of the disturbance variance, not the standard deviation.

E14.2 Command for Nonlinear Regression

The essential command for this estimator provides four sets of information:

```
NLSQ        ; Lhs = dependent variable
            ; Fcn = specification of f(•,•)
            ; Labels = the labels for the model parameters
            ; Start = starting values for the parameters $
```

This requests ordinary, unweighted, nonlinear least squares estimates. To use weighted least squares, instead, add

```
            ; Wts = weighting variable
```

to the command. Your function may contain up to 150 parameters to be estimated.

The Lhs, function and starting values are all mandatory. This list of starting values provides the initial values for the iterations for the estimator, and also tells *LIMDEP* how many parameters are being estimated. Thus, it is essential for you to be accurate in your specification of the starting values. You may use any of the methods discussed elsewhere in this manual to provide the list of starting values. These may appear in a vector or a matrix (read rowwise), a list of specific values, or in scalars that have been defined earlier. For example, the following passes on a set of OLS slopes, the estimated standard deviation of the disturbance, and the value 1.0 as starting values for a model:

```
NAMELIST    ; x = ... $
REGRESS     ; Lhs = y
            ; Rhs = x ; ... $
NLSQ        ; Lhs = ...
            ; Fcn = ...
            ; Start = b, ssqrd, 1.0 ; ...
            ; Labels = ...  $
```

The number of parameters would be two plus the number of variables in the namelist.

The labels are optional. If you do not provide labels for your parameters, they will be automatically named *b*1, *b*2, ..., *bK*, where *K* is the number of starting values you provide. For example, the following specifies a linear regression model:

```
NLSQ        ; Lhs = logy
            ; Fcn = b1 + b2*x2 + b3*x3
            ; Start = 0,0,0 $
```

Note that three parameters are defined by the starting value list. You will usually wish to use your own labels. To do so, use

; Labels = a list of labels, one for each starting value.

TIP: See Section E14.3.5 for an extremely useful device for defining labels in the command. You may also use the labels defined by a **CLIST** command. See Section R6.6.

TIP: Be careful to make sure that the labels you choose are not the same as other items you have created, such as matrices or scalars. In most cases, if you try to use a label that is already the name of a variable or a matrix or a scalar, *LIMDEP* will catch the error and issue an error message. But, there are ways that you can accidently avoid this filter, and this will lead to unexpected (and unwanted) results.

The preceding command could be changed to

```
NLSQ     ; Lhs = logy
         ; Fcn = gamma0 + thetak*x2 + thetal*x3
         ; Labels = gamma0,thetak,thetal
         ; Start = 0,0,0 $
```

LIMDEP will ensure that there is a correspondence between your labels and your starting values. However, it is not possible for the program to ensure that you have used all of the parameters in your function specification. If you define a parameter, but you do not use it in your function definition, then one of two things will occur. Either the iterations will never converge and they will exit on maximum iterations, with one of the parameters not changing from its initial value, or what appears to be convergence will be reached, but the estimated covariance matrix of the estimated parameters will be singular, as it will contain a row and column of zeros corresponding to the unused parameter. Here is an example. Note that the defined model parameter *c*3 does not appear in the regression function.

```
NLSQ     ; Lhs = y
         ; Fcn = c0+c1*x1+c2*x2
         ; Start = 0,0,0,0
         ; Labels = c0,c1,c2,c3
         ; Output = 3 $
```

```
Begin NLSQ iterations. Linearized regression.
Moment matrix has become nonpositive definite.
Switching to BFGS algorithm
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .1000D-05 chg.F   .0000D+00 max|dB|   .0000D+00
Nodes for quadrature: Laguerre=40;Hermite=20.
Replications for GHK simulator= 100
Start values:   .00000D+00   .00000D+00   .00000D+00   .00000D+00
1st derivs.     .13962D+01   .26612D+01  -.31547D+01   .00000D+00
Parameters:     .00000D+00   .00000D+00   .00000D+00   .00000D+00
Itr  1 F=  .5325D+02 gtHg=  .4357D+01 chg.F=  .5325D+02 max|db|=  .3155D+07
1st derivs.    -.47552D+00   .17547D+00  -.62430D-01   .00000D+00
Parameters:    -.20089D-01  -.38291D-01   .45391D-01   .00000D+00
Itr  2 F=  .5311D+02 gtHg=  .5107D+00 chg.F=  .1366D+00 max|db|=  .2367D+02
1st derivs.     .13412D-01   .25228D-01  -.31252D-01   .00000D+00
Parameters:    -.15250D-01  -.40076D-01   .46027D-01   .00000D+00
Itr  3 F=  .5311D+02 gtHg=  .4234D-01 chg.F=  .1327D-02 max|db|=  .8795D+00
1st derivs.     .13412D-01   .25228D-01  -.31252D-01   .00000D+00
Parameters:    -.15250D-01  -.40076D-01   .46027D-01   .00000D+00
Itr  1 F=  .5311D+02 gtHg=  .4234D-01 chg.F=  .5311D+02 max|db|=  .8795D+00
1st derivs.    -.45648D-02   .19336D-02  -.39819D-03   .00000D+00
Parameters:    -.15443D-01  -.40439D-01   .46476D-01   .00000D+00
Itr  2 F=  .5311D+02 gtHg=  .4973D-02 chg.F=  .1290D-04 max|db|=  .2836D+00
1st derivs.     .31087D-05   .91316D-05   .87055D-05   .00000D+00
Parameters:    -.15398D-01  -.40463D-01   .46485D-01   .00000D+00
Itr  3 F=  .5311D+02 gtHg=  .1299D-04 chg.F=  .1271D-06 max|db|=  .2261D-03
1st derivs.     .67446D-12   .86003D-11  -.18233D-10   .00000D+00
Parameters:    -.15398D-01  -.40463D-01   .46485D-01   .00000D+00
Itr  4 F=  .5311D+02 gtHg=  .2425D-11 chg.F=  .8811D-12 max|db|=  .5445D-11
                        * Converged
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit from iterations. Exit status=0.
Function=  .53247681205D+02, at entry,  .53109768475D+02 at exit
Models - estimated variance matrix of estimates is singular
Current estimated covariance matrix for slopes is singular.
```

E14.3 Specification of the Regression Function

NLSQ minimizes the sum of squared residuals. Your function defines the Rhs of the regression for an observation. The **; Fcn** specification is written using the rules and operators of algebra (+, -, *, /, ^). Parentheses may be used freely to force the order of evaluation of expressions. Use as many levels of parentheses as required. Entities which may appear in the specification include:

- variable names,
- any existing scalars,
- matrix elements,
- your parameters, using your labels.

Because there are a variety of named entities which can appear in the function, you should use the

; Labels = list of labels

part of the command to identify which of them are the parameters being estimated. You must then use these labels in the function you specify. Labels may be anything you like, up to eight characters.

WARNING: Use new names! Do not use program names that are in use otherwise, such as *s*, *rho*, *sigma*, *b*, etc., or the names of existing scalars or matrices. Such labels might be accepted when your command is translated, because you are free to use these entities in your function definition to supply specific values. But, later, when *LIMDEP* scans your expression to see what you have specified, it checks all other tables first, and your label list last. For example, if you use *s* as a label, and this command is the first model command that you have given, *s* will simply be taken as the as yet undefined result of a regression. The actual value would, in fact, always be fixed at 0.

The operators are +, -, *, /, ^ (for raise to the power), and @ (for the Box-Cox transformation). The usual rules are observed; ^ and @ are computed first, then * and /, and finally + and -. The CES production function provides an example. The Lhs variable in the equation might be *q*, and the command could be

```
REGRESS    ; Lhs = Log(q) ; Rhs = one,Log(k),Log(l) $
CALC       ; dkl = b(2)/(b(2)+b(3)) ; sc = b(2)+b(3) $
NLSQ       ; Lhs = q
           ; Labels = gamma,delta,r,nu
           ; Fcn = gamma * (delta*k^(-r) + (1-delta)*l^(-r)) ^ (-nu/r)
           ; Start = b(1), dkl, 1.0, sc $
```

Lastly, to use a subscripted matrix element, enclose the subscript in curled brackets, **{ }**, not parentheses. I.e., **gamma(1,1)** will confuse the compiler, use **gamma{1,1}**.

NOTE: This construction, with curled brackets, is specific to the function definition part of the **NLSQ**, **NLSUR**, **MAXIMIZE**, and **MINIMIZE** commands. Elsewhere, such as in **CALC** and **CREATE**, matrix subscripts are indicated with ordinary parentheses. Curled brackets also have a different use in **MATRIX**, but are not used in **CREATE** or **CALC**.

E14.3.1 Parameterization and Reparameterization

We will revisit this issue at several points below. In specifying a nonlinear optimization, it is often helpful to parameterize the model in such a way as to remove some of the nonlinearity. (This discussion applies generally to the several procedures in *LIMDEP* that use a user defined nonlinear optimization, and, more broadly, to optimization in general regardless of what software you might be using.) Nonlinear functions can often be written in several ways. For example:

```
NLSQ    ; Lhs = y
        ; Labels = b0, b1, b2
        ; Start = 0, 0, .5
        ; Fcn = (b0 + b1*x)^(-1/b2)  $
```

is a valid specification of the nonlinear regression model. However, with a particular data set, it is possible that the iterative procedure searching for the parameters could be unable to find a minimizer, and might break down. This can happen for several reasons. One strategy for dealing with the problem, and more generally, for facilitating estimation, is to remove unnecessary nonlinearities, such as the reciprocal of *b*2 that appears above. Even the minus sign is superfluous. The same model could be fit, possibly with greater ease, by specifying it as

```
NLSQ    ; Lhs = y
        ; Labels = b0, b1, b2
        ; Start = 0, 0, -2.0
        ; Fcn = (b0 + b1*x)^c2 $
```

Note that the -1/*b*2 has been replaced with the much simpler *c*2. Also, the starting value has been changed accordingly, from 0.5 to -2.0 = -1/0.5. If the parameter *b*2 were of particular interest, you could follow the **NLSQ** command with a

```
WALD    ; Fn1 = -1/b2  $
```

The other parts of the **WALD** command are automatic if it follows an optimization command.

The opportunities for simplification are sometimes subtle, but it helps to take them when they are available. In the CES function example at the end of the preceding section, there are two superfluous nonlinearities: The function can be specified using

```
CALC    ; sc = -sc $
NLSQ    ; Lhs = q
        ; Labels = gamma,delta,r,theta
        ; Fcn = gamma * (delta*k^(-r) + (1-delta)*l^(-r)) ^ theta
        ; Start = b(1), dkl, 1.0, sc $
```

In this example, the (-*nu*/*r*), which involves a multiplication and a reciprocal is replaced with the simpler parameter, *theta*. In fact, *theta* equals -*nu*/*r*, but since *nu* is a free parameter that appears nowhere else in the function, we can treat -*nu*/*r* as this free parameter. If *nu*, itself, is desired,

```
WALD    ; Fn1 = -r * theta $
```

would compute the estimate as well as an appropriate asymptotic standard error.

E14.3.2 Functions that May Appear in NLSQ Commands

The following functions may be used in the regression specification:

Abs(z)	= absolute value
Ac1(z)	= derivative of Ach(z) = $(z^2 - 1)^{-1/2}$
Ach(z)	= hyperbolic arc cos(z) = log(z + (z^2 – 1))
As1(z)	= derivative of Ash(z) = $(1 + z^2)^{-1/2}$
Ash(z)	= hyperbolic arc sin(z) = log($z + (1 + z^2)^{1/2}$)
At1(z)	= derivative of Ath(z) = $(1 - z^2)^{-1}$
Ath(z)	= hyperbolic arc tan(z) = .5log((1 + z)/(1 – z))
Atn(z)	= arctangent
Bds(z,a,c)	= incomplete beta function; (Bds(0,a,c) = 0, Bds(1,a,c) = 1)
Bv1(z1,z2,ρ)	= bivariate normal CDF derivative wrt x1
Bv2(z1,z2,ρ)	= bivariate normal CDF derivative wrt x2
Bvd(z1,z2,ρ)	= bivariate normal density
Bvn(z1,z2,ρ)	= bivariate normal CDF
Cos(z)	= cosine
Exp(z)	= exponent
Gma(z)	= gamma
Hc1(z)	= derivative of Hcs(z) = Hsn(z)
Hcs(z)	= hyperbolic cos(z) = .5(exp(2z)+1)/exp(z)
Hs1(z)	= derivative of Hsn(z) = Hcs(z)
Hsn(z)	= hyperbolic sin(z) = .5(exp(2z)-1)/exp(z)
Ht1(z)	= derivative of Htn(z) = 1/Hcs^2(z)
Htn(z)	= hyperbolic tan(z) = Hsn(z)/Hcs(z)
Inp(z)	= inverse of standard normal CDF
Lgd(z)	= logit density = Lgp*(1-Lgp)
Lgm(z)	= log of gamma
Lgt(z)	= logit = log(z/(1-z))
Lgp(z)	= logit probability = 1/(1+Exp(-z)) = Prob($Z \le z$)
Lmm(z)	= -N01(z)/Phi(z) = E[z | $z \le 0$] for z ~ N[0,1]
Lmp(z)	= N01(z)/Phi(-z) = E[z | $z \ge 0$] for z ~ N[0,1]
Log(z)	= natural logarithm
Max(z1,z2)	= maximum
Min(z1,z2)	= minimum
N01(z)	= standard normal density
Phi(z)	= standard normal CDF
Psi(z)	= log derivative of Gma, $\Psi = \Gamma'/\Gamma$
Psp(z)	= $\Psi' = \Gamma''/\Gamma - \text{Psi}^2$
Sgn(z)	= signum = -1 if z < 0, 0 if z = 0, +1 if z > 0
Sin(z)	= trigonometric sine
Tvm(z)	= 1 - Lmm × (z + Lmm) = Var[z | z < 0] for z ~ N[0,1]
Tvp(z)	= 1 - Lmp × (z + Lmp) = Var[z | z < 0] for z ~ N[0,1]

The incomplete beta function is

$$\text{Bds}(z,a,c) \;=\; \frac{\Gamma(a)\Gamma(c)}{\Gamma(a+c)}\int_0^z \; t^{a-1}(1-t)^{b-1}dt \text{ for } 0 < z < 1.$$

In the beta and bivariate normal functions, if any of the parameters separated by commas are expressions, it is necessary to enclose them in parentheses. I.e., use **Bvn((1+x'b),z,r)**, not **Bvn(1+x'b,z,r)**. The list may contain variables, labels, scalars, and expressions contained in parentheses. Functions may be nested to any depth and expressions may appear as arguments in the functions, as in

; Fcn = Log (Phi(a1 + a2 * (x/y)^2)).

This would be a valid expression and would evaluate exactly as given.

E14.3.3 Linear Functions and Dot Products

Many expressions in econometric models will involve dot products of parameters and variables. For example, a model built as an extension of a probit model will likely involve an expression of the form **Phi(b'x)**. Dot products may appear in exactly this form in your function definitions. Typically, the '**x**' would be a namelist. To use the parameter vector, use the first name in your labels list. For example, in

```
NAMELIST    ;  x = one,x1,z,p  $
NLSQ        ; Labels = b0,b1,b2,b3
            ; Fcn = ... Phi(b0'x)  ; ... $
```

the term **b0'x** is evaluated as $b0\times one + b1\times x1 + b2\times z + b3\times p$. Once again, in a dot product, the sum is evaluated from left to right using your list of labels in the order in which they appear in **; Labels = list**. If the namelist and the labels list do not have the same number of elements, then the dot product is simply evaluated out to the shorter of the two lists. In the example, if there were additional names in *x*, they would not change **b0'x** because starting at *b*0, there are only four parameters.

NOTE: This replaces the function Dot[.] used in earlier versions of *LIMDEP*. The Dot[.] function is retained for backwards compatibility, though you will probably find it easier to use the more natural syntax. Also, the operation described above does allow a bit more flexibility. For completeness, we note the counterparts to the constructions described above are **Dot[x]** = **b0'x** and **Dot[b3,second]** = **b3'second**. You may use either form.

Suppose you want to pick up just a few of the parameters in a dot product. For example, suppose your parameters are **; Labels = b1,b2,b3,b4,b5,b6,b7** and as part of your function, you want $b3*x14 + b4*xyz + b5*wvs$. You could first define the namelist for the dot product function, with, say,

```
NAMELIST    ; second  = x14,xyz,wvs $
```

Then, to obtain that function, just begin the dot product with **b3** instead of **b1**. Thus, **b3'second** evaluates exactly to $b3*x14 + b4*xyz + b5*wvs$.

It is also possible to skip over parameters in dot products, by putting columns of zeros in your namelists. This may be convenient in specifying your function, especially if it involves many parameters. For example, using the list above, you could obtain $b2 \times x14 + b5 \times xyz$

```
CREATE      ; zero = 0 $
NAMELIST    ; second = x14,zero,zero,xyz $
NLSQ        ; ... b2'second ...
```

Dot products need not be only a mix of variables and parameters. They may also include vectors (matrices) that do not appear elsewhere in the function, and they may be products of variables or parameters. When you are specifying your functions, there are several ways you can shorten your commands by making use of the dot product notation, and using lists. The following constructions can all be used in specifying your functions: Let

a and **d** denote the names of any vectors in your matrix work area,
x and **y** denote the names of any namelists,
cj be any of the labels in your **; Labels = ...** specification.

Then, any of the following can appear in your function

a'a = inner product of the vector,
a'd = dot product of two vectors,
a'x = linear combination of variables, for each observation,
x'y = sum of cross products of the variables, at each observation,
x'x = sum of squares of observation on variables,
cj'a = product of vector elements and parameters,
cj'x = the familiar index function product of coefficients and variables.

Products can be computed beginning with any of the parameters in the list. For example, consider fitting a probit model by least squares (rather than maximum likelihood):

```
NLSQ        ; Lhs = y
            ; Labels = a1,a2,a3
            ; Start = 0,0,0
            ; Fcn = Log(Phi(2*(y-1)*(a1 + a2*x1 + a3*x2))) $
```

Alternatively, with

```
NAMELIST    ; xa = one,x1,x2  ; xb = x1,x2 $
```

Then

```
            ; Fcn = Log(Phi(2*(y-1) * a1'xa))
```

is the same as

```
            ; Fcn = Log(Phi(2*(y-1) * (a1 + xb'a2))).
```

E14.3.4 Bilinear and Quadratic Forms

Bilinear and quadratic forms may also appear in function definitions. Suppose that *c* and *d* indicate elements of the parameter vector, which point to specific parts of the vector, and **z** is a namelist and **A** is a matrix. The following forms may appear in your function definition

(bilinear) **c'**[**z**]**d** $= \Sigma_j \, c_j \, d_j \, z_j$,

c'[**z**]**c** $= \Sigma_j \, c_j^2 z_j$

(quadratic) **c'**[**A**]**c** $= \Sigma_j \Sigma_l \, c_j \, c_l \, A_{jl}$

E14.3.5 Automatically Generating a List of Labels

For large problems, you may use a shortcut for the labels definition,

; Labels = number_label

produces 'number' sequentially numbered repetitions of the label. For example, **5_b** gives **b1,b2,b3,b4,b5**. The number may be a literal value or a scalar. With this device, you can make your model command independent of the size of the model, and you can accommodate a model of any size. For example:

```
NAMELIST    ; xa  =  ... (up to 150 names)
            ; xb  =  ... (up to 150 names) $
CALC        ; ka  =  Col(xa)
            ; kb  =  Col(xb) $
MATRIX      ; ca  =  Init(ka,1,0.)
            ; cb  =  Init(kb,1,0.0) $
NLSQ        ; Lhs =  y
            ; Start =  ca,cb,
            ; Labels =  ka_ba , kb_bb
            ; Fcn =  Index = ba1'xa  +  bb1'xb |
                     ... the rest of the function $
```

This template could be used for a model of any size. Only the namelists would have to be changed from one specification to another.

E14.3.6 Lists of Labels

The label list may be the object of a **CLIST** command. For example,

```
CLIST       ; probfn = pr0,pr1,pr2 $
NLSQ        ; …
            ; Labels = probfn $
```

E14.4 Quadrature and Simulation

You can use the function optimization programs such as **NLSQ** to maximize or minimize functions that contain integrals of the form

$$F(\boldsymbol{\beta}) \;=\; \int_{-\infty}^{\infty} \exp(-v^2)G(\boldsymbol{\beta},v)dv$$

by using Gauss-Hermite quadrature. This is a very accurate approximation which is computed using

$$F(\boldsymbol{\beta}) \;\approx\; \sum_{h=1}^{H} w_h G(\boldsymbol{\beta}, z_h)$$

where H is the number of points for the quadrature, w_h is the weight and z_h is the node at point h of the quadrature. You set the number of points, H for the quadrature. The G(.) function is unrestricted – it can be any function that is allowable in **NLSQ**, **NLSURE**, **MINIMIZE**, or **MAXIMIZE**. The variable of the integration, v, may or may not actually appear in the function. (Exp($-v^2$) integrates to sqr(π), so if v does not appear in G(.,.), then $F(\boldsymbol{\beta})$ will equal sqr(π)$G(\boldsymbol{\beta})$.) You can also include functions of the form

$$F(\boldsymbol{\beta}) \;=\; \int_{0}^{\infty} \exp(-v\ \ )G(\boldsymbol{\beta},v)dv$$

(notice that the exponent is exp($-v$) rather than exp($-v^2$), and the range of integration is from 0 to +∞ rather than from -∞ to ∞. Integrals of this form are accurately approximated using Gauss-Laguerre integration, rather than Gauss-Hermite integration. Finally, you can include functions that include subfunctions that are expectations of the form

$$F(\boldsymbol{\beta}) \;=\; E_v\,[F(\boldsymbol{\beta},v)].$$

where v is distributed as standard normal. These can be approximated quite accurately by simulation, by using

$$F(\boldsymbol{\beta}) \;\approx\; (1/R)\sum_{r=1}^{R} \;F(\boldsymbol{\beta},v_r)$$

where v_r is one of a sufficiently large R random draws from the standard normal distribution.

To use one of these integrals in your regression function, you must set up the operation as follows:

NLSQ **; Lhs = dependent variable**
; Fcn = name = Ntg(the function to be integrated) | the rest of the function, which will probably involve 'name'
; Hrq = name of the variable over which integration is done for Hermite integration
or **; Glq = name** of the variable over which integration is done for Gauss-Laguerre integration
or **; Sim = name** of the variable over which summation is done for integration by simulation
; Start = parameter values, as usual
; Labels = labels for parameters in the model, as usual
; other options $

Note the ' | ' at the end of the first line of the function definition. The function is being defined recursively. Recursive function definitions are described in the next section. The following requirements apply:

- You can have more than one integral in the final function, but each must be a named subfunction. If you specify 'Ntg(...)' within a function definition, an error will occur during compilation claiming that you have an unidentified symbol.

- Integrals should not be functions of other integrals. The results will be unpredictable, and almost certainly incorrect.

- You may have only one kind of integral in your function definition. Each Hrq, Glq, or Sim which appears in a command overrides previous ones.

To set the number of points for the approximation, you will use (as with other applications)

; Hpt = number of points for Hermite quadrature
; Lpt = number of points for Laguerre quadrature
; Pts = number of points for simulations

See Section R26.7 for discussion of the available values for these parameters.

NOTE: The seed for the random number generator is set to the same value each time a computation is done for a specific individual. Thus, you can replicate a computation done earlier by setting the main seed for the program before estimation.

Two examples follow. Note that this is not necessarily a 'good model,' and unless the data actually do satisfy the assumptions of the model, estimation will not produce very appealing results. (One would not normally fit a probit model by nonlinear least squares.) The example is intended only to illustrate use of the tools.

Heterogeneity in a Probit Model

Consider a probit model in which there is normally distributed, unobserved individual heterogeneity,

$$y = 0 \text{ or } 1,$$

$$\text{Prob}[\, y = 1 \mid v\,] = \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta v) \text{ where } v \text{ is standard normally distributed.}$$

(A nonunitary standard deviation of v would be absorbed into the free parameter θ.) The probability that enters the log likelihood is Prob[y = 1] = E_v [Prob[y = 1 | v]]. The expectation is exactly equal to

$$\text{Prob}\,[\, y = 1\,] = \int_{-\infty}^{\infty} (1/\sqrt{2\pi})\exp(-v^2/2)\Phi(\boldsymbol{\beta}'\mathbf{x} + \theta v)dv = P(y).$$

In the integral, let $u = v/\sqrt{2}$, so $v = u\sqrt{2}$. Make the change of variable in the integral, to produce

$$\text{Prob}\,[\, y = 1\,] = \int_{-\infty}^{\infty} (1/\sqrt{\pi})\exp(-u^2)\Phi(\boldsymbol{\beta}'\mathbf{x} + \theta u\sqrt{2})du = P(\text{y})$$

This is now exactly in the form noted earlier for Hermite quadrature. (It can be simplified a bit more by defining $\gamma = \sqrt{2}\,\theta$.) The model could be estimated with the commands

```
NAMELIST    ; x = ... list of variables $
PROBIT      ; Lhs = y ; Rhs = x $
CALC        ; kx = Col(x) $
NLSQ        ; Lhs = y              ? Note the separator for the subfunction.
            ; Labels = kx_b, c
            ; Start = b, 0
            ; Fcn = Prob = Ntg(1/Sqr(pi) * Phi(b1'x + c*u)) | Prob
            ; Hrq = u
            ; Hpt = 20 $
```

A second way to approximate the expected value would be by simulation and averaging. That is, the probability can be approximated by averaging the probabilities obtained with a sample of random draws from the distribution of v. The change in the preceding would be only to the method of integration. The resulting **NLSQ** command would be

```
NLSQ        ; Lhs = y
            ; Fcn = Prob = Ntg(1/Sqr(pi) * Phi(b1'x + t*Sqr(2)*u)) | Prob
            ; Start = b, 0
            ; Labels = kx_b, t
            ; Sim = u
            ; Pts = 100 $
```

E14.5 Recursive Functions

There are many settings in which certain parts of a regression function or a minimand involve constructions that appear more than once in a function. For example, consider the nonlinear regression function

$$E[y|\mathbf{x}] = \boldsymbol{\beta}'\mathbf{x} + \sigma \times \phi((L\text{-}\boldsymbol{\beta}'\mathbf{x})/\sigma) / \Phi((L\text{-}\boldsymbol{\beta}'\mathbf{x})/\sigma)$$

(this is the conditional mean function for a truncated regression model). Based on the preceding, the function could be specified with

```
NAMELIST    ; x = ... $
NLSQ        ; Lhs = y
            ; Start = the set of starting values
            ; Labels = b0,b1,b2,...,sg
            ; Fcn = b0'x + sg * N01((L-b0'x)/sg) / Phi((L-b0'x)/sg) $
```

The string **(L-b0'x)/sg** appears twice in the function definition. (In some settings, this sort of construction could appear many times.) You can build up such a function recursively by defining parts of it by name, then using the names of the parts later in the function. For the example above, an alternative form would be

```
; Fcn = dev = (L-b0'x)/sg                  |
        b0'X + sg*N01(dev)/Phi(dev)
```

The general form of a recursive definition is

```
; Fcn = name = string                      |
        ... next string can use name ...   |
        ...                                |
        function
```

Note that the ‘ | ’ character is used to separate the named strings from the parts that use them later. The last substring to be evaluated must produce the desired function. You may define up to 49 substrings in this fashion – the 50th would have to give the function, itself. Also, subfunctions may use earlier subfunctions. For examples,

```
; Fcn = cx   = c0+c1*x     |
        e    = y-cx        |
        e^2
```

or

```
; Fcn = cx   = c0+c1*x     |
        e    = y-cx        |
        esq  =  e^2        |
        e^2 + esq^2
```

NOTE: Functions are interpreted from left to right, or top to bottom. If you use a name which is defined *after* the function you are defining, an error will occur in which the name you are using does not appear to have been defined.

E14.6 Providing Analytic Derivatives

In computing nonlinear least squares estimates, *LIMDEP* uses numeric (synthetic) derivatives of the sum of squares. You can provide your own derivatives for some or all of the parameters in the function as follows: The regression function is $f(\beta_1,...,\beta_K$, anything else). You can provide derivatives for the regression function (not the sum of squares). Derivatives are specified in the same fashion as the subfunctions above, except that the name is the parameter label, preceded by an underscore. For example, suppose we were estimating a probit model by nonlinear least squares. The conditional mean would be $\Phi(\boldsymbol{\beta}'\mathbf{x})$, and the derivative vector is $\phi(\boldsymbol{\beta}'\mathbf{x}) \times \mathbf{x}$. Use

```
NLSQ      ; ... ; Labels = c0, c1
          ; Fcn = cx    = c0 + c1*x   |
                  fcx   = N01(cx)     |
                  _c0   = fcx         |
                  _c1   = fcx *x      |   Phi(cx) $
```

The function definition above contains both a subfunction and analytic derivatives. You may provide derivatives for any of the parameters, or all of them. Any derivatives that are not provided are evaluated numerically. Regardless of the complexity of the function, there is no difference in the amount of time saved by giving explicit derivatives for one parameter as opposed to another. Some time is saved if all derivatives are provided compared to just some of them.

WARNING: If the derivatives that you provide do not match the function, the optimization procedure will eventually break down, claiming to be unable to minimize the function.

To illustrate the use of analytic derivatives, we use one of the National Institute of Standards and Technology (NIST) benchmark problems for testing nonlinear regression programs. This test problem may be found at http://www.itl.nist.gov/div898/strd/nls/data/hahn1.shtml. The data are at the same site. (See Section E14.13 for further details.) The nonlinear regression model is

$$\text{Hahn1 Function: } y = \frac{\beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3}{1 + \beta_5 x + \beta_6 x^2 + \beta_7 x^3} + \varepsilon \quad \text{(average level of difficulty)}$$

```
Description:   These data are the result of a NIST study involving
               the thermal expansion of copper.  The response
               variable is the coefficient of thermal expansion, and
               the predictor variable is temperature in degrees kelvin.
Reference:     Hahn, T., NIST (197?). Copper Thermal Expansion Study.
Data:          1 Response  (y = coefficient of thermal expansion)
               1 Predictor (x = temperature, degrees kelvin)
               236 Observations
Model:         Rational Class (cubic/cubic)
               7 Parameters (b1 to b7)
     Starting values                  Certified Values
        Start 1     Start 2           Parameter     Standard Deviation
  b1 =   10          1             1.0776351733E+00  1.7070154742E-01
  b2 =   -1         -0.1          -1.2269296921E-01  1.2000289189E-02
  b3 =    0.05       0.005         4.0863750610E-03  2.2508314937E-04
  b4 =   -0.00001   -0.000001     -1.4262662514E-06  2.7578037666E-07
  b5 =   -0.05      -0.005        -5.7609940901E-03  2.4712888219E-04
  b6 =    0.001      0.0001        2.4053735503E-04  1.0449373768E-05
  b7 =   -0.000001  -0.0000001    -1.2314450199E-07  1.3027335327E-08
```

```
Residual Sum of Squares:                    1.5324382854E+00
Residual Standard Deviation:                8.1803852243E-02
Degrees of Freedom:                              229
Number of Observations:                          236
```

Data on *y* are followed by data on *x* in the listing below.

```
  0.591   1.547   2.902   2.894   4.703   6.307   7.030   7.898   9.470   9.484  10.072
 10.163  11.615  12.005  12.478  12.982  12.970  13.926  14.452  14.404  15.190  15.550
 15.528  15.499  16.131  16.438  16.387  16.549  16.872  16.830  16.926  16.907  16.966
 17.060  17.122  17.311  17.355  17.668  17.767  17.803  17.765  17.768  17.736  17.858
 17.877  17.912  18.046  18.085  18.291  18.357  18.426  18.584  18.610  18.870  18.795
 19.111   0.367   0.796   0.892   1.903   2.150   3.697   5.870   6.421   7.422   9.944
 11.023  11.870  12.786  14.067  13.974  14.462  14.464  15.381  15.483  15.590  16.075
 16.347  16.181  16.915  17.003  16.978  17.756  17.808  17.868  18.481  18.486  19.090
 16.062  16.337  16.345  16.388  17.159  17.116  17.164  17.123  17.979  17.974  18.007
 17.993  18.523  18.669  18.617  19.371  19.330   0.080   0.248   1.089   1.418   2.278
  3.624   4.574   5.556   7.267   7.695   9.136   9.959   9.957  11.600  13.138  13.564
 13.871  13.994  14.947  15.473  15.379  15.455  15.908  16.114  17.071  17.135  17.282
 17.368  17.483  17.764  18.185  18.271  18.236  18.237  18.523  18.627  18.665  19.086
  0.214   0.943   1.429   2.241   2.951   3.782   4.757   5.602   7.169   8.920  10.055
 12.035  12.861  13.436  14.167  14.755  15.168  15.651  15.746  16.216  16.445  16.965
 17.121  17.206  17.250  17.339  17.793  18.123  18.490  18.566  18.645  18.706  18.924
 19.100   0.375   0.471   1.504   2.204   2.813   4.765   9.835  10.040  11.946  12.596
 13.303  13.922  14.440  14.951  15.627  15.639  15.814  16.315  16.334  16.430  16.423
 17.024  17.009  17.165  17.134  17.349  17.576  17.848  18.090  18.276  18.404  18.519
 19.133  19.074  19.239  19.280  19.101  19.398  19.252  19.890  20.007  19.929  19.268
 19.324  20.049  20.107  20.062  20.065  19.286  19.972  20.088  20.743  20.830  20.935
 21.035  20.930  21.074  21.085  20.935
 24.410  34.820  44.090  45.070  54.980  65.510  70.530  75.700  89.570  91.140  96.400
 97.190 114.260 120.250 127.080 133.550 133.610 158.670 172.740 171.310 202.140 220.550
221.050 221.390 250.990 268.990 271.800 271.970 321.310 321.690 330.140 333.030 333.470
340.770 345.650 373.110 373.790 411.820 419.510 421.590 422.020 422.470 422.610 441.750
447.410 448.700 472.890 476.690 522.470 522.620 524.430 546.750 549.530 575.290 576.000
625.550  20.150  28.780  29.570  37.410  39.120  50.240  61.380  66.250  73.420  95.520
107.320 122.040 134.030 163.190 163.480 175.700 179.860 211.270 217.780 219.140 262.520
268.010 268.620 336.250 337.230 339.330 427.380 428.580 432.680 528.990 531.080 628.340
253.240 273.130 273.660 282.100 346.620 347.190 348.780 351.180 450.100 450.350 451.920
455.560 552.220 553.560 555.740 652.590 656.200  14.130  20.410  31.300  33.840  39.700
 48.830  54.500  60.410  72.770  75.250  86.840  94.880  96.400 117.370 139.080 147.730
158.630 161.840 192.110 206.760 209.070 213.320 226.440 237.120 330.900 358.720 370.770
372.720 396.240 416.590 484.020 495.470 514.780 515.650 519.470 544.470 560.110 620.770
 18.970  28.930  33.910  40.030  44.660  49.870  55.160  60.900  72.080  85.150  97.060
119.630 133.270 143.840 161.910 180.670 198.440 226.860 229.650 258.270 273.770 339.150
350.130 362.750 371.030 393.320 448.530 473.780 511.120 524.700 548.750 551.640 574.020
623.860  21.460  24.330  33.430  39.220  44.180  55.020  94.330  96.440 118.820 128.480
141.940 156.920 171.650 190.000 223.260 223.880 231.500 265.050 269.440 271.780 273.460
334.610 339.790 349.520 358.180 377.980 394.770 429.660 468.220 487.270 519.540 523.030
612.990 638.590 641.360 622.050 631.500 663.970 646.900 748.290 749.210 750.140 647.040
646.890 746.900 748.430 747.350 749.270 647.610 747.780 750.510 851.370 845.970 847.540
849.930 851.610 849.750 850.980 848.230
```

The NIST problems are provided with two sets of starting values for the iterations, as seen in the example above. The first set of values are always farther from the solution than the second, so estimation beginning with the first set is always ‘harder’ than with the second set. In the following, we solve the problem from the first set of starting values. (We chose this particular problem out of the 27 problems provided because *LIMDEP* is able to solve all but three, including this one, using the default settings without using analytic derivatives.)

The direct approach is the 'default' solution: (the **; Output** and **; Dfc** options are described in Section E14.7.4.) *LIMDEP* is not able to solve this problem with the direct solution.

READ **; Nobs = 236 ; Nvar = 2 ; Names = y,x ; By Variables $**
(The data follow, exactly as shown above)
CREATE **; x2 = x*x ; x3 = x2*x$**
NLSQ **; Lhs = y**
; Fcn = (b1+b2*x+b3*x2+b4*x3)/(1+b5*x+b6*x2+b7*x3)
; Labels = b1,b2,b3,b4,b5,b6,b7
; Output = 1
; Dfc
; Maxit = 500
; Start = 10,-1,.05,-.00001,-.05,.001,-.000001$

After 500 iterations, this formulation exits at a point that is nowhere near the correct solution:

```
Iteration=500; Sum of squares=  7.99101926    ; Gradient=  .106836976E-02
Maximum iterations exceeded
```

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=Y        Mean                 =      14.21530
             Standard deviation   =       5.76869
             Number of observs.   =           236
Model size   Parameters           =             7
             Degrees of freedom   =           229
Residuals    Sum of squares       =       7.99102
             Standard error of e  =        .18680
Fit          R-squared            =        .99898
             Adjusted R-squared   =        .99895
Model test   F[  6,   229] (prob) = 37313.0(.0000)
Diagnostic   Log likelihood       =      64.62109
             Restricted(b=0)      =    -747.94531
             Chi-sq [  6]  (prob) =1625.1(  .0000)
Info criter. Akaike Info. Criter. =      -3.32619
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|    10.3875***      .40345    25.75  .0000      9.5968   11.1783
      B2|    -.92729***      .02498   -37.12  .0000     -.97626   -.87832
      B3|     .02148***      .00041    52.11  .0000      .02067    .02229
      B4|-.19842D-04***   .1451D-06  -136.72  .0000 -.20126D-04  -.19557D-04
      B5|     .00600***      .00117     5.11  .0000      .00370    .00830
      B6|     .00102***   .2402D-04    42.61  .0000      .00098    .00107
      B7|-.10164D-05***   .2415D-10 ********  .0000 -.10164D-05  -.10163D-05
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

We now reformulate the problem with analytic derivatives. Note that the derivatives for *b*6 and *b*7 are formulated recursively, using their predecessors as if they were ordinary, predefined functions (which they are).

```
NLSQ    ; Lhs = y
        ; Fcn = top  = (b1+b2*x+b3*x2+b4*x3)   |
                bot  = (1+b5*x+b6*x2+b7*x3)    |
                q    = 1/bot                   |
                _b1 = q                        |
                _b2 = x*q                      |
                _b3 = x2*q                     |
                _b4 = x3*q                     |
                _b5 = -top * q * q * x         |
                _b6 = _b5 * x                  |
                _b7 = _b6 * x                  |
                top * q   ?  This defines the regression function.
        ; Labels = b1,b2,b3,b4,b5,b6,b7
        ; Output= 1
        ; Dfc
        ; Start = 10,-1,.05,-.00001,-.05,.001,-.000001$
```

By changing the estimation to use analytic derivatives, we now obtain the NIST solution, after 98 iterations:

```
Begin NLSQ iterations. Linearized regression.
Iteration=  1; Sum of squares=  3097556.53    ; Gradient=  3097518.97
Iteration=  2; Sum of squares=  82100.0867    ; Gradient=  82037.2061
Iteration=  3; Sum of squares=  8813.71742    ; Gradient=  8799.72594
Iteration=  4; Sum of squares=  1142.50038    ; Gradient=  1139.39827
Iteration=  5; Sum of squares=  25.2098528    ; Gradient=  21.9074256
```

(Iterations 6 - 19 omitted)

```
Iteration= 20; Sum of squares=  42.5269570    ; Gradient=  14.0951380
Iteration= 21; Sum of squares=  41.4433864    ; Gradient=  12.9637663
Moment matrix has become nonpositive definite.
Switching to BFGS algorithm
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .1000D-05 chg.F   .0000D+00 max|dB|   .0000D+00
Nodes for quadrature: Laguerre=40;Hermite=20.
Replications for GHK simulator= 100
```

(Initial iterations to improve starting values)

```
Itr  1 F=  .1549D+07 gtHg=  .1625D+14 chg.F=  .1549D+07 max|db|=  .1625D+20
Itr  2 F=  .2776D+05 gtHg=  .3355D+05 chg.F=  .1521D+07 max|db|=  .1036D+09
Itr  3 F=  .3913D+04 gtHg=  .2232D+04 chg.F=  .2385D+05 max|db|=  .1598D+07
Itr  4 F=  .3912D+04 gtHg=  .3306D+04 chg.F=  .1536D+01 max|db|=  .7356D+06
Itr  5 F=  .3910D+04 gtHg=  .2138D+04 chg.F=  .1731D+01 max|db|=  .6182D+06
```

(Iterative gradient method search for function optimizers)

```
Itr  1 F=  .3910D+04 gtHg=  .2138D+04 chg.F=  .3910D+04 max|db|=  .6182D+06
Itr  2 F=  .3908D+04 gtHg=  .3304D+04 chg.F=  .1701D+01 max|db|=  .1416D+07
Itr  3 F=  .2201D+04 gtHg=  .1220D+03 chg.F=  .1708D+04 max|db|=  .7837D+03
Itr  4 F=  .2197D+04 gtHg=  .5820D+02 chg.F=  .3989D+01 max|db|=  .1619D+04
Itr  5 F=  .1094D+04 gtHg=  .6585D+02 chg.F=  .1102D+04 max|db|=  .3680D+01
```

(Iterations 6 - 96 omitted)

```
Itr 97 F=  .7662D+00 gtHg=  .1386D-05 chg.F=  .4288D-10 max|db|=  .4059D-06
Itr 98 F=  .7662D+00 gtHg=  .9207D-07 chg.F=  .1668D-11 max|db|=  .1557D-06
                        * Converged
Normal exit from iterations. Exit status=0.
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=Y        Mean                 =      14.21530
             Standard deviation   =       5.76869
             Number of observs.   =           236
Model size   Parameters           =             7
             Degrees of freedom   =           229
Residuals    Sum of squares       =       1.53244
             Standard error of e  =        .08180
Fit          R-squared            =        .99980
             Adjusted R-squared   =        .99980
Model test   F[  6,   229] (prob) =194732.2(.0000)
Diagnostic   Log likelihood       =     259.49316
             Restricted(b=0)      =    -747.94531
             Chi-sq [  6]  (prob) =2014.9(  .0000)
Info criter. Akaike Info. Criter. =      -4.97765
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|    1.07764***      .17070     6.31  .0000      .74307   1.41220
      B2|    -.12269***      .01200   -10.22  .0000     -.14621   -.09917
      B3|     .00409***      .00023    18.15  .0000      .00365    .00453
      B4|-.14263D-05***   .2758D-06    -5.17  .0000 -.19668D-05  -.88575D-06
      B5|    -.00576***      .00025   -23.31  .0000     -.00625   -.00528
      B6|     .00024***   .1045D-04    23.02  .0000      .00022    .00026
      B7|-.12314D-06***   .1303D-07    -9.45  .0000 -.14868D-06  -.97611D-07
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E14.7 Options for Nonlinear Least Squares

The following lists the options available for the nonlinear least squares program.

E14.7.1 Fixing Some of the Parameters

It is sometimes necessary to minimize the sum of squares while holding some of the parameters fixed at known values, for example for hypothesis testing. To do so, you can simply specify the problem in the command exactly as if the parameters were not to be fixed. Include the known values in the appropriate places in the list of starting values. Then, add the specification

; Fix = parm1,parm2,...

where the names in the **; Fix = ...** list will be some of those in your **; Labels** list. For example, suppose you wanted to obtain a CES model with constant returns to scale. This is done by setting *nu* at 1.0.

```
NLSQ      ; Lhs = Q
          ; Labels = gamma,delta,r,nu
          ; Start = 2.3, .3,  .1,  1.
          ; Fix = nu
          ; Fcn = gamma*(delta*k^(-r) + (1-delta)*l^(-r))^(-nu/r) $
```

If you have fixed the values of parameters, with the **; Fix** option, these will be among the values placed in the matrix *b* when the model results are kept for later use.

You might want to compute the sum of squares function at a particular set of parameters. You can do this by specifying that *all* parameters are to be fixed at the starting values. This is

; Fix all

(not **; Fix = all**). The full set of output will be produced, but no iterations will be done. Note that **; Fix all** is the same as **; Maxit = 0**, except the latter also produces a Lagrange multiplier statistic.

Standard Model Specifications for the Nonlinear Regression Model

This is the full list of general specifications from Chapter E1. See Chapter E1 and references noted there for further details on these specifications.

Controlling Output from Model Commands

; Table = name saves model results to be combined later in output tables.
; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Robust Asymptotic Covariance Matrices

	; Cluster = spec	cluster form of corrected covariance estimator.
or	**; Robust**	requests a ‘sandwich’ estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Lpt = n	sets the number of points to use for Laguerre quadrature.
; Hpt = n	sets the number of points to use for Hermite quadrature.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; CML: spec defines a constrained maximum likelihood estimator.
; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions – same as **; Test: spec**.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values sets up Lagrange multiplier test of restrictions.
; Fix = list fixes the named parameters at the start values.

E14.7.2 Setting the Algorithm

Six algorithms are available: Gauss-Marquardt (the default), Davidon-Fletcher-Powell, Newton's method, Berndt, Hall, Hall, and Hausman, BFGS, and steepest descent. Unless your problem is globally convex (which is unlikely) you will probably want to use the first of these, which is the default. This is a very effective algorithm, which has been used in a wide variety of settings. Newton's method may require somewhat less computing owing to the necessity of BFGS and DFP to do a line search at each iteration; Newton's method uses a step length on 1.0. However, Newton's method is very likely to overshoot and subsequently to diverge. Choose the algorithm with

; Alg = Newton or BHHH or BFGS or DFP or Steepest Descent

Other options,

; Output = setting
; Tlb ; Tlf and ; Tlg
; Maxit = maximum
; Covariance Matrix (or ; Printvc)

all operate as usual. A setting that is specific to nonlinear least squares is

; Tln = convergence tolerance for Gauss-Marquardt method

E14.7.3 Heteroscedasticity Robust Covariance Matrix

The asymptotic covariance matrix for the nonlinear least squares estimator is estimated with

$$\mathbf{V}_{nls} = \frac{\mathbf{e}'\mathbf{e}}{n}\left(\mathbf{X^0}'\mathbf{X^0}\right)^{-1}$$

where the ith row of $\mathbf{X^0}$ is the vector of derivatives of the regression function,

$$\mathbf{x}_i^0 = \frac{\partial f(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$$

This is the counterpart to the usual calculation for linear least squares. You may request a heteroscedasticity robust covariance matrix of the form of the White estimator by adding

; Heteroscedasticity

to the **NLSQ** command. The matrix computed is

$$\mathbf{V}_{robust} = \left(\mathbf{X^0}'\mathbf{X^0}\right)^{-1} \left[\sum_{i=1}^{n}\left(e_i\mathbf{x}_i^0\right)\left(e_i\mathbf{x}_i^0\right)'\right] \left(\mathbf{X^0}'\mathbf{X^0}\right)^{-1}$$

(There is no counterpart for the Newey-West autocorrelation consistent estimator.)

E14.7.4 Degrees of Freedom Correction

In the expression above, the disturbance variance is estimated with the mean of the nonlinear least squares residuals, with no correction for degrees of freedom. Some authors prefer to make this correction, which produces the estimator

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-K}.$$

The program default is to use n rather than n-K in this computation. The estimator is consistent either way, but is not unbiased in either case. Computer programs differ in this computation, and users should check the documentation of other programs that they might be using if apparent inconsistencies arise. You can request this computation with

; Dfc (degrees of freedom correction)

added to the **NLSQ** command. (Note that this form is used in the NIST benchmark tests, so we have requested it in the illustration using the NIST problem.)

E14.8 Model Output and Retrievable Results

Output from **NLSQ** consists of a table of diagnostic statistics similar to that presented for the linear regression model and a standard table of results for the parameters. The one produced by the preceding example is shown below.

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=Y         Mean                     =       14.21530
              Standard deviation       =        5.76869
              Number of observs.       =            236
Model size    Parameters               =              7
              Degrees of freedom       =            229
Residuals     Sum of squares           =        1.53244
              Standard error of e      =         .08180
Fit           R-squared                =         .99980
              Adjusted R-squared       =         .99980
Model test    F[  6,    229] (prob)  =194732.2(.0000)
Diagnostic    Log likelihood           =      259.49316
              Restricted(b=0)          =     -747.94531
              Chi-sq [   6]  (prob)  =2014.9(  .0000)
Info criter. Akaike Info. Criter. =          -4.97765
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error        z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
      B1|    1.07764***      .17070     6.31  .0000       .74307   1.41220
      B2|    -.12269***      .01200   -10.22  .0000      -.14621   -.09917
      B3|     .00409***      .00023    18.15  .0000       .00365    .00453
      B4|-.14263D-05***   .2758D-06    -5.17  .0000 -.19668D-05  -.88575D-06
      B5|    -.00576***      .00025   -23.31  .0000      -.00625   -.00528
      B6|     .00024***   .1045D-04    23.02  .0000       .00022    .00026
      B7|-.12314D-06***   .1303D-07    -9.45  .0000 -.14868D-06  -.97611D-07
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

(Note that the R^2 computed as above is not bounded by zero as it would be if this were a linear least squares computation. Likewise, the F statistic, when it is produced – the one above is huge – should be ignored.) The saved results are:

Matrices:	*b*	=	the parameters
	varb	=	estimated asymptotic covariance matrix
	gradient	=	the vector of first derivatives

If you have fixed parameters, *varb* will contain rows and columns of zeros. Unless you have fixed some or all of the parameters, limited the number of iterations to less than necessary to obtain convergence, or the optimization fails, *gradient* will be approximately zero (to within the limit of your convergence criterion).

Scalars: *s* = $\sqrt{}$(1/n×sum of squared residuals,)
(The denominator is n - K if **; Dfc** is specified.)
ssqrd = s squared
ybar = mean of Lhs variable
sy = standard deviation of Lhs variable
sumsqdev = sum of squared residuals
rsqrd = 1 - *sumsqdev* / ((n-1)×**SY**×**SY**)
rho = 0.0
degfrdm = n - number of free (not fixed) parameters
kreg = total number of parameters
nreg = current sample size
logl = log likelihood = $-(n/2)[1+\log 2\pi+\log(\mathbf{e}'\mathbf{e}/n)]$

Last Function: See Section E14.9

The standard options,

; List	to display predicted values
; Keep = name	to retain for predictions, $F(\mathbf{x}_i,\boldsymbol{\beta})$
; Res = name	to retain for residuals, $y - F(\mathbf{x}_i,\boldsymbol{\beta})$
; Fill	to fill in values for unused observations

all operate as usual. If you use **; List**, your listing will contain a list of the Lhs variable and the values of $F(\mathbf{x}_i,\boldsymbol{\beta})$. Under the heading '95% Confidence Interval' will be two columns of zeros which can be ignored. The values of $F(\mathbf{x}_i,\boldsymbol{\beta})$ evaluated at the final estimates are the fitted values for this command.

For purposes of using **WALD**, the *Last Model* kept uses the labels in your **; Labels** list. To continue the earlier application, for example, to test the hypothesis that ρ equals 0 and ν equals 1 in the CES function, you could use

```
NLSQ    ; Lhs = Q
        ; Labels = gamma,delta,r,nu
        ; Start = 2.3, .3,  .1,  1.
        ; Fcn = gamma*(delta*K^(-r) + (1-delta)*L^(-r))^(-nu/r) $
WALD    ; Fn1 = r ; Fn2 = nu - 1 $
```

E14.9 Partial Effects for Nonlinear Regressions

You can obtain partial effects for any function. NLSQ does not save the function definition for computation of the partial effects or for a simulation. However, your command editor contains the function definition that you need to obtain these. NLSQ does save the parameter vector and the covariance matrix as the *Last Function*. You can use these just by reconstructing the function in the **PARTIALS** or **SIMULATE** command and using the parameters and covariance from the estimation step.

Consider the earlier example,

```
NLSQ        ; Lhs = y
            ; Fcn = top = (b1+b2*x+b3*x2+b4*x3)      |
                    bot = (1+b5*x+b6*x2+b7*x3)       |
                    q   = 1/bot                      |
            (derivatives omitted)
                    top * q   ?  This defines the regression function.
            ; Labels = b1,b2,b3,b4,b5,b6,b7
            ; Dfc
            ; Start = 10,-1,.05,-.00001,-.05,.001,-.000001$
```

We used a simplification to fit the function, substituting $x2 = x^2$ for $x*x$ and $x3 = x^3$ for $x*x*x$ in the optimization. However, the regression is, ultimately, a function only of *x*. To examine the behavior of the function after nonlinear least squares, we used the following:

```
SIMULATE    ; Scenario: & x = 0(50)1000
            ; Plot (ci)
            ; Function = (b1+b2*x+b3*x*x+b4*x*x*x) /
                         (1 + b5*x+b6*x*x+b7*x*x*x)
            ; Labels = b1,b2,b3,b4,b5,b6,b7 $
PARTIALS    ; Effects: x  & x = 0(50)1000
            ; Plot
            ; Function = (b1+b2*x+b3*x*x+b4*x*x*x) /
                         (1 + b5*x+b6*x*x+b7*x*x*x)
            ; Labels = b1,b2,b3,b4,b5,b6,b7 $
```

Note that we have replaced the shortcuts for the powers of *x*. This was useful for the optimization, but not for the simulation which needs the actual functional form to get the right derivatives. The parameter values and covariance matrix have been stored when the results were reported. The scenarios examine the values of *x* ranging from zero to 1,000 in steps of 50, which is roughly the range of *x* in the data. The simulation produces a plot with a confidence interval.

```
-----------------------------------------------------------------------
Model Simulation Analysis for User Specified Function
-----------------------------------------------------------------------
Simulations are computed by average over sample observations
-----------------------------------------------------------------------
User Function       Function   Standard
(Delta method)       Value      Error      |t|  95% Confidence Interval
-----------------------------------------------------------------------
Avrg. Function     14.21530     .50208   28.31    13.23123     15.19936
X        =   .00    1.07764     .17070    6.31      .74306      1.41221
X        = 50.00    3.83746     .01635  234.66     3.80541      3.86951
X        =100.00   10.43772     .01605  650.26    10.40626     10.46918
X        =150.00   13.60074     .08362  162.66    13.43685     13.76463
X        =200.00   15.15580    1.49077   10.17    12.23388     18.07772
X        =250.00   16.06395     .75142   21.38    14.59116     17.53673
X        =300.00   16.68267     .57523   29.00    15.55522     17.81013
X        =350.00   17.15886     .57946   29.61    16.02312     18.29460
X        =400.00   17.56087     .62862   27.94    16.32877     18.79298
X        =450.00   17.92426     .70036   25.59    16.55155     19.29697
X        =500.00   18.26942     .78809   23.18    16.72476     19.81408
X        =550.00   18.60921     .88957   20.92    16.86564     20.35277
X        =600.00   18.95252    1.00427   18.87    16.98415     20.92090
X        =650.00   19.30615    1.13254   17.05    17.08637     21.52594
X        =700.00   19.67581    1.27532   15.43    17.17617     22.17544
X        =750.00   20.06672    1.43404   13.99    17.25599     22.87744
X        =800.00   20.48407    1.61061   12.72    17.32728     23.64087
X        =850.00   20.93336    1.80747   11.58    17.39072     24.47600
X        =900.00   21.42062    2.02769   10.56    17.44635     25.39489
X        =950.00   21.95278    2.27512    9.65    17.49355     26.41200
X        =******   22.53797    2.55458    8.82    17.53099     27.54495
```

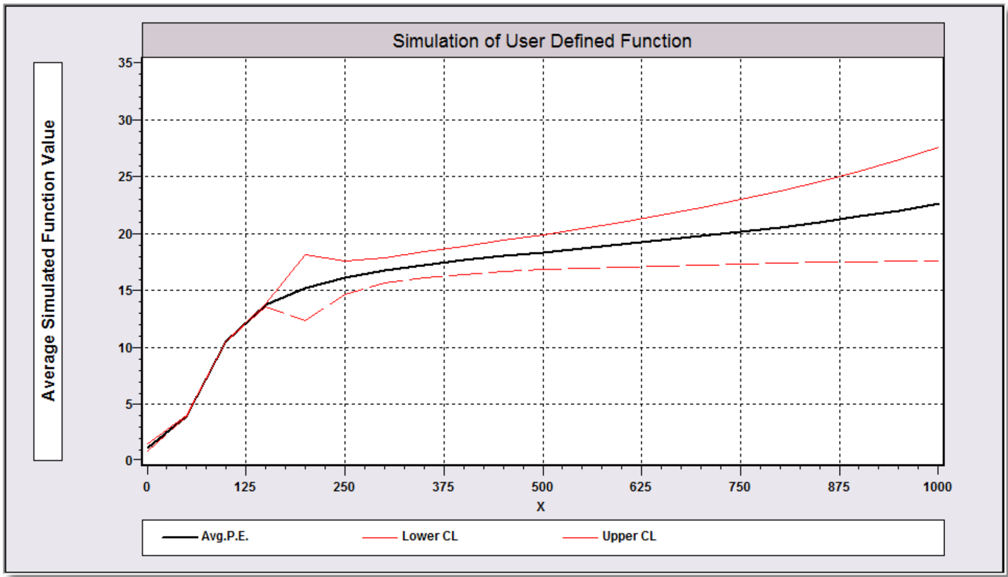


Figure E14.1 Simulation of Nonlinear Regression

```
-----------------------------------------------------------------------
Partial Effects  Analysis for User Specified Function
-----------------------------------------------------------------------
Effects on function with respect to X
Results are computed by average over sample observations
Partial effects for continuous X        computed by differentiation
Effect is computed as derivative     = df(.)/dx
-----------------------------------------------------------------------
df/dX                  Partial    Standard
(Delta method)         Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
APE. Function          .03999      .16168      .25      -.27690      .35688
X        =    .00    -.11648      .01103    10.56      -.13810     -.09487
X        = 50.00      .16072      .00101   159.70       .15874      .16269
X        =100.00      .09177      .00036   258.36       .09108      .09247
X        =150.00      .04226      .00379    11.16       .03484      .04968
X        =200.00      .02284      .17518      .13      -.32051      .36619
X        =250.00      .01457      .00962     1.51      -.00429      .03342
X        =300.00      .01064      .00077    13.84       .00913      .01214
X        =350.00      .00863      .00067    12.97       .00732      .00993
X        =400.00      .00757      .00124     6.08       .00513      .01001
X        =450.00      .00703      .00161     4.38       .00388      .01018
X        =500.00      .00681      .00190     3.59       .00310      .01053
X        =550.00      .00681      .00216     3.15       .00257      .01104
X        =600.00      .00695      .00243     2.86       .00219      .01171
X        =650.00      .00721      .00271     2.67       .00191      .01252
X        =700.00      .00759      .00301     2.52       .00169      .01349
X        =750.00      .00807      .00335     2.41       .00151      .01463
X        =800.00      .00865      .00373     2.32       .00134      .01595
X        =850.00      .00934      .00416     2.25       .00119      .01750
X        =900.00      .01017      .00467     2.18       .00103      .01931
X        =950.00      .01114      .00526     2.12       .00084      .02144
X        =******      .01230      .00595     2.07       .00063      .02397
```

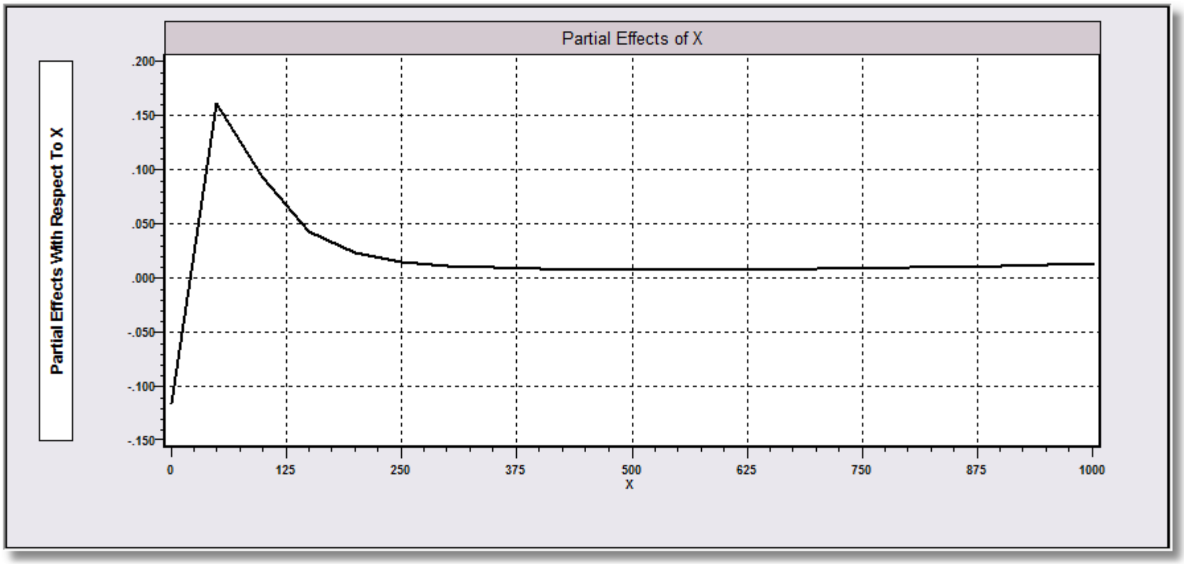


Figure E14.2 Partial Effects in Nonlinear Regression

E14.10 Imposing Restrictions and Testing Hypotheses

Since you are fully specifying the function for **NLSQ**, there is no need for a method of imposing restrictions on the parameters. For any that you desire, just build them into the function. If you wish to fix parameters at known values for the purpose of inference, use the **; Fix** option described in Section E14.7.1. After estimation, results can be retrieved for carrying out tests of restrictions on the parameters of the regression model. There are several methods of testing hypotheses with the results of **NLSQ**.

The Lagrange multiplier test described in Section R13.5 can be used if the starting values are restricted estimates from some other specification. Then,

; Maxit = 0

will produce a full set of results and the appropriate chi squared statistics.

With normally distributed disturbances, you can carry out *likelihood ratio tests* exactly as shown in Section R13.4. The scalar *logl* will contain the appropriate value for each model that you estimate.

Wald tests can be carried out by using the *Last Model* construction. See Section R13.3 for details. The labels to be used for the tests are those that appear in your **; Labels** list.

A form of F test can be based on the sum of squared residuals. The scalar *sumsqdev* contains the necessary statistic. The procedure would be

```
NLSQ        ; ... unrestricted model $
CALC        ; eeu = sumsqdev
            ; ku = kreg  $
NLSQ        ; ... restricted model $
CALC        ; eer = sumsqdev
            ; kr = kreg
            ; List
            ; dfn = ku - kr
            ; dfd = n - ku
            ; f = ((eer - eeu)/dfn) / (eeu/dfd)
            ; prob = 1 - Fds(f,dfn,dfd)  $
```

The theory surrounding this statistic for testing hypotheses is not so definitive as that for the Wald test. (See, e.g., Greene (2011) for discussion.)

The **; Test: ... restrictions** option is also available for **NLSQ**, but it will be a bit cumbersome to use. For purposes of this option if you wish to use it, the parameters in your **; Labels** list are renamed, as usual, $b(1),...,b(K)$. It will be easier to use **WALD** to obtain the same results.

E14.11 An Application

To illustrate the nonlinear least squares computation, we use a small set of data based on the Poisson regression model. This specification for a discrete random variable is

$$\text{Prob}[y_i = j] = e^{-\lambda_i} \lambda_i^{y_i} / y_i! \quad \text{where} \quad \lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}.$$

The conditional mean function is $E[y_i] = \lambda_i$.

IMPORT $
y,x1,x2,x3

```
 1  -0.545   0.160   0.033
 0   0.892   0.125   1.476
 2   1.647   0.619  -0.262
 2   1.749  -1.446   0.310
 2   0.362  -0.589  -1.404
 0   0.531  -0.606   0.777
 2   0.003  -0.800  -0.897
 0   0.260   0.597  -0.640
 3   1.502  -0.309   0.112
 0   0.613   0.273  -0.845
 0  -1.028  -0.307  -1.170
 2   0.155  -0.262  -0.534
 1  -1.795  -2.051  -0.398
 0  -1.007   1.974   0.189
 1   0.596  -0.493  -1.369
```

NAMELIST ; x = one,x1,x2,x3 $

We compute the linear regression twice, with **REGRESS** and with **NLSQ**

```
REGRESS     ; Lhs = y
            ; Rhs = x $
NLSQ        ; Lhs = y
            ; Start = 0,0,0,0
            ; Labels = b1,b2,b3,b4
            ; Fcn = b1'x
            ; Dfc $
```

The nonlinear regression is based on $E[y|\mathbf{x}] = \exp(\boldsymbol{\beta}'\mathbf{x})$

```
NLSQ        ; Lhs = y
            ; Start = 0,0,0,0
            ; Labels = b1,b2,b3,b4
            ; Fcn = Exp(b1'x)
            ; Output= 4
            ; Dfc $
```

```
--------------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                 =       1.06667
             Standard deviation   =       1.03280
             No. of observations  =            15  Degrees of freedom
Regression   Sum of Squares       =       5.78935           3
Residual     Sum of Squares       =       9.14398          11
Total        Sum of Squares       =       14.9333          14
             Standard error of e  =        .91174
Fit          R-squared            =        .38768  R-bar squared =   .22068
Model test   F[  3,    11]        =       2.32148  Prob F > F*   =  .13153
Diagnostic   Log likelihood       =     -17.57192  Akaike I.C.   =  .03838
             Restricted (b=0)     =     -21.25067  Bayes  I.C.   =  .22719
             Chi squared [  3]    =       7.35750  Prob C2 > C2* =  .06134
--------+-----------------------------------------------------------------------
        |              Standard            Prob.      95% Confidence
       Y|  Coefficient      Error       t    |t|>T*        Interval
--------+-----------------------------------------------------------------------
Constant|     .75513**      .27109     2.79  .0177      .22381   1.28646
      X1|     .50861*       .24599     2.07  .0630      .02648    .99074
      X2|    -.37994        .26275    -1.45  .1761    -.89493    .13505
      X3|    -.32196        .31345    -1.03  .3264    -.93632    .29240
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=Y        Mean                 =       1.06667
             Standard deviation   =       1.03280
             Number of observs.   =            15
Model size   Parameters           =             4
             Degrees of freedom   =            11
Residuals    Sum of squares       =       9.14398
             Standard error of e  =        .91174
Fit          R-squared            =        .38768
             Adjusted R-squared   =        .22068
Model test   F[  3,    11] (prob) =    2.3(.1315)
Diagnostic   Log likelihood       =     -17.57192
             Restricted(b=0)      =     -21.25067
             Chi-sq [  3]  (prob) =   7.4(  .0613)
Info criter. Akaike Info. Criter. =        .03838
Not using OLS or no constant. Rsqrd & F may be < 0
--------+-----------------------------------------------------------------------
(Estimates identical to those given above are omitted.)
--------------------------------------------------------------------------------
Begin NLSQ iterations. Linearized regression.
Iteration=  1; Sum of squares=  15.0000000    ; Gradient=  5.85601636
Iteration=  2; Sum of squares=  9.46277971    ; Gradient=  .601416926
Iteration=  3; Sum of squares=  8.86223571    ; Gradient=  .995192240E-02
Iteration=  4; Sum of squares=  8.84925264    ; Gradient=  .153166021E-02
Iteration=  5; Sum of squares=  8.84704954    ; Gradient=  .313593091E-03
Iteration=  6; Sum of squares=  8.84659536    ; Gradient=  .646589430E-04
Iteration=  7; Sum of squares=  8.84650149    ; Gradient=  .133647352E-04
Iteration=  8; Sum of squares=  8.84648206    ; Gradient=  .276532978E-05
Iteration=  9; Sum of squares=  8.84647804    ; Gradient=  .572424008E-06
Convergence achieved
```

```
-----------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ...........
LHS=Y        Mean                    =       1.06667
             Standard deviation      =       1.03280
             Number of observs.      =            15
Model size   Parameters              =             4
             Degrees of freedom      =            11
Residuals    Sum of squares          =       8.84648
             Standard error of e     =        .89679
Fit          R-squared               =        .40760
             Adjusted R-squared      =        .24604
Model test   F[  3,    11] (prob) =     2.5(.1116)
Diagnostic   Log likelihood          =     -17.32385
             Restricted(b=0)         =     -21.25067
             Chi-sq [  3]  (prob) =   7.9(  .0491)
Info criter. Akaike Info. Criter. =          .00530
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------
      B1|    -.42591        .47138     -.90  .3662    -1.34979    .49798
      B2|     .58062*       .31447     1.85  .0648     -.03574   1.19697
      B3|    -.30570        .27612    -1.11  .2682     -.84689    .23548
      B4|    -.36097        .33212    -1.09  .2771    -1.01192    .28997
--------+--------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------
```

E14.12 Technical Details

LIMDEP uses the Gauss-Marquardt method for nonlinear least squares. The iteration is

$$\mathbf{b}(k+1) = \mathbf{b}(k) + \left(\mathbf{X^0}(k)'\mathbf{X^0}(k)\right)^{-1}\mathbf{X^0}'(k)\mathbf{e^0}(k)$$

where the matrix of pseudo-regressors from the linearized regression has ith row equal to the transpose of

$$\mathbf{x}_i^0(k) = \frac{\partial f(\mathbf{x}_i, \mathbf{b}(k))}{\partial \mathbf{b}(k)'}$$

and $$e_i^0(k) = y_i - f(\mathbf{x}_i, \mathbf{b}(k)).$$

Convergence is measured by the ‘gradient’ measure,

$$\delta(k) = \mathbf{e^0}(k)'\left(\mathbf{X^0}(k)'\mathbf{X^0}(k)\right)^{-1}\mathbf{e^0}(k)$$

The tolerance value is 1.D-20. (Note the convergence assessment in the preceding example.

This iterative procedure can break down for two reasons. Since it is a Newton-like method without a controlled line search, it is possible for the estimated parameter vector to diverge, or to drift to a place in the parameter space where the regression function cannot be computed. If this happens, the message

```
Function is no longer computable.
```

will appear in the output. The second failure can occur when the current values of the parameters causes the second moment matrix to become nonpositive definite. For example, a regression function that includes exp($\mathbf{b}'\mathbf{x}$) might degenerate to a column of zeros for an extreme value of the parameters. In this case, the message

```
Moment matrix has become nonpositive definite.
```

will appear in your results. In both of these cases, *LIMDEP* will make one more attempt to fit the model, using a different algorithm. You will see the message

```
Switching to BFGS algorithm
Nonlinear Estimation of Model Parameter
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg .1000D-05 chg.F .0000D+00 max|dB| .0000D+00
Nodes for quadrature: Laguerre=40;Hermite=20.
Replications for GHK simulator= 100
```

followed by the standard output for nonlinear optimization. At this point, *LIMDEP* will have abandoned the procedure used only for nonlinear least squares, and switched to minimizing the sum of squares as an ordinary optimization problem. The starting values for this second attempt will be the ones initially used for the Gauss-Marquardt method. This is what occurred in the NIST example Hahn1 in Section E14.6.

All derivatives that you do not supply explicitly are obtained by numerical approximation, using a symmetric (two sided) rate of change. The Hessian is approximated by the summed outer products of the first derivatives. This is the BHHH estimator. Thus, it is always nonnegative definite. But, it can be singular in a particularly difficult problem. If you request Newton’s method, the Hessian is computed using two sided approximations to the second derivatives. For most problems, this will be sufficiently accurate. But, this matrix is not guaranteed to be positive definite, so you may get a diagnostic for a singular Hessian when using this method. If this happens, use one of the other methods.

E14.13 The NIST Accuracy Benchmarks

The National Institute of Standards and Technology (NIST) has published a suite of 27 benchmark problems for testing nonlinear regression programs. The NIST site contains statements of the problems and the datasets for their solutions, all of which can be downloaded from their website: http://www.itl.nist.gov/div898/strd/nls/nls_main.shtml. McCullough (1999) has extensively documented the performance of several econometric software packages, including *LIMDEP*, in solving the NIST benchmarks. Many of the NIST datasets and *LIMDEP* commands to carry out the tests are included with the *LIMDEP* program, and can be found in the last book of the Help file. Click Help, Help Topics, then double click The Nist Benchmarks book. The NIST datasets and files may also be found in the C:\LIMDEP10\Command Files folder created with program installation. All of the data and code needed to carry out the nonlinear least squares tests with *LIMDEP* are provided in both locations.

McCullough's (1999) survey suggests how the tests results can be summarized. As noted earlier, NIST provides two sets of starting values, a 'difficult' set (Start 1) and an easier set (Start 2) that is closer to the correct solution. McCullough provides a measure of the accuracy of the solutions using either set based on the number of correct digits produced, compared to the NIST certified solution. He also suggests that some other benchmarks be used to evaluate software, namely comparisons of solutions using program default settings vs. user modifications of the defaults, such as changes in the algorithms, provision of derivatives, or changes in the convergence tolerances.

The listing below shows some program code and a summary of the solutions to these problems as implemented in *LIMDEP*. A full listing of all the test runs would occupy an inordinate amount of space in this manual. The following summarizes the results:

- 26 of the 27 tests can be solved with *LIMDEP* to greater than nine digit accuracy. The exception is the Hahn1 problem examined in detail above, which is solved to 6.8 digit accuracy. 22 of the 27 solutions are obtained to greater than 10 digit accuracy.

- Of the 27 solutions obtained, all but five are reachable from Start 1.

- Using the basic defaults and Start 1, 19 of the 27 problems are solvable. The accuracy exceeds nine digits in 15 of these 19, exceeds six digits in three of the remaining four, and is at less than six digits in only one of these solutions.

In every case in which a solution is obtained, that solution is improved by including the analytic derivatives in the command. In general, this appears to be a good idea when it is feasible. This is clear with the Hahn1 problem shown earlier, for which the solution is not attainable without supplying the derivatives. The other modification which produces some benefit is to switch to the BFGS algorithm, then tighten the gradient convergence rule to, say, 1.D-12. Finally, it is occasionally necessary to increase the maximum number of iterations. Iterations in the hundreds are not uncommon.

E14.13.1 Setting up the NIST Benchmarks

To illustrate the test procedure, we show the full setup for the Hahn1 problem examined in Section E14.6. The data are presented there. We now **READ** them as variables *yh*1 and *xh*1. The parameters for the test are placed in a matrix arranged in the form

```
Number of observations                             Solutions
Number of parameters        Start 1     Start 2   Slopes    Std errors
       0               b1(0)_1     b1(0)_2  b1(*)     s1(*)
       0               b2(0)_1     b2(0)_2  b2(*)     s2(*)
       ...             ...         ...    ...      ...
Correct sum of squares   0           0        0         0

MATRIX;Hahn1=[
236,     10,          1,             1.0776351733E+00, 1.7070154742E-01/
  7,     -1,         -0.1,          -1.2269296921E-01, 1.2000289189E-02/
  0,      0.05,       0.005,         4.0863750610E-03, 2.2508314937E-04/
  0,     -0.00001,   -0.000001,     -1.4262662514E-06, 2.7578037666E-07/
  0,     -0.05,      -0.005,        -5.7609940901E-03, 2.4712888219E-04/
  0,      0.001,      0.0001,        2.4053735503E-04, 1.0449373768E-05/
  0,     -0.000001,  -0.0000001,    -1.2314450199E-07, 1.3027335327E-08/
1.5324382854E+00, 0,  0,             0,                0               ]$
```

We use a procedure to prepare the data set and matrix elements for the execution:

```
PROC = Setup(Problem) $
CALC          ; ndata  = problem(1,1)
              ; nparm = problem(2,1)
              ; np1 = nparm+1$
MATRIX        ; start1 = Part(problem,1,nparm,2,2)
              ; start2 = Part(problem,1,nparm,3,3)
              ; trueb  = Part(problem,1,nparm,4,4)
              ; trues  = Part(problem,1,nparm,5,5)
              ; trueee= Part(problem,np1,np1,1,1) $
CALC          ; truess = trueee $
SAMPLE        ; 1 - nparm $
CREATE        ; truebeta  = trueb $
CREATE        ; truese  = trues $
SAMPLE        ; 1 - ndata $
ENDPROC
```

The procedure sets up the specific NIST nonlinear least squares problems. It sets the sample size and problem size and extracts from the setup matrix the two sets of starting values into matrices. The two sets of 'true' values are also placed in matrices, and copied into variables in preparation for the LRE (log relative error) score routine that is executed after estimation. The procedure must be executed with

```
EXEC          ; Proc = Setup(problem name) $
```

before the **NLSQ** commands can be carried out.

The next step is to compute the nonlinear least squares estimates, first using the program default values and Start 1, second using whatever methods are available – what McCullough labels an ‘all out assault’ on the solution.

Step 1. Set up the problem:

```
EXECUTE     ; Proc = Setup(Hahn1) $
TITLE       ; Nonlinear Least Squares Estimation. Data set = Hahn1 $
CREATE      ; xh12 = xh1*xh1 ; xh13 = xh12*xh1$
```

Step 2. Solution attempt using program defaults:

```
NLSQ        ; Lhs = yh1
            ; Fcn = (b1+b2*xh1+b3*xh12+b4*xh13) /
                    (1+b5*xh1+b6*xh12+b7*xh13)
            ; Labels  = b1,b2,b3,b4,b5,b6,b7
            ; Output= 0 ; Dfc ; Start = Start 1 $
```

After fitting the model, we use another procedure to ‘score’ the solution in terms of its agreement with the NIST certified solution:

```
Maximum iterations exceeded
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=YH1       Mean                    =       14.21530
              Standard deviation      =        5.76869
              Number of observs.      =            236
Model size    Parameters              =              7
              Degrees of freedom      =            229
Residuals     Sum of squares          =        7.91052
              Standard error of e     =         .18586
Fit           R-squared               =         .99899
              Adjusted R-squared      =         .99896
Model test    F[  6,    229] (prob) = 37693.1(.0000)
Diagnostic    Log likelihood          =       65.81589
              Restricted(b=0)         =     -747.94531
              Chi-sq [   6]  (prob) =1627.5(  .0000)
Info criter. Akaike Info. Criter. =       -3.33632
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
UserFunc|  Coefficient       Error        z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
      B1|    10.2621***      .39880    25.73  .0000      9.4805   11.0438
      B2|    -.91579***      .02468   -37.10  .0000     -.96416   -.86741
      B3|     .02122***      .00041    52.11  .0000      .02042    .02202
      B4|-.19559D-04***   .1436D-06  -136.22  .0000 -.19841D-04  -.19278D-04
      B5|     .00575***      .00116     4.96  .0000      .00348    .00803
      B6|     .00101***   .2373D-04    42.65  .0000      .00097    .00106
      B7|-.10030D-05***   .8758D-10 ********  .0000 -.10032D-05  -.10028D-05
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This routine evaluates the quality of the solution to the **NLSQ** routine for each of the NIST problems. The computation is documented in McCullough (1999). It effectively measures the number of correct digits in the solution. The certified NIST values are given with 11 digits, so that is the maximum value. LRE scores are computed for coefficient vectors and for standard errors. The overall score for a problem is then the minimum value in the vector of LREs. No value is computed for the sum of squares, since it will be redundant. A correct solution for the coefficient vector implies a correct solution for the sum of squares, and vice versa (unique solution at the true minimum).

```
PROC  = Score $
MATRIX      ; betahat1 = b ; sehat1 = Diag(varb)
            ; sehat1 = Esqr(sehat1) ; sehat1 = Vecd(sehat1) $
SAMPLE      ; 1 - nparm $
CREATE      ; betahat = betahat1 $
CREATE      ; sehat = sehat1 $
CREATE      ; If((betahat - truebeta)#0)
              lrebeta = -Log(Abs(betahat - truebeta)/Abs(truebeta))/Log(10)
            ; (Else) lrebeta = 11.0$
CREATE      ; If(lrebeta > 11) lrebeta = 11 ; If(lrebeta < 0) lrebeta = 0 $
CREATE      ; If((sehat-truese)#0)
              lrese = -Log(Abs(sehat - truese)/ Abs(truese))/Log(10)
            ; (Else) lrere = 11.0$
CREATE      ; If(lrese > 11)lrese = 11 ; If(lrese < 0)lrese = 0 $
TYPE        ; Coefficient Estimates $
TYPE        ;    True          Estimated         LRE $
WRITE       ; truebeta, betahat, lrebeta ; Format = (3G20.11) $
CALC        ; List ; lreb = Min(lrebeta) $
TYPE        ; Estimated Standard Errors $
TYPE        ;    True          Estimated         LRE $
WRITE       ; truese, sehat, lrese ; Format = (3G20.11) $
CALC        ; List ; lres = Min(lrese)   $
CALC        ; If((sumsqdev - truess) #0 )
              lress = -Log(Abs(sumsqdev -truess) / Abs(truess))/Log(10)
            ; (Else) lress = 11.0$
CALC        ; If(lress > 11)lress = 11 ; If(lress < 0)lress = 0 $
SAMPLE      ; 1 $
CREATE      ; Cert_ss = truess ; actualss = sumsqdev ; scoress = lress $
TYPE        ; Sum of Squared Deviations $
TYPE        ;    True          Estimated         LRE $
WRITE       ; cert_ss , actualss , scoress ; Format = (3G20.11) $
SAMPLE      ; 1 - ndata $
ENDPROC
```

After estimation, the routine is invoked with

```
EXECUTE     ; Proc = Score $
```

It displays the results below the model output. Note that **SCORE** is wired to **SETUP**, and will only operate properly with respect to the immediately preceding **SETUP** and **NLSQ** commands.

This procedure requires only **EXECUTE ; Proc = SCORE $** to obtain the results. As noted earlier, this default solution ends up nowhere near the correct solution. The **SCORE** routine provides the following results: Note that the LRE score is essentially 0. This is the number of correct digits, with 11.0 being a perfect score.

```
Coefficient Estimates
      True              Estimated            LRE
    1.0776351733        10.262109935        .00000000000
   -.12269296921       -.91578640611        .00000000000
    .40863750610E-02    .21219532055E-01    .00000000000
   -.14262662514E-05   -.19559299652E-04    .00000000000
   -.57609940901E-02    .57544847240E-02    .00000000000
    .24053735503E-03    .10120072962E-02    .00000000000
   -.12314450199E-06   -.10030135584E-05    .00000000000
    LREB    =  .00000000000000000D+00
 Estimated Standard Errors
      True              Estimated            LRE
    .17070154742        .39880401139        .00000000000
    .12000289189E-01    .24682520390E-01    .00000000000
    .22508314937E-03    .40716767594E-03    .92069941905E-01
    .27578037666E-06    .14358433722E-06    .31934491643
    .24712888219E-03    .11599360500E-02    .00000000000
    .10449373768E-04    .23726005203E-04    .00000000000
    .13027335327E-07    .87579253224E-10    .29295025797E-02
    LRES    =  .00000000000000000D+00
 Sum of Squared Deviations
      True              Estimated            LRE
    1.5324382854        7.9105155265        .00000000000
```

We then try to solve the problem using some options to improve the search, and reevaluate the score.

```
NLSQ    ; Lhs = yh1
        ; Fcn = top  = (b1+b2*xh1+b3*xh12+b4*xh13)  |
                bot  = (1+b5*xh1+b6*xh12+b7*xh13)   |
                q    = 1/bot                        |
                _b1 = q                             |
                _b2 = xh1*q                         |
                _b3 = xh12*q                        |
                _b4 = xh13*q                        |
                _b5 = -top * q * q * xh1            |
                _b6 = _b5 * xh1                     |
                _b7 = _b6 * xh1                     |
                top * q
        ; Labels  = b1,b2,b3,b4,b5,b6,b7
        ; Output= 0 ; Alg = bfgs ; Dfc ; Start = Start 1 $
```

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=YH1       Mean                    =      14.21530
              Standard deviation      =       5.76869
              Number of observs.      =           236
Model size    Parameters              =             7
              Degrees of freedom      =           229
Residuals     Sum of squares          =       1.53244
              Standard error of e     =        .08180
Fit           R-squared               =        .99980
              Adjusted R-squared      =        .99980
Model test    F[  6,    229] (prob) =194732.2(.0000)
Diagnostic    Log likelihood          =     259.49316
              Restricted(b=0)         =    -747.94531
              Chi-sq [  6]  (prob) =2014.9(  .0000)
Info criter. Akaike Info. Criter. =       -4.97765
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|    1.07764***       .17070     6.31  .0000       .74307   1.41220
      B2|    -.12269***       .01200   -10.22  .0000      -.14621   -.09917
      B3|     .00409***       .00023    18.15  .0000       .00365    .00453
      B4|-.14263D-05***   .2758D-06    -5.17  .0000 -.19668D-05  -.88575D-06
      B5|    -.00576***       .00025   -23.31  .0000      -.00625   -.00528
      B6|     .00024***   .1045D-04    23.02  .0000       .00022    .00026
      B7|-.12314D-06***   .1303D-07    -9.45  .0000 -.14868D-06  -.97611D-07
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Coefficient Estimates
      True                  Estimated             LRE
    1.0776351733          1.0776360397        6.0947662175
   -.12269296921         -.12269298842        6.8052541456
    .40863750610E-02      .40863750498E-02    8.5617610439
   -.14262662514E-05     -.14262669557E-05    6.3064221751
   -.57609940901E-02     -.57609949011E-02    6.8514956358
    .24053735503E-03      .24053734756E-03    7.5081180728
   -.12314450199E-06     -.12314452770E-06    6.6802617659
[CALC] LREB     =      6.0947662
 Estimated Standard Errors
      True                  Estimated             LRE
    .17070154742          .17070153182        7.0389926603
    .12000289189E-01      .12000287674E-01    6.8987668645
    .22508314937E-03      .22508311606E-03    6.8297481252
    .27578037666E-06      .27578034073E-06    6.8851424753
    .24712888219E-03      .24712885330E-03    6.9322284075
    .10449373768E-04      .10449372086E-04    6.7933913363
    .13027335327E-07      .13027333567E-07    6.8694475538
[CALC] LRES     =      6.7933913
 Sum of Squared Deviations
      True                  Estimated             LRE
    1.5324382854          1.5324382854        10.628709885
```

This produces much greater agreement with the certified solutions.

E14.13.2 Application – Dan Wood

The Dan Wood problem is much easier to solve than Hahn1, as the following shows.

```
DAN WOOD
Model:          Miscellaneous Class
                2 Parameters (b1 and b2)
                y  = b1*x**b2  +  e
                Lower Level of Difficulty
                                    ----[ NIST Certified Solutions]---
        Start 1     Start 2            Parameter     Standard Deviation
  b1 =   1           0.7            7.6886226176E-01  1.8281973860E-02
  b2 =   5           4              3.8604055871E+00  5.1726610913E-02
Residual Sum of Squares:                     4.3173084083E-03
Residual Standard Deviation:                 3.2853114039E-02
Degrees of Freedom:                                 4
Number of Observations:                             6
Data:  y              x
      2.138         1.309
      3.421         1.471
      3.597         1.49
      4.34          1.565
      4.882         1.611
      5.66          1.68
```

```
MATRIX      ; DanWood = [
              6,  1,  0.7,   7.6886226176E-01, 1.8281973860E-02/
              2,  5,  4,     3.8604055871E+00, 5.1726610913E-02/
                             4.3173084083E-03, 0, 0, 0, 0] $
EXECUTE     ; Proc = Setup(DanWood) $
TITLE       ; Nonlinear Least Squares Estimation. Data set = DanWood $
NLSQ        ; Lhs= ydw
            ; Fcn= b1 * xdw^b2
            ; Labels = b1,b2
            ; Dfc ; Start = Start 1 ; Tln = 1.d-20 $
```

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=YDW       Mean                     =       4.00633
              Standard deviation       =       1.23398
              Number of observs.       =             6
Model size    Parameters               =             2
              Degrees of freedom       =             4
Residuals     Sum of squares           =    .431731E-02
              Standard error of e      =        .03285
Fit           R-squared                =        .99943
              Adjusted R-squared       =        .99929
Model test    F[  1,     4] (prob) =  7050.0(.0000)
Diagnostic    Log likelihood           =      13.19702
              Restricted(b=0)          =      -9.22815
              Chi-sq [  1]  (prob) =  44.9(  .0000)
Info criter. Akaike Info. Criter. =      -6.57022
Not using OLS or no constant. Rsqrd & F may be < 0
```

```
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
      B1|     .76886***      .01828    42.06  .0000      .73303    .80469
      B2|    3.86041***      .05173    74.63  .0000     3.75902   3.96179
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
Coefficient Estimates
      True                 Estimated             LRE
    .76886226176         .76886226185        9.9426221544
    3.8604055871         3.8604055868        10.171162836
[CALC] LREB    =      9.9426222
 Estimated Standard Errors
      True                 Estimated             LRE
    .18281973860E-01     .18281973717E-01    8.1071779144
    .51726610913E-01     .51726610215E-01    7.8695705878
[CALC] LRES    =      7.8695706
 Sum of Squared Deviations
      True                 Estimated             LRE
    .43173084083E-02     .43173084083E-02    11.000000000
```

```
NLSQ       ; Lhs = ydw
           ; Fcn = _b1 = xdw^b2                  |
                   _b2 = b1 * _b1 * Log(xdw)   |
                   b1 * xdw^b2
           ; Labels = b1,b2
           ; Dfc ; Start = Start 1 ; Tln = 1.d-20 $
```

```
----------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=YDW      Mean                 =        4.00633
             Standard deviation   =        1.23398
             Number of observs.   =              6
Model size   Parameters           =              2
             Degrees of freedom   =              4
Residuals    Sum of squares       =    .431731E-02
             Standard error of e  =         .03285
Fit          R-squared            =         .99943
             Adjusted R-squared   =         .99929
Model test   F[  1,      4] (prob) =  7050.0(.0000)
Diagnostic   Log likelihood       =       13.19702
             Restricted(b=0)      =       -9.22815
             Chi-sq [  1]  (prob) =  44.9(  .0000)
Info criter. Akaike Info. Criter. =       -6.57022
Not using OLS or no constant. Rsqrd & F may be < 0
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
      B1|     .76886***      .01828    42.06  .0000      .73303    .80469
      B2|    3.86041***      .05173    74.63  .0000     3.75902   3.96179
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

```
Coefficient Estimates
      True              Estimated             LRE
    .76886226176        .76886226176        11.000000000
    3.8604055871        3.8604055871        11.000000000
[CALC] LREB    =     11.0000000
 Estimated Standard Errors
      True              Estimated             LRE
    .18281973860E-01    .18281973860E-01    11.000000000
    .51726610913E-01    .51726610913E-01    11.000000000
[CALC] LRES    =     11.0000000
 Sum of Squared Deviations
      True              Estimated             LRE
    .43173084083E-02    .43173084083E-02    11.000000000
```

E15: Linear Models for Time Series/Cross Section Data

E15.1 Introduction

The models described in this chapter are of a type based on balanced panels in which the regression specification accommodates the natural grouping in the data in a more structured (and less flexible) fashion than the panel data models to follow in Chapters E16-E19. They are accessed with the command

TSCS ; Lhs = dependent variable
; Rhs = independent variables
; Pds = number of observations per group $

for the first case, in which the panel structure of the data is built into the disturbance covariance matrix, and

REGRESS ; Lhs = dependent variable
; Rhs = independent variables
; Pds = number of observations per group
; RCM $

for the second case in which the panel data aspect of the model appears in variation of the regression parameters.

The essential structure of the first form of the model is

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

The *groupwise covariance structures* model considered in this chapter is

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, \text{Cov}[\varepsilon_{it},\varepsilon_{js}] = \sigma_{ij}\times\mathbf{1}(t = s).$$

In this model, the regression function is assumed to be the same for all groups, and the structure of the model is built into the pattern of heteroscedasticity and contemporaneous correlation across groups. This model also allows for first order autocorrelation of the form Corr$[\varepsilon_{it},\varepsilon_{is}] = \rho_i^{|t-s|}$. The second form of the model is

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it}$$

where the variation in the parameter vector is built on a random parameter structure,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{w}_i, \mathbf{w}_i \sim f(\mathbf{0},\boldsymbol{\Gamma}).$$

E15.2 Panel Data Arrangement and Setup

Your data for this model are assumed to consist of variables:

$$y_{it}, x_{1it}, x_{2it}, ..., x_{Kit}, I_{it}, \; i=1,...,N, t = 1,...,T,$$

y_{it} = dependent variable,
$\mathbf{x}_{it}$ = set of independent variables,
K = number of regressors, including *one*,
N = number of groups,
T = fixed number of observations in group 'i.'

The data set for all panel data models in *LIMDEP* will normally consist of multiple observations, denoted $t = 1,...,T_i$, on each of $i = 1,...,N$ observation units, or 'groups.' A typical data set would include observations on several persons or countries each observed at several points in time, T_i, for each country. In the following, we use 't' to symbolize 'time' purely for convenience. The panel could consist of N cross sections observed at different locations or N time series drawn at different times, or, most commonly, a cross section of N time series, each of length T_i. The estimation routines are structured to accommodate large values of N, such as in the national longitudinal data sets, with T_i being as large or small as dictated by the study but not directly relevant to the internal capacity of the estimator. Data for the panel data estimators in *LIMDEP* are assumed to be arranged contiguously in the data set. Logically, you will have

$$Nobs = \sum_{i=1}^{N} T_i$$

observations on your independent variables, arranged in a data matrix

$$\mathbf{X} = \begin{bmatrix} T_1 \text{ observations for group 1} \\ T_2 \text{ observations for group 2} \\ \cdots \\ T_N \text{ observations for group } N \end{bmatrix}$$

and likewise for the data on $\mathbf{y}$, the dependent variable. When you first read the data into your program, you should treat them as a cross section with *Nobs* observations. The partitioning of the data for panel data estimators is done at estimation time. Chapter R5 contains further details on how to set up and use panel data sets.

NOTE: The estimators described in this chapter require a *balanced* panel. T_i must be the same for all i. These are the only panel data models in *LIMDEP* that have this restriction.

E15.3 Groupwise Heteroscedasticity, Correlation and Autocorrelation

This chapter describes estimation of a form of panel data model in which data are (typically) observed for a relatively large number of periods and for a relatively small number of cross sectional units – the reverse of the more familiar panel arrangement. Typical applications involve cross country studies, such as the 30 OECD countries, observed for a relatively large number of years. The model is

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ i = 1,...,N,\ t = 1,...,T.$$

The subscript 'i' indexes groups, 't' indexes periods. The coefficient vector is assumed to be constant over time and for all groups. The model allows for:

- groupwise heteroscedasticity, $E[\varepsilon_{it}^2] = \sigma_{ii}$,
- cross group correlation, $\text{Cov}[\varepsilon_{it},\varepsilon_{jt}] = \sigma_{ij}$,
- within group autocorrelation, $\varepsilon_{it} = \rho_i\varepsilon_{i,t-1} + u_{it}$.

For the nonautocorrelated models, the estimator may be two step FGLS or iterated FGLS which produces a maximum likelihood estimator. For the models with autocorrelation, the estimator may be three step GLS or iterated GLS, which though convergent, does not produce the MLE because of the Jacobian term.

The number of cross sectional units, N, is limited to 100. The number of periods per group must be fixed at some T. The full sample is limited to 200,000 observations for this estimator.

The TSCS model formulation provides three forms of the regression, labeled

S0 = homoscedastic and uncorrelated across groups (linear regression),
S1 = groupwise heteroscedastic,
S2 = groupwise heteroscedastic and correlated across groups.

The basic model command requests estimation and display of all three forms of the model. There are also three forms of the autocorrelation model:

R0 = no autocorrelation,
R1 = common autocorrelation coefficient, ρ,
R2 = group specific autocorrelation coefficient, ρ_i.

With no further modification, this creates nine different variants of the model, S0,R0, S0,R1, etc. Results presented with this model will include a full set of results for all models, the three base cases, or the nine permutations if you specify **; AR1**. You may limit estimation to one of the specific models with the **; Model = Ss,Rd** specification in the command, for example, **; Model = S1,R1**. All forms up to this model in the order, S0,R0, S1,R0, S2,R0, S0,R1, ... are estimated, but only the one you request is actually reported. Estimation stops at that point, so any saved results are based on the model you specify. Otherwise, the saved results are based on S2,R0 if you do not specify **; AR1** or S2,R2 if you do.

E15.3.1 Command and Options

The basic command for the 'Time Series/Cross Section' estimator is

TSCS ; Lhs = ...
; Rhs = ...
; Pds = number of periods $

You may use **REGRESS ; TSCS** instead of **TSCS** if you prefer. These are synonyms. Although this is a panel data estimator, its construction differs slightly from the ones in the preceding sections.

- The panel must be balanced. The construction of the disturbance covariance matrix, **Σ** requires this.
- There is no stratification indicator. The data set consists of *N* blocks of *T* observations.

Options are

; AR1	to request the models of autocorrelation
; MLE	to request the iterative estimators (also for AR1)
; Model = S0,R0 or **S1,R0**, etc.	

The **; Model = type** specification allows you to stop estimation at, and save the results for a particular specification rather than estimating all forms of the model and saving the last one estimated. The standard options available for **TSCS** also include:

; Wts =	an optional weighting variable
; Res = name	to retain residuals
; List	to display fitted values, residuals, and forecast intervals
; Keep = name	to retain predictions
; Output = 4	to list various covariance matrices
; Labels =	a list of names for the groups

Labels are used to label the rows and columns of residual covariance and correlation matrices. They can also be used in the next specification, which is used to place some rows and columns of zeros in the disturbance covariance matrix;

; Group = the list labels or numbers of groups that are freely correlated.

For the **; Group** option, any group not included in the list is assumed to be uncorrelated with all groups in the list as well as all other groups not in the list. If you have not used **; Labels**, then use simple group numbers. If you have used labels, then the list should use the labels you supply. For example: In the Grunfeld data, five of the ten groups (firms), GM, GE, Chrysler, U.S. Steel, and Westinghouse. The following two specifications would be equivalent:

```
TSCS        ; Lhs = i
            ; Rhs = one,f,c
            ; Pds = 20
            ; Group = 1,3,4
            ; Output = 4 $
```

and

```
TSCS        ; Lhs =  i
            ; Rhs = one,f,c
            ; Pds = 20
            ; Labels = gm,ge,chrysler,us_steel,westnghs
            ; Group = gm,chrysler,us_steel
            ; Output = 4 $
```

The correlation matrix reported would be as follows for the second form – the first would use simple firm numbers 1-5 instead of the names:

```
Matrix r(ei,ej) has  5 rows and  5 columns.
          GM                   GE         CHRYSLER     US_STEEL     WESTNGHS
        +---------------------------------------------------------------------
GM      |    1.0000    .0000000D+00    -.3576       -.4907    .0000000D+00
GE      | .0000000D+00    1.0000    .0000000D+00 .0000000D+00 .0000000D+00
CHRYSLER|    -.3576    .0000000D+00    1.0000        .8837    .0000000D+00
US_STEEL|    -.4907    .0000000D+00     .8837       1.0000    .0000000D+00
WESTNGHS| .0000000D+00 .0000000D+00 .0000000D+00 .0000000D+00    1.0000
```

NOTE: The zero correlation assumption is forced on the estimator, so GLS is done with the zeros specified in the covariance matrix.

E15.3.2 Results

The results from this model are extensive, as illustrated in the next section. Results which are kept for later use, as well as the residuals and predictions are based on the last model estimated, which will be S2,R0 or S2,R2 or on the model indicated in the **; Model = type** specification. These are:

Matrices: *b* = coefficient vector
varb = estimated asymptotic covariance matrix for **B**
sigma = $N \times N$ covariance matrix of *nonautocorrelated* disturbances
tscs_rho = $N \times 1$ vector of estimated autocorrelation coefficients.
The values of ρ_i in *tscs_rho* are always estimated, even without **; AR1**.

Scalars: *rho* = average autocorrelation for the groups (see below),
kreg = number of regressors
nreg = number of periods (not total number of observations)

Last Model: *b_variable* labels to use for Wald tests

Last Function: Conditional mean = **b′x**

In the general form for the TSCS model, the appropriate asymptotic covariance matrix for the ordinary least squares estimator would be

$$\text{Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where, since this is OLS, $\mathbf{X}$ would be the stack of $\mathbf{X}$ and $\mathbf{\Omega}$ would be the block matrix, with *ij*th block equal to $\sigma_{ij}\mathbf{I}$. Expanding this gives

$$\text{Var}[\mathbf{b}] = [\Sigma_i\mathbf{X}_i'\mathbf{X}_i]^{-1} \times [\Sigma_i\Sigma_j\, \sigma_{ij}\, \mathbf{X}_i'\mathbf{X}_j] \times [\Sigma_i\mathbf{X}_i'\mathbf{X}_i]^{-1}$$

A consistent estimator of this is easily obtained by just using $s_{ij} = \mathbf{e}_i'\mathbf{e}_j/T$ for σ_{ij}. By default, the reported covariance matrix for the first (S0,R0) OLS estimates in this model is just $s^2(\mathbf{X}'\mathbf{X})^{-1}$. You can request the preceding alternative estimator by adding

; PCSE (panel corrected standard errors)

NOTE: This correction (which originates with Beck and Katz (1995)) only applies to the standard errors computed for the OLS estimates (form S0,R0).

The panel corrected standard errors estimator is a special case. Panel corrected standard errors (PCSE) in TSCS allows more than 100 groups; there is no limit on the number of groups; the limit is only 200,000 observations in total. The reason is that it is not necessary to compute the $N\times N$ matrix $\mathbf{\Sigma}$ for this calculation. The example below for the Grunfeld data (Greene, 2011, Table F10.4) illustrates the difference.

TSCS ; Lhs = i ; Rhs = one,f,c ; Pds = 20 ; Model = S0,R0 $

```
----------------------------------------------------------------------------
Groupwise Regression Models
Estimator =                    2 Step GLS
Homoscedastic Regression              (S0)
Nonautocorrelated disturbances        (R0)
Pooled OLS residual variance (SS/nT)   8779.2524
Test statistics for homoscedasticity:
Deg.Fr. =    9 C*(.95) =  16.92 C*(.99) =  21.67
Lagrange multiplier statistic      =    209.7588
Wald                statistic      =  17051.7892
Likelihood ratio    statistic      =    268.5190
Log-likelihood function =          -1191.802360
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|   -42.7144***     9.44007    -4.52  .0000    -61.2166   -24.2122
       F|     .11556***      .00579    19.95  .0000      .10421     .12691
       C|     .23068***      .02528     9.12  .0000      .18112     .28023
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

TSCS ; Lhs = i ; Rhs = one,f,c ; Pds = 20 ; Model = S0,R0 ; PCSE $

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=ONE      Mean                 =      145.95825
             Standard deviation   =      216.87530
             No. of observations  =            200  Degrees of freedom
Regression   Sum of Squares       =    .760409E+07           2
Residual     Sum of Squares       =    .175585E+07         197
Total        Sum of Squares       =    .935994E+07         199
             Standard error of e  =       94.40840
Fit          R-squared            =         .81241  R-bar squared =   .81050
Model test   F[  2,    197]       =      426.57573  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    -1191.80236  Akaike I.C.   =  9.11015
             Restricted (b=0)     =    -1359.15096  Bayes  I.C.   =  9.15962
             Chi squared [  2]    =      334.69719  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error         z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -42.7144***     6.78096     -6.30  .0000     -56.0048   -29.4239
       F|      .11556***     .00721     16.02  .0000       .10143     .12970
       C|      .23068***     .02789      8.27  .0000       .17602     .28533
--------+--------------------------------------------------------------------
```

If you include **; MLE** in your command, then the FGLS estimators will be iterated to convergence. This gives the maximum likelihood estimators for the R0 cases. When there is autocorrelation, this is only approximately MLE. The problem is the first observation and the Jacobian term in the log likelihood. In any event, with MLE in the command, the results will contain an additional table, like the one shown below. Note that the log likelihood for all nine cells is calculated correctly. However, the R1 and R2 parameter estimates are not the true MLEs, as noted above. Thus, one might want to be careful in using these for any kind of formal testing. Note, for example, that the value of logL shown for S2,R2 is less than that for S2,R1, whereas if the estimates were true MLEs, the reverse would be true.

Log-likelihood functions for estimated models

	R0 Log-L	R0 Parameters	R1 Log-L	R1 Parameters	R2 Log-L	R2 Parameters
S0	-1191.802	4	-1116.263	5	-1119.031	14
S1	-956.689	13	-889.176	14	-896.558	23
S2	-798.888	58	-782.462	59	-800.379	68

NOTE: In the GLS estimator for TSCS, it is necessary to invert the *N*x*N* covariance matrix of the group specific residuals, **S**. This matrix has rank less than or equal to the minimum of *N* and *T*. Since **S** is a sum of *T* rank one matrices, its rank cannot exceed *T*. If *T* is less than *N*, **S** must be singular, and GLS cannot be computed. In words, if you have more periods than groups (e.g., countries), then GLS will not be possible. The condition $N > T$ is autodetected. When this condition is detected, a long warning is given, then the routine switches to PCSE with least squares, and halts the estimation. The following illustrates. The Grunfeld data set we are using contains ten firms. In the regression below, we use all $N = 10$ firms, and the first $T = 8$ observations to trigger the warning. The ‘correction’ noted in the warning is the PCSE estimator.

```
+-----------------------------------------------------------+
| You have T =   8 periods and N =    10 groups. When N > T |
| the NxN residual covariance matrix EtE/T has rank   T < N |
| and cannot be inverted. The full GLS estimator cannot be  |
| computed. The results below include only OLS with a cor-  |
| rection to the estimated covariance matrix.               |
+-----------------------------------------------------------+

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=ONE      Mean                 =      103.82050
             Standard deviation   =      142.24309
             No. of observations  =             80  Degrees of freedom
Regression   Sum of Squares       =    .119178E+07          2
Residual     Sum of Squares       =        406639.         77
Total        Sum of Squares       =    .159841E+07         79
             Standard error of e  =       72.67064
Fit          R-squared            =         .74560  R-bar squared =   .73899
Model test   F[  2,    77]        =      112.83574  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     -454.86123  Akaike I.C.   =  8.60865
             Restricted (b=0)     =     -509.61493  Bayes  I.C.   =  8.69798
             Chi squared [  2]    =      109.50740  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -13.4943**      6.11660    -2.21  .0274    -25.4826   -1.5060
       F|     .09100***      .00651    13.98  .0000      .07825    .10376
       C|     .18164***      .05254     3.46  .0005      .07866    .28462
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E15.3.3 Application

We continue the application developed in the previous sections. The data and model are the same. The command, which requests most of the available output, is the first one below. In the second, we allow a different constant term for each firm (which makes this equivalent to a 'fixed effects' model). The results from the regressions are largely the same as those shown in the earlier examples, and are not shown. The **PLOT** command displays the residuals from the two regressions.

```
INCLUDE    ; New ; Firm <= 5 $
CLIST      ; firms = gm,ge,chrysler,us_steel,wstnghse $
CREATE     ; Expand(firm) = d1,d2,d3,d4,d5 $
TSCS       ; Lhs = i ; Rhs = one,f,c ; AR1
           ; Pds = 20 ; Output = 4 ; Res = e
           ; Labels = firms $
TSCS       ; Lhs = i ; Rhs = d1,d2,d3,d4,d5, f, c ; AR1
           ; Pds = 20 ; Labels = firms ; Res = e _i  $
PLOT       ; Rhs = e,e_i ; Spikes = 20.5,40.5,60.5,80.5 ; Bars = 0
           ; Title = Residuals for TSCS Regression
           ; Vaxis = Residual $
```

The following figure is a plot of the residuals from the two regressions. The figure is particularly revealing. Not only does it show quite clearly the groupwise heteroscedasticity induced by the model (same coefficient vector for all firms), but also the autocorrelation. It is also clear that assuming the same coefficient vector applies to all firms induces quite a serious misspecification for the third and fifth firms. Allowing the constant terms to differ by firm partly mitigates the effect, but even with this extension, this would still appear to be a badly specified model.

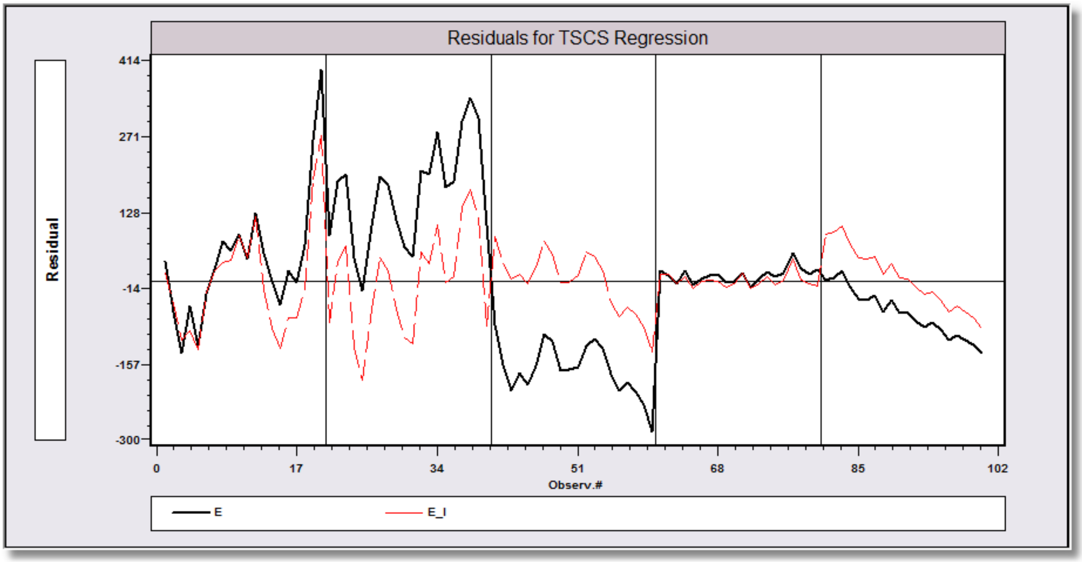


Figure E15.1 TSCS Residuals

E15.3.4 Technical Details

We use the same general procedure for all computations:

Let $\boldsymbol{\Sigma}$ = $N \times N$ period specific covariance matrix, $[\sigma_{ij}]$.

There are three cases:

S0: $\boldsymbol{\Sigma}$ = $\sigma^2\mathbf{I}$, homoscedastic regression,
S1: $\boldsymbol{\Sigma}$ = diag$[\sigma_{11},\sigma_{22},...,\sigma_{NN}]$, groupwise heteroscedastic,
S2: $\boldsymbol{\Sigma}$ = an $N \times N$ positive definite matrix, groupwise heteroscedastic and cross group correlated.

Let $\boldsymbol{\rho}$ = $N \times 1$ vector of group specific autocorrelation coefficients.

There are also three cases:

R0: $\boldsymbol{\rho}$ = $\mathbf{0}$, nonautocorrelated,
R1: $\boldsymbol{\rho}$ = $(\rho,\rho,...,\rho)$, common autocorrelation coefficient,
R2: $\boldsymbol{\rho}$ = $(\rho_1,\rho_2,...,\rho_N)$.

Thus, there are nine models when all three contemporaneous covariance specifications ($\mathbf{\Sigma}$) are crossed with the three autocorrelation specifications. Our approach is to compute all of them as restrictions on the model (S2,R2). The computations are as follows:

Step 1. Ordinary, pooled least squares.

Substep 1. Use OLS residuals to estimate $\boldsymbol{\rho}$. The procedures are as follows:

R0: Set $\rho_i = 0$.
R1: The common ρ is estimated as $(1/N)\Sigma_i r_i$ where r_i is the group specific residual autocorrelation.
R2: Use r_i as noted above.

Substep 2. With $\boldsymbol{\rho}$ in hand, first transform the data using the Prais-Winsten transformation.

Step 2. Compute OLS estimates using the data which have been transformed to remove the autocorrelation. Use the OLS residual sum of squares and cross products to compute

$$\mathbf{S} = [s_{ij}] = \mathbf{e}_i'\mathbf{e}_j/(T-l), \text{ where } l=0 \text{ for R0 and } l=1 \text{ for R1 and R2.}$$

Step 3. FGLS regression.

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{y} = [\Sigma_i\Sigma_j s^{ij}\mathbf{X}_i'\mathbf{X}_j]^{-1}[\Sigma_i\Sigma_j s^{ij}\mathbf{X}_i'\mathbf{y}_j],$$

and

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\beta}}] = [\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}.$$

The different specifications are estimated by restricting $\mathbf{\Sigma}$ and/or $\boldsymbol{\rho}$.

Three diagnostic statistics are computed for testing the hypothesis of the restrictions S0 on S1 and S1 on S2. For testing homoscedasticity as a restriction on S1,

$$\textbf{Wald} = (T/2)\Sigma_i\,[s^2/s_{ii} - 1]^2,$$
$$\textbf{LM} = (T/2)\Sigma_i\,[s_{ii}/s^2 - 1]^2, \text{ and}$$
$$\textbf{LR} = T(N\log s^2 - \Sigma_i \log s_{ii})$$

in which s^2 is the pooled OLS residual variance. All have limiting chi squared distributions with N-1 degrees of freedom under the hypothesis of homoscedasticity. The LR statistic is computed using estimates from S0 and S1, while both Wald and LM are based on S1. (LM should, in fact, be based on S0, so this is an approximation.) For testing groupwise heteroscedasticity as a restriction on S2, we compute

$$\textbf{LM} = T\Sigma_i\Sigma_{j<i}\,[s_{jj}^2/(s_{ii}s_{jj})] \text{ (the squared cross group correlation),}$$

and

$$\textbf{LR} = T(\Sigma_i \log s_{ii} - \log|\mathbf{S}|).$$

LM is computed using the final GLS estimates while LR is based on both S1 and S2.

No specific test is given for autocorrelation. One can test the significance of the estimated correlations, themselves, by referring $(T\text{-}1)r/(1 - r^2) \approx \chi^2[1]$ to the value 3.84, which is the 95% critical value from the chi squared distribution with one degree of freedom.

As shown in the example above, the TSCS estimator produces a large amount of output. With **; AR1**, there are nine regressions. In addition, several different forms of **Σ** are shown:

- For all estimators, the untransformed covariance matrix of the residuals for the model. For the three specifications, these are:

  S0: **Σ** = σ^2**I**. We display the estimate of σ^2.
  S1: **Σ** = a diagonal matrix. The diagonal matrix is displayed. Note that, in fact, $\mathbf{e}_{j\,i}'\mathbf{e}_j/T$ will not equal zero, but the model assumes σ_{ij} = zero.
  S2: **Σ** = a positive definite matrix. We display the full matrix. For this case, we also display **R** = the matrix of cross sectional correlations.

- For Models R1 and R2, the first display is the covariance matrix of the nonautocorrelated residuals. These are the residuals that result from the Cochrane-Orcutt transformation. Then, we also display the derived covariance matrix of the autocorrelated residuals. The computations are as follows: The disturbances are $\varepsilon_{it} = \rho_i\,\varepsilon_{it} + u_{it}$. The first matrix displayed is the estimate of **Σ** = Cov[u_{it},u_{jt}]. This is followed by the estimate of

  $$\mathbf{\Sigma}^{**} = \text{Cov}[\varepsilon_{it},\varepsilon_{jt}] = \sigma_{ij} / [(1 - \rho_i)(1 - \rho_j)].$$

  These matrices are normally not displayed with the standard output. To request them as part of the output, add

  ; Output = 4 to the **TSCS** command.

If you would prefer to see only a particular set of results for one of the nine models, use

; Model = S0,R0 or **S1,R0**, etc.,

for any of the nine forms of the model. Only that set of results will be displayed, and the final results saved as *b*, *varb*, etc. will be based on that model.

Two Important Technical Points

The covariance matrix, **S**, for Model S2 (the full model) is based on the residuals computed from the results of Model S1, the groupwise heteroscedastic regression. The standard textbook treatment of this model prescribes computing **S** using the results of OLS at the first step. The former is appropriate both under the null of the classical model and the alternative of the full model, so in terms of the asymptotic properties of the estimator, it makes no difference. But, the difference will be noticeable numerically in a finite sample. To use the textbook variant, add

; OLS

to the **TSCS** command. This requests the estimator of **Σ** based on the OLS residuals. There is no wisdom on which is a preferable estimator. In some limited experiments, we have found that the iterated FGLS estimator converged more readily using the default estimator rather than the OLS estimator, but we could find no theoretical basis on which to explain the finding.

The second point relates to the computation of **S** in the S2 formulation of the model. (This is one of our most frequently asked questions.) Let $\mathbf{e}_t$ denote the column vector of N residuals for all N groups at a particular time, t. For Model S2, **S** is computed using the formula

$$\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t{}'$$

This $N\times N$ matrix has rank less than or equal to the minimum of N and T – it is a sum of T rank 1 $N\times N$ matrices. This means that *if you have more countries than periods, it is not possible to invert* **S**, *so it is not possible to compute Model S2*. This is a general result having nothing to do with the software. The example shown earlier at the end of Section E15.3.2 shows the program reaction to finding that your model has this problem.

E15.4 Hildreth, Houck, and Swamy's Random Coefficients Model

The Hildreth/Houck/Swamy variant of the random coefficients model (*RCM*) is

$$\begin{aligned}
\mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i,\ i = 1,\ldots,N \text{ groups},\\
E[\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i] &= \mathbf{0},\ \text{Var}[\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i] = \sigma_i^2\mathbf{I},\\
\boldsymbol{\beta}_i &= \boldsymbol{\beta} + \mathbf{v}_i,\\
E[\mathbf{v}_i \mid \mathbf{X}_i] &= \mathbf{0},\ \text{Var}[\mathbf{v}_i \mid \mathbf{X}_i] = \boldsymbol{\Gamma}.
\end{aligned}$$

(See Swamy (1971, 1974), Hsiao (1986), and Hildreth and Houck (1968).) A linear regression model applies to each group. The coefficient vector within the group is a random draw from a distribution with overall mean, $\boldsymbol{\beta}$ which we seek to estimate. The reduced form of the model is

$$\begin{aligned}
\mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i + \mathbf{X}_i\mathbf{v}_i)\\
&= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i,
\end{aligned}$$

with $E[\mathbf{w}_i \mid \mathbf{X}_i] = \mathbf{0}$

and $\text{Var}[\mathbf{w}_i \mid \mathbf{X}_i] = \sigma^2\mathbf{I} + \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i{}' = \mathbf{A}_i.$

This is a groupwise heteroscedastic and autocorrelated (correlation across observations within groups) regression model.

E15.4.1 Command

This model is set up as a panel data regression model. The command is

```
REGRESS     ; Lhs = ... ; Rhs = ...
            ; Panel
            ; RCM
            ; Str = the stratification indicator
       or   ; Pds = group size $
```

In the example of the previous section, the stratification indicator would be *firm*. As part of the labeling of your output, your command may include

; Labels = a set of alphanumeric labels for up to 500 groups.

In this setting, there may be any number of groups, $i = 1,...,N$. (The limit of 20,000 in earlier versions of *LIMDEP* no longer applies.) The number of coefficients is limited to 150 as usual. The panel may be unbalanced, with the number of observations in group i equal to T_i. As before, there is no limit on T_i. Once again, the stratification variable need not be the set of consecutive integers. It can be any set of distinct values, so long as each i has a value. (See Chapter R5 for discussion of setting up panel data for estimation.)

The estimation results will include only the feasible GLS (FGLS) estimates. A deeper analysis involving display of results for each group, including the group specific prediction of $\boldsymbol{\beta}_i$, is requested with

; Output = 4

The **REGRESS** command builder dialog boxes will construct the instruction for the random coefficients model. The regression is set up as described in Section E16.2 for the panel data, linear regression model. On the Options page, first select Panel data model, then in the Model type window, select Random Coefficients. Next, click Settings to open the dialog box of options for the random coefficients model.

- Other options for the classical model apply as usual, but **; AR1** is not available.
- **; Res** and **; Keep** for residuals and predictions are based on the estimated overall mean,
- Values saved include matrices *b* and *varb* as usual. The variance weighted average of the OLS values is saved as a $K\times 1$ vector *beta_hat*. If the number of groups+1 times K is less than 50,000, then the individual predictions for $\boldsymbol{\beta}_i$ are saved as columns in the matrix *bt*.

The method of computing the individual predictions is given in Section E7.3.2.

The results saved are:

Scalars: *sumsqdev, ssqrd, s* are based on the sum of squares from $\hat{\boldsymbol{\beta}}$,

rsqrd

rho, logl		are returned as zero (not computed)
ybar, sy		are based on the full, pooled data set
kreg	=	the number of Rhs variables
nreg	=	the total number of observations, $\Sigma_i T_i$
degfrdm	=	*nreg* - *kreg*
ngroup	=	the number of groups
exitcode	=	zero unless data are collinear or a setup error occurs

Matrices:

b	=	$\hat{\boldsymbol{\beta}}$, the feasible GLS estimate of the mean $\boldsymbol{\beta}$
varb	=	the estimated asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}$
bt	=	$K\times N$ matrix whose *i*th column is the estimate of $\boldsymbol{\beta}_i$
gamma	=	the estimate of $\boldsymbol{\Gamma}$
beta_hat	=	the variance matrix weighted average of the least squares vectors $\mathbf{b}_i$. (Sets $\boldsymbol{\Gamma} = \mathbf{0}$ in $\mathbf{W}_i$ in the expression in Section E15.4.3.)

Last Model: *b_variables*

Last Function: Conditional mean function = **β'x**

You may specify that the coefficients on certain variables are not random. These coefficients will still be estimated by FGLS. But, under this specification, there are rows and columns of zeros in **Γ**, the covariance matrix for the slopes, for these coefficients. Use the specification,

; Rh2 = list of variables whose coefficients are not random.

You may also specify that there is only a single common σ^2 instead of group specific disturbance variances, σ_i^2. The estimate is the pooled OLS estimator, based on a single OLS vector of coefficients. A different estimate, s^2* can also be requested. This estimate is based on the sum of group specific sums of squares, based on group specific OLS coefficient vectors. This estimate will always be smaller than s^2 based on a pooled OLS estimate. Use

	; Alg = constant	to force a common σ^2
or	**; Alg = group**	for the second estimator

We note, this estimator requires that it be possible to compute a least squares estimator, $\mathbf{b}_i$, for each group. This limits the applicability of this estimator. The Grunfeld data used earlier, and below are a natural application. The data are described in Section E7.10.1.

E15.4.2 Application

Listed below are the results of applying the random coefficients program to the five firm Grunfeld data used for our earlier examples. The command requests the preliminary OLS results (**; All**) and the individual predictions of the group specific coefficient vectors (**; Output = 4**). After estimation, we compare the residuals from three sets of estimates. The thick line in the figure tracks the GLS residuals from the model with a single coefficient vector. The lighter dashed line shows the residuals from a least squares regression in which each firm is allowed to have a separate constant term. The improvement in the fit is obvious. The lightest dashed line shows the residuals from the firm specific estimates derived in the next section. As might be expected, these appear to produce the best fit of the three. The commands used in the analysis are as follows:

```
SAMPLE      ; 1-200 $
SETPANEL    ; Group = firm ; Pds = ti $
NAMELIST    ; x = one,f,c $
```

Compute the Hildreth and Houck random coefficients regression.

```
REGRESS     ; Lhs = i ; Rhs = x ; Panel ; RCM
            ; All ; Output = 4 ; Res = e $
CREATE      ; Expand(firm) = d1,d2,d3,d4,d5,d6,d7,d8,d9,d10 $
CREATE      ; ef = 0 $
```

Compute the least squares results with firm specific constant terms.

```
REGRESS     ; Lhs = i ; Rhs = d*,f,c ; Res = e_firm $
```

Extract the individual vectors in *bt* and create the residuals using these.

```
PROCEDURE $
INCLUDE     ; New ; firm = j $
MATRIX      ; bf = bt(1:3,j:j) $
CREATE      ; ef = i - x'bf $
ENDPROC     $
EXECUTE     ; j = 1,10 $
```

Plot the three sets of residuals in the same figure.

```
SAMPLE      ; 1-200 $
PLOT        ; Rhs = e,e_firm,ef
            ; Spikes = 40.4,80.5,120.5,160.5 ; Bars = 0
            ; Fill ; Symbols ; Endpoints = 0,200
            ; Title = Residuals by Firm: FGLS and Dummy Variables
            ; Yaxis = Residual $
```

```
+-------------------------------------------------------------------+
| Variable = ____________ Variable Groups     Max     Min    Average |
| TI          Group sizes  FIRM           10      20      20       20.0 |
+-------------------------------------------------------------------+

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                   =      145.95825
             Standard deviation     =      216.87530
             No. of observations    =            200  Degrees of freedom
Regression   Sum of Squares         =     .760409E+07           2
Residual     Sum of Squares         =     .175585E+07         197
Total        Sum of Squares         =     .935994E+07         199
             Standard error of e    =       94.40840
Fit          R-squared              =         .81241  R-bar squared =    .81050
Model test   F[  2,    197]         =      426.57573  Prob F > F*   =    .00000
Diagnostic   Log likelihood         =    -1191.80236  Akaike I.C.   =   9.11015
             Restricted (b=0)       =    -1359.15096  Bayes  I.C.   =   9.15962
             Chi squared [  2]      =      334.69719  Prob C2 > C2* =    .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -42.7144***     9.51168     -4.49  .0000    -61.3569   -24.0718
       F|     .11556***      .00584     19.80  .0000      .10412     .12700
       C|     .23068***      .02548      9.05  .0000      .18075     .28061
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Random Coefficients Model
Number of groups                  =        10
Full sample statistics based on GLS:
Mean of dependent variable        =  145.9582
Std. Dev. of dependent variable   =  216.8753
Residual standard deviation       =  105.5392
R squared                         =     .7656
Chi-squared for homogeneity test  =    901.43
Degrees of freedom                =        27
Probability value for chi-squared=    .000000
X means below are var. weighted OLS slopes.
Heterosc. e(i,t). s(i) based on b(i,ols)
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
CONSTANT|   -9.62929       17.03504     -.57  .5719   -43.01735  23.75878
       F|     .08459***      .01996     4.24  .0000      .04547    .12370
       C|     .19942***      .05265     3.79  .0002      .09622    .30262
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

--------------------------------------------------------------------------------
        | Estimate of the underlying distribution of
        | beta. Estimated mean is b(GLS). Estimated
        | covar. matrix is sample estimate of Gamma.
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
CONSTANT|   -9.62929       48.41739     -.20  .8424  -104.52563  85.26706
       F|     .08459         .05584     1.51  .1298     -.02486    .19403
       C|     .19942         .15647     1.27  .2025     -.10725    .50609
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

--------------------------------------------------------------------------------
Group specific coefficient estimates
Prediction for group  1   GROUP001
Number of Observations  =           20.0
Group Mean of LHS       =      608.02000
Group Std. Dev. of LHS  =      309.57463
Fit Measures for the Estimators
(When not OLS, Rsqrd = 1-ee/yy may be < 0!)
Estimator    Sum of Squares    R-squared
OLS            143205.877411      .921354
GLS            737640.023023      .594902
Prediction     150205.440285      .917510
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|   -55.4418       34.12527    -1.62  .1042   -122.3261   11.4425
       F|     .09815*        .05504     1.78  .0746     -.00974    .20603
       C|     .37225**       .15344     2.43  .0153      .07152    .67298
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
(Firms 2 - 9 omitted)
--------------------------------------------------------------------------------
Group specific coefficient estimates
Prediction for group 10   GROUP010
Number of Observations  =           20.0
Group Mean of LHS       =        3.08450
Group Std. Dev. of LHS  =        1.71866
Fit Measures for the Estimators
(When not OLS, Rsqrd = 1-ee/yy may be < 0!)
Estimator     Sum of Squares    R-squared
OLS                20.026732      .643158
GLS               656.047378   -10.689645
Prediction         20.790983      .629540
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|    -.18988       48.39062      .00  .9969   -95.03376  94.65399
       F|     .01396         .05169      .27  .7870     -.08735    .11528
       C|     .38426***      .14453     2.66  .0078      .10099    .66753
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                 =      145.95825
             Standard deviation   =      216.87530
             Number of observs.   =            200
Model size   Parameters           =             12
             Degrees of freedom   =            188
Residuals    Sum of squares       =        523478.
             Standard error of e  =       52.76797
Fit          R-squared            =         .94407
             Adjusted R-squared   =         .94080
Model test   F[ 11,    188] (prob) =   288.5(.0000)
Diagnostic   Log likelihood       =    -1070.78103
             Restricted(b=0)      =    -1359.15096
             Chi-sq [ 11]  (prob) = 576.7(  .0000)
Info criter. Akaike Info. Criter. =        7.98993
Not using OLS or no constant. Rsqrd & F may be < 0
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       I|  Coefficient       Error       t    |t|>T*         Interval
--------+-----------------------------------------------------------------------
      D1|   -70.2967       49.70796    -1.41  .1590   -167.7225   27.1291
      D2|    101.906***    24.93832     4.09  .0001      53.028   150.784
      D3|   -235.572***    24.43162    -9.64  .0000    -283.457  -187.687
      D4|   -27.8093**     14.07775    -1.98  .0497    -55.4012    -.2174
      D5|   -114.617***    14.16543    -8.09  .0000    -142.381   -86.853
      D6|   -23.1613*      12.66874    -1.83  .0691    -47.9916    1.6690
      D7|   -66.5535***    12.84297    -5.18  .0000    -91.7252  -41.3817
      D8|   -57.5457***    13.99315    -4.11  .0001    -84.9717  -30.1196
      D9|   -87.2223***    12.89189    -6.77  .0000   -112.4899  -61.9546
     D10|   -6.56784       11.82689     -.56  .5793   -29.74812  16.61244
       F|     .11012***      .01186     9.29  .0000      .08689    .13336
       C|     .31007***      .01735    17.87  .0000      .27605    .34408
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

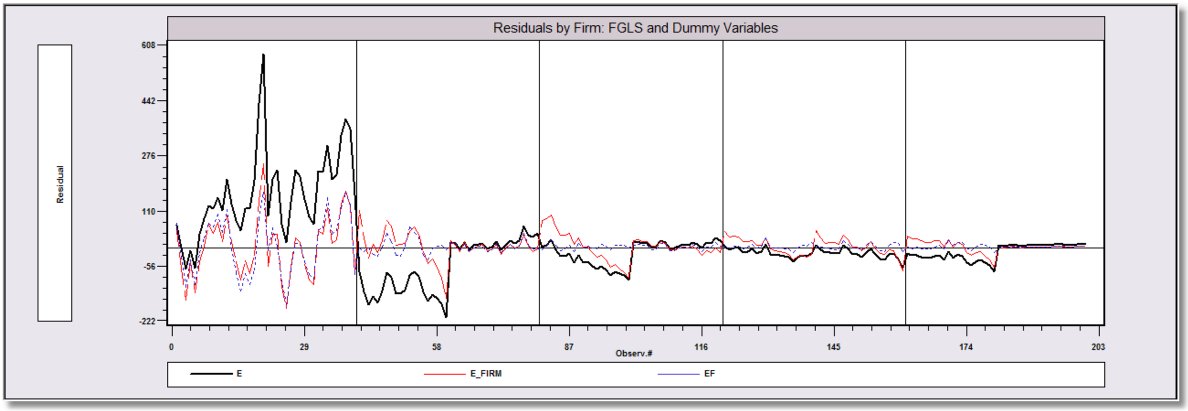


Figure E15.2 Residual Plot for Random Parameters Regression

E15.4.3 Technical Details for the Random Coefficients Estimator

Feasible Generalized Least Squares

The FGLS estimator in this model is computed as follows: Let

$$\boldsymbol{\Phi}_i = \sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}.$$

The FGLS estimator is

$$\hat{\boldsymbol{\beta}} = \Sigma_i \mathbf{W}_i \mathbf{b}_i,$$

where $\mathbf{b}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i$

and $\mathbf{W}_i = [\Sigma_i (\boldsymbol{\Gamma} + \boldsymbol{\Phi}_i)^{-1}]^{-1}(\boldsymbol{\Gamma} + \boldsymbol{\Phi}_i)^{-1}.$

Note that $\Sigma_i \mathbf{W}_i = \mathbf{I}$. Estimation is done in two steps, by estimating $\boldsymbol{\Gamma}$ first, then accumulating the matrix weighted average of the ordinary least squares coefficient vectors. The following technique is suggested (by Swamy (1974)). Let $\bar{\mathbf{b}} = (1/N)\Sigma_i \mathbf{b}_i$. Then,

$$\mathbf{V}_i = s_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1},$$

where $s_i^2 = \mathbf{e}_i'\mathbf{e}_i/(T_i - K).$

Estimate $\boldsymbol{\Gamma}$ with $\mathbf{G} = [1/(N\text{-}1)][\Sigma_i \mathbf{b}_i\mathbf{b}_i' - N\bar{\mathbf{b}}\,\bar{\mathbf{b}}'] - (1/N)\Sigma_i \mathbf{V}_i.$

Then, the remaining computations are straightforward. A problem may arise in that **G** may not be positive definite if the second matrix is too large. Several fixes have been suggested; the simplest is to omit the second matrix, which vanishes asymptotically anyway.

A chi squared test of the model against the alternative of the classical regression (no randomness of the coefficients) can be based on

$$\chi^2[(N\text{-}1)K] = \Sigma_i (\mathbf{b}_i - \mathbf{b}_*)'\mathbf{V}_i^{-1}(\mathbf{b}_i - \mathbf{b}_*),$$

where $\mathbf{b}_* = [\Sigma_i \mathbf{V}_i^{-1}]^{-1}\Sigma_i \mathbf{V}_i^{-1}\mathbf{b}_i.$

This is reported with the model output in the second frame of results. From the application above,

```
| Chi-squared for homogeneity test =    655.50   |
| Degrees of freedom               =        12   |
```

Predicting Group Specific Coefficient Vectors

The individual predictions of the group specific coefficient vectors are matrix weighted averages of the GLS estimator, $\hat{\boldsymbol{\beta}}$, and the group specific OLS estimates, $\mathbf{b}_i$,

$$\hat{\boldsymbol{\beta}}_i = \mathbf{Q}_i\hat{\boldsymbol{\beta}} + [\mathbf{I} - \mathbf{Q}_i]\mathbf{b}_i,$$

where $$\mathbf{Q}_i = [(1/s_i^2)\mathbf{X}_i'\mathbf{X}_i + \mathbf{G}^{-1}]^{-1}\mathbf{G}^{-1}.$$

(It can be shown that the weights in this average are proportional to the inverses of the asymptotic covariance matrices of the two parts. If there is sufficient space ($KN \leq 50{,}000$), these estimates are saved as the columns of the matrix *bt*. *LIMDEP* will also report full statistical results for the individual group predictions of $\boldsymbol{\beta}_i$ with standard errors. This produces a full set of output for each group in the sample, which can be substantial. Request this option with

; Output = 4

added to the **REGRESS ; ... ; RCM $** command. If you would like a separate report of the OLS results for the pooled sample, use

; All

E16: Linear Regression Models for Panel Data

E16.1 Introduction

Chapters E16-E23 document estimators for linear models using panel data. This chapter will detail some basic elements of the framework. The essential structure for most of the models is an 'effects' model,

$$y_{it} = \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

in which variation across groups (individuals) or time is captured in simple shifts of the regression function – i.e., changes in the intercepts. Chapter E17 describes the *fixed effects* (FE) model. The *random effects* (RE) models are detailed in Chapter E18. Several variations on this structure can be analyzed with this estimator, including both one and two factor models, models of autocorrelation, and simultaneous equations models. Chapter E18 also presents some major extensions including multifactor random effects models. More general forms of random parameter models are documented in Chapter E19. Chapters E20 and E21 show how to fit models with endogenous right hand side variables using two stage least squares in Chapter E22, and the Hausman and Taylor estimator for random effects and the Arellano, Bond and Bover estimator for dynamic panel data models in Chapter E23.

E16.2 Commands for Panel Data Regressions

The commands for estimation of these models are variants of the basic structure

```
SETPANEL    ; Group = group identifier variable
            ; Pds = variable to use for counts $
```

Then,

```
REGRESS     ; Lhs = y
            ; Rhs = x ...
            ; Panel
            ; ... other options $
```

The **SETPANEL** command is a global setting that needs only to be invoked once before your various panel data analyses. (See Chapter R5 for details.) Earlier versions of *LIMDEP* used one of

; Str = the name of a stratification variable

or **; Pds = specification of the number of periods, variable or fixed**

in each model command to specify the panel data structure. This construction may still be used in Version 10.

You may specify the **REGRESS** command using the command builder by selecting Model:Linear Models/Regression. The Lhs variable and the single Rhs variable are specified on the Main page.

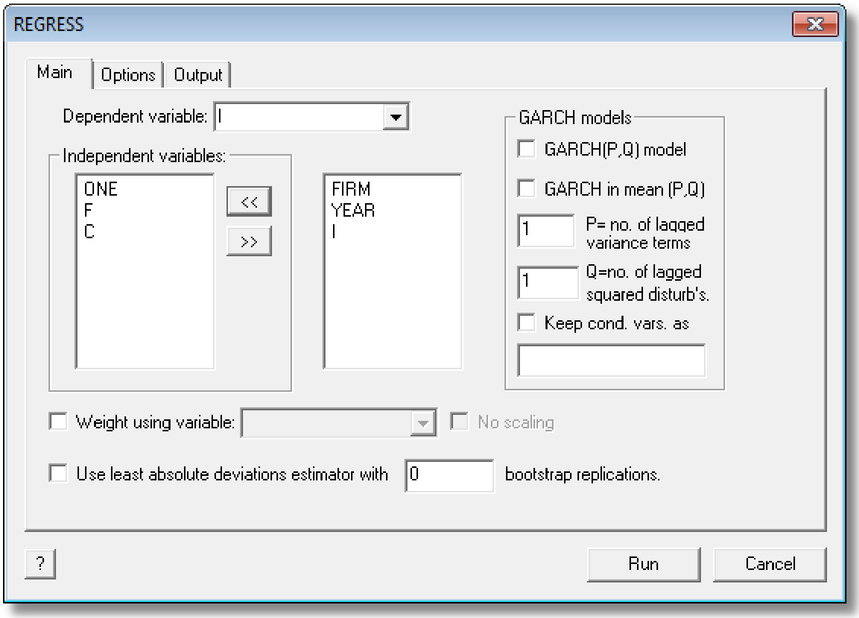


Figure E16.1 Main Page of Command Builder for REGRESS

The panel data model, and either a stratification variable or a fixed number of periods (only one would be used) are specified on the Options page.

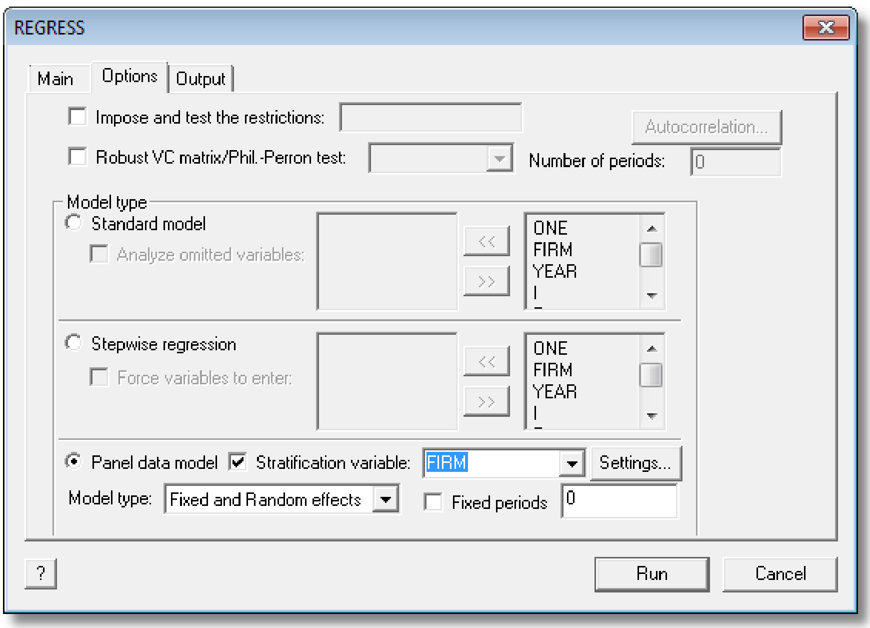


Figure E16.2 Options Page of Command Builder for REGRESS

HINT: If you have a fixed number of periods, be sure to click the Fixed periods checkbox before you enter the number of periods in the editing window. *LIMDEP* will allow you to enter the number of periods, but if you do not also click the checkbox, the panel data estimator will not be used.

E16.3 One Way Analysis of Variance

The simple one way analysis of variance for a variable is produced by the fixed effects regression model specified without covariates:

$$\begin{aligned} y_{it} &= \alpha_1 d_{1it} + \alpha_2 d_{2it} + \ldots + \varepsilon_{it} \\ &= \alpha_i + \varepsilon_{it}. \end{aligned}$$

The command for this computation is simply

REGRESS ; Lhs = the variable ; Rhs = one ; Panel $

If you have not used **SETPANEL** first, you can use

; ... stratification, either ; Str = variable or ; Pds = T $

E16.3.1 Computations and Saved Results

For the one way analysis of variance, results are based on the following computations:

N = total number of groups

T_i = number of observations in group i

$Nobs$ = $\sum_{i=1}^{N} T_i$ = total number of observations in the sample

$\bar{y}_i$ = sample mean of observations in group i

$\bar{y}$ = $\left[\sum_{i=1}^{N}\sum_{t=1}^{T_i} y_{it}\right] / Nobs$ = overall sample mean

$\sum_{i=1}^{N} T_i\left(\bar{y}_i - \bar{y}\right)^2$ = between groups sum of squares = SSB

$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(y_{it} - \bar{y}_i\right)^2$ = within groups sum of squares = SSW

$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(y_{it} - \bar{y}\right)^2$ = total sum of squares = SST = $SSB + SSW$

Sums of squares are computed in deviation form to achieve the maximum accuracy. Reported results are based on these statistics. Results saved for later use are:

Scalars:

ssqrd = regression variance = $SSW / (Nobs - N)$
rsqrd = proportion explained = SSB / SST
s = square root of *SSQRD*
sumsqdev = *SSB*
degfrdm = $Nobs - N$
ybar = $\bar{y}$
sy = square root of $SST / (Nobs - 1)$
kreg = N
nreg = $Nobs$

E16.3.2 Applications

We illustrate the computation with two of the NIST benchmarks for software accuracy, one each for low, medium, and high level of difficulty. See Section E2.12 for discussion of this suite of test problems. There are nine analyses of variance test problems. The problem statements given below are verbatim from the NIST website.

```
Dataset Name:    Atomic Weight of Silver   (agwt.dat)
Procedure:       Analysis of Variance
Reference:       Powell, L.J., Murphy, T.J. and Gramlich, J.W. (1982).
                 "The Absolute Isotopic Abundance & Atomic Weight
                 of a Reference Sample of Silver".
                 NBS Journal of Research, 87, pp. 9-19.
Data:            1 Factor
                 2 Treatments
                 24 Replicates/Cell
                 48 Observations
                 7 Constant Leading Digits
                 Average Level of Difficulty
                 Observed Data
Model:           3 Parameters (mu, tau_1, tau_2)
                 y_{ij} = mu + tau_i + epsilon_{ij}
Certified Values:
Source of                  Sums of               Mean
Variation          df      Squares             Squares             F Statistic
Between     Instrument         1     3.63834187500000E-09     3.63834187500000E-09
1.59467335677930E+01
Within Instrument  46 1.04951729166667E-08 2.28155932971014E-10
                   Certified R-Squared 2.57426544538321E-01
                   Certified Residual
                   Standard Deviation  1.51048314446410E-05
Data:  Instrument           AgWt
Read ; Nobs=48 ; Nvar=1 ; Names=y ; ByVariables $
107.8681568 107.8681465 107.8681572 107.8681785 107.8681446 107.8681903
107.8681526 107.8681494 107.8681616 107.8681587 107.8681519 107.8681486
107.8681419 107.8681569 107.8681508 107.8681672 107.8681385 107.8681518
107.8681662 107.8681424 107.8681360 107.8681333 107.8681610 107.8681477
107.8681079 107.8681344 107.8681513 107.8681197 107.8681604 107.8681385
107.8681642 107.8681365 107.8681151 107.8681082 107.8681517 107.8681448
107.8681198 107.8681482 107.8681334 107.8681609 107.8681101 107.8681512
107.8681469 107.8681360 107.8681254 107.8681261 107.8681450 107.8681368
Regress ; Lhs=y ; Rhs=one ; Pds = 24 ; Panel$
```

```
----------------------------------------------------------------------------
Analysis of Variance for         Y
Stratification Variable          _STRATUM
Total Sample Size                                  48      Group Sizes
Number of Groups                                    2      Max =    24
Number of groups with no data                       0      Min =    24
Overall Sample Mean                       107.8681451      Avg =  24.0
Total Sample Minimum                      107.8681079
Total Sample Maximum                      107.8681903
Sample Standard Deviation                    .0000173
Total Sample Variance                        .0000000
```

```
Source of Variation             Variation     Deg.Fr.      Mean Square
Between Groups             .3638341875D-08          1  .3638341875D-08
Within  Groups             .1049517292D-07         46  .2281559330D-09
Total                      .1413351479D-07         47  .3007130807D-09
Residual S.D.              .1510483144D-04
R-squared                      .2574265445
F ratio                      15.9467335680     P value           .00001
--------------------------------------------------------------------------
```

```
Dataset Name:   Si_Resistivity     (NIST-si_resistivity.dat)
Procedure:      Analysis of Variance
Reference:      Ehrstein, James and Croarkin, M. Carroll.
                Unpublished NIST dataset.
Data:           1 Factor
                5 Treatments
                5  Replicates/Cell
                25 Observations
                3 Constant Leading Digits
                Lower Level of Difficulty
                Observed Data
Model:          6 Parameters (mu,tau_1, ... , tau_5)
                y_{ij} = mu + tau_i + epsilon_{ij}
Certified Values:
Source of                Sums of                Mean
Variation          df      Squares                Squares             F Statistic
Between     Instrument          4     5.11462616000000E-02     1.27865654000000E-02
1.18046237440255E+00
Within Instrument  20 2.16636560000000E-01 1.08318280000000E-02
                   Certified R-Squared 1.90999039051129E-01
                   Certified Residual
                   Standard Deviation  1.04076068334656E-01
Data:  Instrument   Resistance
Read ; Nobs=25 ; Nvar=1 ; Names=y ; ByVariables$
196.3052  196.1240  196.1890  196.2569  196.3403
196.3042  196.3825  196.1669  196.3257  196.0422
196.1303  196.2005  196.2889  196.0343  196.1811
196.2795  196.1748  196.1494  196.1485  195.9885
196.2119  196.1051  196.1850  196.0052  196.2090
Regress; Lhs=y ; Rhs=one ; Pds = 5 ;  Panel $
```

```
--------------------------------------------------------------------------
Analysis of Variance for         Y
Stratification Variable          _STRATUM
Total Sample Size                                  25      Group Sizes
Number of Groups                                    5      Max =     5
Number of groups with no data                       0      Min =     5
Overall Sample Mean                       196.1891560      Avg =   5.0
Total Sample Minimum                      195.9885000
Total Sample Maximum                      196.3825000
Sample Standard Deviation                    .1056296
Total Sample Variance                        .0111576
Source of Variation             Variation     Deg.Fr.      Mean Square
Between Groups             .5114626160D-01          4  .1278656540D-01
Within  Groups             .2166365600D+00         20  .1083182800D-01
Total                      .2677828216D+00         24  .1115761757D-01
Residual S.D.              .1040760683D+00
R-squared                      .1909990391
F ratio                       1.1804623744     P value           .34781
--------------------------------------------------------------------------
```

E16.4 The Group Means Estimator

The regression model in terms of group means is specified as

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

so

$$\bar{y}_i = \alpha_i + \boldsymbol{\beta}'\bar{\mathbf{x}}_i + \bar{\varepsilon}_i$$

This is a possibly heteroscedastic regression, Var[$\bar{\varepsilon}_i$] = σ^2/T_i. The coefficients are estimated by weighted least squares.

The group means estimator is computed as part of the computation of the random effects model. Since it is only an intermediate result, it is discarded at the end of estimation. If you wish to produce this as an estimator in its own right, use

; Means

in the command. In this case, the group means estimator is the only estimator produced.

For our previous application, now using all 10 firms in the sample, we obtain the following:

REGRESS ; Lhs = i ; Rhs = one,f,c ; Panel ; Pds = 20 ; Means $

```
-----------------------------------------------------------------------------
Group Means Regression............................
Ordinary     least squares regression ............
LHS=YBAR(i.) Mean                    =      145.95825
             Standard deviation      =      198.82421
WTS=NTi/Nobs Number of observs.      =             10
Model size   Parameters              =              3
             Degrees of freedom      =              7
Residuals    Sum of squares          =        50603.2
             Standard error of e     =       85.02366
Fit          R-squared               =         .85777
             Adjusted R-squared      =         .81713
Model test   F[  2,      7] (prob) =    21.1(.0011)
Diagnostic   Log likelihood          =      -56.83531
             Restricted(b=0)         =      -66.58679
             Chi-sq [  2]  (prob) =  19.5(  .0001)
Info criter. Akaike Info. Criter. =        9.12918
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
       F|     .13465***       .02875     4.68  .0000      .07831     .19099
       C|     .03203          .19094      .17  .8668    -.34220     .40626
Constant|   -8.52711        47.51531     -.18  .8576  -101.65541   84.60118
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

All of the standard results for regression models are saved by this estimator. Among the scalars, however, the *logl* should be ignored. Also, *rho* is not computed.

E16.5 The Pooled Regression

The pooled regression treats the panel as if it were a single cross section. This can obviously be obtained with the **REGRESS** command discussed in Chapters E7 and E8. You can use, instead,

REGRESS ; Lhs = dependent variable
; Rhs = independent variables
; Panel
; Robust ? This option requests the cluster correction
; Pooled $

to obtain additional output related to the panel. The example below illustrates. In addition to the standard regression results, this regression contains the univariate analysis of variance for the dependent variable and three specification tests for the model. The test statistics are detailed in Section E16.6.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                   =      145.95825
             Standard deviation     =      216.87530
             No. of observations    =            200  Degrees of freedom
Regression   Sum of Squares         =    .760409E+07          2
Residual     Sum of Squares         =    .175585E+07        197
Total        Sum of Squares         =    .935994E+07        199
             Standard error of e    =       94.40840
Fit          R-squared              =         .81241  R-bar squared =   .81050
Model test   F[  2,    197]         =      426.57573  Prob F > F*   =   .00000
Diagnostic   Log likelihood         =    -1191.80236  Akaike I.C.   =  9.11015
             Restricted (b=0)       =    -1359.15096  Bayes  I.C.   =  9.15962
             Chi squared [  2]      =      334.69719  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]      =      798.16155  Prob C2 > C2* =   .00000
[High values of  LM favor FEM/REM over base model]
Baltagi-Li form of LM Statistic     =      798.16155  [= BP if balanced panel]
Moulton/Randolph form:SLM N[0,1]    =      35.715641
Robust cluster corrected covariance matrix used
--------------------------------------------------------
Panel Data Analysis of I                   [ONE way]
               Unconditional ANOVA (No regressors)
Source          Variation  Deg. Free.    Mean Square
Between    7115591.65455          9.   790621.29495
Residual   2244352.27433        190.    11812.38039
Total      9359943.92889        199.    47034.89412
--------+--------------------------------------------------------------------
        |                  Standard             Prob.      95% Confidence
       I|  Coefficient       Error       t     |t|>T*          Interval
--------+--------------------------------------------------------------------
       F|     .11556***      .01627     7.10  .0000      .08367     .14745
       C|     .23068***      .08698     2.65  .0086      .06021     .40115
Constant|   -42.7144**     20.90839    -2.04  .0424    -83.6941    -1.7347
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E16.6 Specification Test for the One Factor Panel Models

Breusch and Pagan's Lagrange multiplier statistic,

$$\text{LM} = \frac{1}{2}\frac{\left(\sum_t T_i\right)^2}{\sum_i T_i(T_i-1)}\left[\frac{\sum_i\left(\sum_t e_{it}\right)^2}{\sum_i\sum_t e_{it}^{\ 2}} \ \text{-1}\right]^2$$

is used to test the null hypothesis that there are no group effects in the random effects model. Arguably, a rejection of the null hypothesis is as likely to be due to the presence of fixed effects. The statistic is computed from the ordinary least squares residuals from a pooled regression. Large values of LM favor the effects model over the classical model with no common effects. The Breusch and Pagan LM statistic is presented with the pooled regression results as shown in the preceding example. This is a chi squared statistic with one degree of freedom.

Two alternative forms of the LM statistic are presented in the pooled regression results. The Baltagi and Li (1990) version of LM is

$$\text{BL - LM} = \frac{1}{2}\left[\frac{\left(N\overline{T}\right)^2}{\left(\sum_i T_i^2\right)-N\overline{T}}\right]\left[\frac{\sum_i\left(\sum_t e_{it}\right)^2}{\sum_i\sum_t e_{it}^{\ 2}} \ \text{-1}\right]^2, \ \overline{T} = \frac{N}{\sum_i(1/T_i)}.$$

This statistic is identical to the Breusch and Pagan statistic when the panel is balanced. The authors argue that the small sample performance is better for unbalanced panels. A second alternative is the Moulton and Randolph (1989) statistic, which is more involved: It is computed as follows

$$\mathbf{z}_i = T_i\overline{\mathbf{x}}_i, \qquad tr_i = T_i + \mathbf{z}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}_i,$$

$$\text{Trace} = \sum_i tr_i, \quad \text{Trace2} = \sum_i tr_i^2$$

$$\text{E} = \frac{Trace}{\sum_i T_i}, \qquad V = \frac{2\left[\left(\left(\sum_i T_i\right)Trace2\right) - Trace^2\right]}{\left(\sum_i T_i\right)^2\left[\left(\sum_i T_i\right)+2\right]},$$

$$\text{MR} = \frac{\left(\dfrac{\sum_i\left(\sum_t e_{it}\right)}{\sum_i\sum_t e_{it}^2}\right) - \text{E}}{\sqrt{V}}.$$

The limiting distribution of MR is standard normal, so values in excess of 1.96 weigh against the base regression model. The Baltagi and Li and the Moulton and Randolph statistics are presented in the results for the pooled regression.

VERSION NOTE: In earlier versions of *LIMDEP*, the Breusch and Pagan and the Baltagi and Li statistics were reported with the results for the random effects model. They have been moved to the results for the pooled regression since they are computed and used with reference to the pooled model (and are more useful there) and, in addition, they can be computed and reported without actually computing the random effects estimator. The Moulton and Randolph statistic is new with Version 10.

E16.7 One Way Fixed and Random Effects Models

The next two chapters consider formulation and estimation of one way common effects models,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

The fixed effects model is

$$\begin{aligned} y_{it} &= \alpha_1 d_{1it} + \alpha_2 d_{2it} + \ldots + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} \\ &= \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, \end{aligned}$$

where $E[\varepsilon_{it}|\mathbf{X}_i] = 0$, $\text{Var}[\varepsilon_{it}|\mathbf{X}_i] = \sigma^2$, $\text{Cov}[\varepsilon_{it},\varepsilon_{js}|\mathbf{X}_i,\mathbf{X}_j] = 0$ for all i,j,

$$\text{Cov}[\alpha_i,\mathbf{x}_{it}] \neq \mathbf{0}.$$

The efficient estimator for this model in the base case is least squares. This model is documented in Chapter E17. The random effects model is

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$$

where $E[u_i] = 0$, $\text{Var}[u_i] = \sigma_u^2$ $\text{Cov}[\varepsilon_{it},u_i] = 0$.

$$\text{Var}[\varepsilon_{it} + u_i] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2 .$$

For a given i, the disturbances in different periods are correlated because of their common component, u_i,

$$\text{Corr}[\varepsilon_{it} + u_i,\varepsilon_{is} + u_i] = \rho = \sigma_u^2 / \sigma^2.$$

The efficient estimator is generalized least squares. This model is developed in Chapter E18.

E17: Fixed Effects Linear Regression

E17.1 Introduction

This chapter will detail estimation of linear regression models with fixed effects. The essential structure is,

$$y_{it} = \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

in which variation across groups (individuals) or time is captured in simple shifts of the regression function – i.e., changes in the intercepts. The user is referred to textbook treatments such as Greene (2011) or Wooldridge (2010) for background theory of the models. The two models estimated with this program are 'one way' or 'one factor' designs of the form

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

where ε_{it} is a classical disturbance with $E[\varepsilon_{it}|\mathrm{x}_{\mathrm{it}}] = 0$ and $\mathrm{Var}[\varepsilon_{it}|\mathrm{x}_{\mathrm{it}}] = \sigma_\varepsilon^2$ and 'two way' or 'two factor' models as shown in the first equation above.

E17.2 One Way Fixed Effects Model

In the fixed effects model (FEM), α_i is a separate constant term for each unit. Thus, the model may be written

$$\begin{aligned} y_{it} &= \alpha_1 d_{1it} + \alpha_2 d_{2it} + \ldots + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} \\ &= \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, \end{aligned}$$

where the α_is are individual specific constants, and the d_js are group specific dummy variables which equal one only when $j = i$. The fixed effects model is an ordinary linear regression model. The complication for the least squares procedure is that N may be very large so that the usual formulas for computing least squares coefficients are cumbersome (or impossible) to apply. The model may be estimated in a simpler form by exploiting the algebra of least squares.

E17.2.1 Command for One Factor Models

The one way FEM is a linear regression with N dummy variables (and no overall constant term). To invoke this procedure, use the command the panel is set up with

SETPANEL ; Group = the group identifier ; Pds = variable to use for group counts $

Then,

```
REGRESS   ; Lhs = y
          ; Rhs = list of regressors
          ; Panel
          ; Fixed Effects $
```

You need not include *one* among your regressors. The constant is placed in the regression automatically when it is needed. You may also use:

; Output = 2

to list fixed effects in an output file. This will also produce estimated standard errors for the fixed effects. If the number of groups is large, the amount of output can be very large.

The basic command can be constructed using the **REGRESS** command builder shown earlier. You can also use the command builder for many optional features. Select Panel data model on the Options page to activate the Model type window and the Settings button. Select your model type, then click Settings to open a dialog box of optional features for that type of model. Options for fixed and random effects models are listed in the same dialog box, as shown in Figure E17.1.

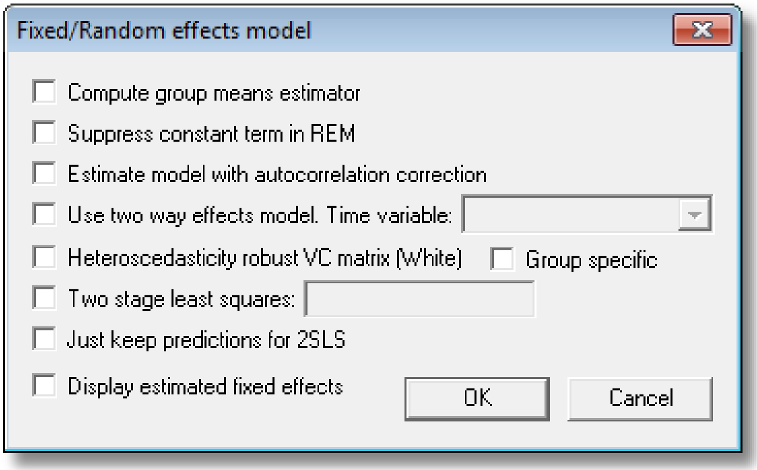


Figure E17.1 Command Builder Options for Common Effects Models

Standard options for residuals and fitted values, include the following. All of

; List	to display residuals and fitted values
; Keep = name	to retain predictions
; Res = name	to retain residuals
; Var =	a submatrix of the parameter VC matrix
; Fill	(missing observations)
; Wts = weighting variable	
; Covariance Matrix (or ; Printvc)	

are available as usual. If your stratification indicators are set up properly for out of sample observations, **; Fill** will allow you to extrapolate outside the estimation sample.

WARNING: If you do not have a stratification indicator already in use, **; Fill** will not work. The *_stratum* variable is set up only for the estimation sample. Thus, with **; Pds = T**, you cannot extrapolate outside the sample.

E17.2.2 Program Output for One Way Fixed Effects Models

Two full sets of estimates are computed by this estimator:

1. **Pooled Regression:** The fixed effects model above with all of the individual specific constants assumed equal is $y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$. This model is estimated by simple ordinary least squares.

2. **Least Squares Dummy Variable:** The fixed effects model with individual specific constant terms is estimated by partitioned ordinary least squares. For the one factor models, we formulate this model with N group specific constants and no overall constant.

The second constitutes the results for the fixed effects estimator. Define the four models:

Model	1	$y_{it} = \alpha + \varepsilon_{it}$	(no group effects or xs),
Model	2	$y_{it} = \alpha_i + \varepsilon_{it}$	(group dummies only),
Model	3	$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$	(regressors only),
Model	4	$y_{it} = \alpha_I + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$	(regressors and group effects).

Output from this program, in the order in which it will appear, is as follows:

1. Pooled linear regression of y on a single constant and the regressors, $x_1,...,x_K$. *These K variables do not include one*. This is Model 3 above. Output consists of the standard results for least squares regression. The diagnostic statistics in this regression output will also include the unconditional analysis of variance for the dependent variable. This is the usual ANOVA for the groups, ignoring the regressors. The output from this procedure could be used to test the hypothesis that the unconditional mean of y is the same in all groups. (This test is done by the program. See part 3 below.) Results at this step also include the Breusch and Pagan test statistic for common effects and two alternatives. See Section E16.6 and the application below for details.

2. Ordinary least squares estimates of Model 4 above. Output is the same as in part 1, the usual for a least squares regression. The estimates of the dummy variable coefficients and the estimated standard errors are listed in the output file if requested with **; Output = 2**. (There may be hundreds or thousands of them!)

3. Test statistics for the various classical models. The table contains

   a. For Models 1-4, the log likelihood function, sum of squared residuals based on the least squares estimates, and R^2.

   b. Chi squared statistics based on the likelihood functions and F statistics based on the sums of squares for testing the restrictions of:

      - Model 1 as a restriction on Model 2 (no group effects on the mean of y),
      - Model 1 as a restriction on Model 3 (no fit in the regression of y on xs),
      - Model 1 as a restriction on Model 4 (no group effects or fit in regression),
      - Model 2 as a restriction on Model 4 (group effects but no fit in regression),
      - Model 3 as a restriction on Model 4 (fit in regression but no group effects).

The statistic, degrees of freedom, and prob value (probability that the statistic would be equaled or exceeded by the chi squared or F random variable) are given for each hypothesis.

E17.2.3 Saved Results

The retrievable results are:

Matrices:	*b* and *varb*		
	alphafe		contains the estimates of the fixed effects, α_i. This matrix is limited to 50,000 cells, so if your data have more than 20,000 groups, *alphafe* will contain the first 50,000 fixed effects computed.
Scalars:	*ssqrd*	=	s^2 from least squares dummy variable (LSDV)
	rsqrd	=	R^2 from LSDV
	s	=	$\sqrt{s^2}$ from LSDV
	sumsqdev	=	sum of squared residuals from LSDV
	rho	=	estimated disturbance autocorrelation from whatever model is fit last
	degfrdm	=	$\Sigma_i T_i - K$
	sy	=	standard deviation of Lhs variable
	ybar	=	mean of Lhs variable
	kreg	=	K
	nreg	=	total number observations
	logl	=	log likelihood from LSDV model
	exitcode	=	0.0 if the model was estimable
	ngroup	=	number of groups
	nperiod	=	number of periods. This will be 0.0 if you fit a one way model.
Last Model:	*b_variable*		constructed as usual

Last Function: Conditional mean **= b'x**

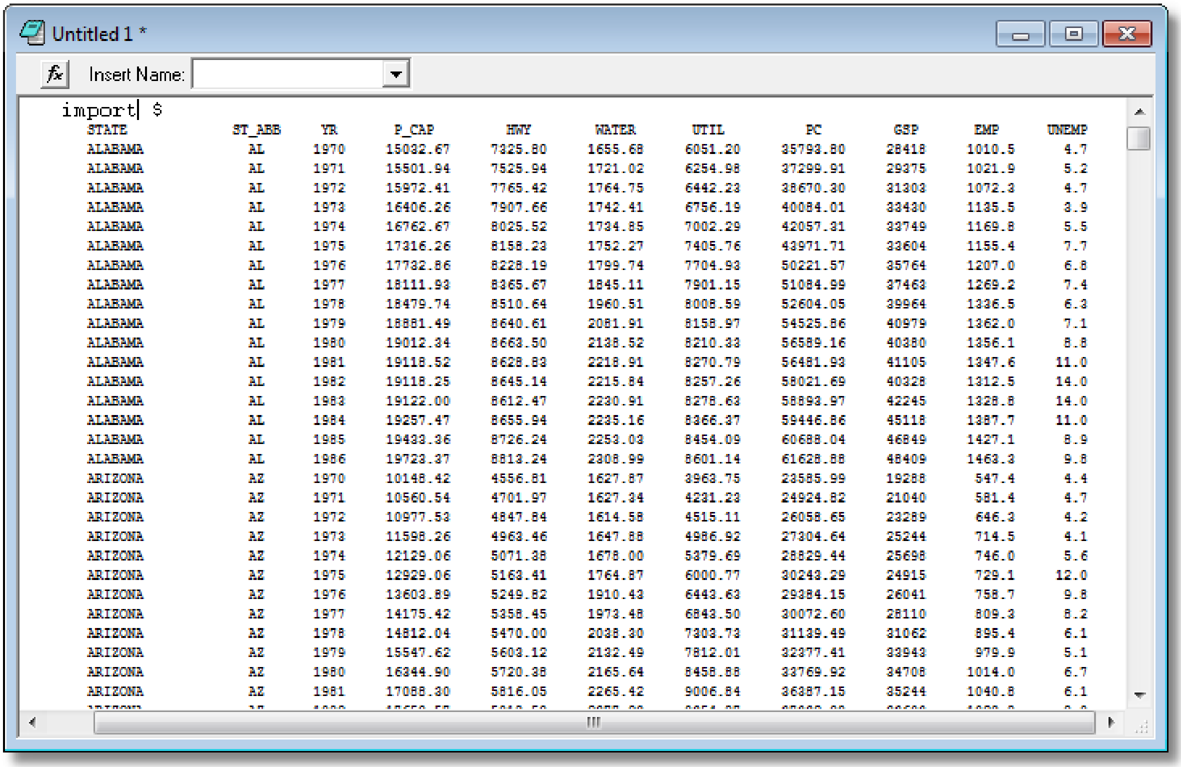
Predicted values are based on the last model estimated, one or two way, fixed or random. Predictions are not listed when you use the group means estimator, but they can be computed with **MATRIX**. Estimates of the variances or standard errors of the fixed effects are not kept. But, a simple method of computing them is given below.

Note, the implication of not storing the constants (there could be thousands of them) is that because the model is linear, **PARTIALS** will give you the right answer for partial effects even when there are interactions or nonlinearities in the model. However, **SIMULATE** will not give the correct predicted values – the appropriate function would be a_i + $\mathbf{b'x}_{it}$. Predictions using **; List** and **; Keep** and residuals requested with **; Res** are computed correctly using a_i as indicated.

E17.2.4 Application

The examples to follow are based on an application in Baltagi (2005) which describes Munnell's (1990) study of statewide productivity. The data were downloaded from the website for the text: http://www.wiley.com/legacy/wileychi/baltagi/supp/PRODUC.prn. The data are a balanced (17 years) panel of observations on the 'lower' 48 states. Variables in the data set are

state	=	state ID (changed from the name in the original)
yr	=	year, 1970,...,1986
p_cap	=	public capital
hwy	=	highway capital
water	=	water utility capital
util	=	utility capital
pc	=	private capital
gsp	=	gross state product
emp	=	employment
unemp	=	unemployment rate



STATE	ST_ABB	YR	P_CAP	HWY	WATER	UTIL	PC	GSP	EMP	UNEMP
ALABAMA	AL	1970	15032.67	7325.80	1655.68	6051.20	35793.80	28418	1010.5	4.7
ALABAMA	AL	1971	15501.94	7525.94	1721.02	6254.98	37299.91	29375	1021.9	5.2
ALABAMA	AL	1972	15972.41	7765.42	1764.75	6442.23	38670.30	31303	1072.3	4.7
ALABAMA	AL	1973	16406.26	7907.66	1742.41	6756.19	40084.01	33430	1135.5	3.9
ALABAMA	AL	1974	16762.67	8025.52	1734.85	7002.29	42057.31	33749	1169.8	5.5
ALABAMA	AL	1975	17316.26	8158.23	1752.27	7405.76	43971.71	33604	1155.4	7.7
ALABAMA	AL	1976	17732.86	8228.19	1799.74	7704.93	50221.57	35764	1207.0	6.8
ALABAMA	AL	1977	18111.93	8365.67	1845.11	7901.15	51084.99	37463	1269.2	7.4
ALABAMA	AL	1978	18479.74	8510.64	1960.51	8008.59	52604.05	39964	1336.5	6.3
ALABAMA	AL	1979	18881.49	8640.61	2081.91	8158.97	54525.86	40979	1362.0	7.1
ALABAMA	AL	1980	19012.34	8663.50	2138.52	8210.33	56589.16	40380	1356.1	8.8
ALABAMA	AL	1981	19118.52	8628.83	2218.91	8270.79	56481.93	41105	1347.6	11.0
ALABAMA	AL	1982	19118.25	8645.14	2215.84	8257.26	58021.69	40328	1312.5	14.0
ALABAMA	AL	1983	19122.00	8612.47	2230.91	8278.63	58893.97	42245	1328.8	14.0
ALABAMA	AL	1984	19257.47	8655.94	2235.16	8366.37	59446.86	45118	1387.7	11.0
ALABAMA	AL	1985	19433.36	8726.24	2253.03	8454.09	60688.04	46849	1427.1	8.9
ALABAMA	AL	1986	19723.37	8813.24	2308.99	8601.14	61628.88	48409	1463.3	9.8
ARIZONA	AZ	1970	10148.42	4556.81	1627.87	3963.75	23585.99	19288	547.4	4.4
ARIZONA	AZ	1971	10560.54	4701.97	1627.34	4231.23	24924.82	21040	581.4	4.7
ARIZONA	AZ	1972	10977.53	4847.84	1614.58	4515.11	26058.65	23289	646.3	4.2
ARIZONA	AZ	1973	11598.26	4963.46	1647.88	4986.92	27304.64	25244	714.5	4.1
ARIZONA	AZ	1974	12129.06	5071.38	1678.00	5379.69	28829.44	25698	746.0	5.6
ARIZONA	AZ	1975	12929.06	5163.41	1764.87	6000.77	30243.29	24915	729.1	12.0
ARIZONA	AZ	1976	13603.89	5249.82	1910.43	6443.63	29384.15	26041	758.7	9.8
ARIZONA	AZ	1977	14175.42	5358.45	1973.48	6843.50	30072.60	28110	809.3	8.2
ARIZONA	AZ	1978	14812.04	5470.00	2038.30	7303.73	31139.49	31062	895.4	6.1
ARIZONA	AZ	1979	15547.62	5603.12	2132.49	7812.01	32377.41	33943	979.9	5.1
ARIZONA	AZ	1980	16344.90	5720.38	2165.64	8458.88	33769.92	34708	1014.0	6.7
ARIZONA	AZ	1981	17088.30	5816.05	2265.42	9006.84	36387.15	35244	1040.8	6.1

Figure E17.2 Importing Munnell State Production Data

We will fit the loglinear regression model

$$\log gsp_{it} = \alpha_i + \beta_1 \log p\_cap_{it} + \beta_2 \log hwy_{it} + \beta_3 \log water_{it} + \beta_4 \log util_{it} + \beta_5 \log pc_{it} + \beta_6 \log emp_{it} + \varepsilon_{it}$$

Estimates of the fixed effects model follow:

```
CREATE      ; loggsp  = Log(gsp)
            ; logkp   = Log(p_cap)
            ; loghwy  = Log(hwy)
            ; logh2o  = Log(water)
            ; logutil = Log(util)
            ; logemp  = Log(emp)$
NAMELIST    ; x = logkp,loghwy,logh2o,logutil,logemp $
CREATE      ; stateid = Trn(17,0) $
SETPANEL    ; Group = stateid ; Pds = ti $
REGRESS     ; Lhs = loggsp ; Rhs = x,one
            ; Panel ; Fixed Effects ; Parameters ; Output = 2 $
```

```
+-------------------------------------------------------------------+
| Variable = ____________ Variable Groups    Max    Min    Average |
| TI          Group sizes  STATEID      48     17     17       17.0 |
+-------------------------------------------------------------------+

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGGSP   Mean                    =      10.50885
             Standard deviation      =       1.02113
             No. of observations     =           816  Degrees of freedom
Regression   Sum of Squares          =       838.219          5
Residual     Sum of Squares          =       11.5898        810
Total        Sum of Squares          =       849.809        815
             Standard error of e     =        .11962
Fit          R-squared               =        .98636  R-bar squared =   .98628
Model test   F[  5,   810]           =   11716.51224  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =     577.89764  Akaike I.C.   = -4.23959
             Restricted (b=0)        =   -1174.41748  Bayes  I.C.   = -4.20500
             Chi squared [  5]       =    3504.63025  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]       =    4462.89033  Prob C2 > C2* =   .00000
[High values of  LM favor FEM/REM over base model]
Baltagi-Li form of LM Statistic      =    4462.89033  [= BP if balanced panel]
Moulton/Randolph form:SLM N[0,1]     =     75.214155
------------------------------------------------------------
Panel Data Analysis of LOGGSP              [ONE way]
              Unconditional ANOVA (No regressors)
Source        Variation  Deg. Free.   Mean Square
Between       830.86743         47.      17.67803
Residual       18.94145        768.        .02466
Total         849.80888        815.       1.04271
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .45392***     .15355    2.96  .0031      .15298     .75487
  LOGHWY|     .08572        .08184    1.05  .2949     -.07468     .24612
  LOGH2O|     .08663***     .02479    3.50  .0005      .03805     .13521
 LOGUTIL|    -.18742***     .06580   -2.85  .0044     -.31639    -.05845
  LOGEMP|     .61908***     .02281   27.14  .0000      .57437     .66380
Constant|    2.01100***     .15245   13.19  .0000     1.71221    2.30979
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
LSDV          least squares with fixed effects ....
LHS=LOGGSP    Mean                 =      10.50885
              Standard deviation   =       1.02113
              No. of observations  =           816  Degrees of freedom
Regression    Sum of Squares       =       848.708          52
Residual      Sum of Squares       =       1.10080         763
Total         Sum of Squares       =       849.809         815
              Standard error of e  =        .03798
Fit           R-squared            =        .99870  R-bar squared =   .99862
Model test    F[ 52,   763]        =   11312.81044  Prob F > F*   =   .00000
Diagnostic    Log likelihood       =    1538.36346  Akaike I.C.   = -6.47847
              Restricted (b=0)     =   -1174.41748  Bayes  I.C.   = -6.17292
              Chi squared [ 52]    =    5425.56189  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)    =       .725563
-------------------------------------------------------
Panel:Groups Empty      0,      Valid data        48
              Smallest  17,      Largest           17
              Average group size in panel     17.00
Variances     Effects a(i)        Residuals e(i,t)
               .022553                      .001443
--------+-------------------------------------------------------------------
        |              Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
   LOGKP|     .56623***      .10400     5.44  .0000      .36240    .77007
  LOGHWY|    -.23193***      .06325    -3.67  .0002     -.35590   -.10796
  LOGH2O|     .05375***      .01871     2.87  .0041      .01707    .09043
 LOGUTIL|    -.33878***      .04340    -7.81  .0000     -.42385   -.25372
  LOGEMP|    1.00378***      .02010    49.94  .0000      .96439   1.04318
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

```
        Estimated Fixed Effects
        Group       Coefficient       Standard Error       t-ratio
            1           2.54918               .20118      12.67133
            2           2.67049               .19329      13.81593
            3           2.65383               .18995      13.97131
            4           2.66292               .23066      11.54497
            5           2.62603               .19259      13.63501
           (Rows 6 – 47 omitted)
           48           3.23311               .19324      16.73130
```

```
+----------------------------------------------------------------------+
|              Test Statistics for the Regression Model                |
+----------------------------------------------------------------------+
|      Model              Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only     -1174.41747       849.80888      .00000 |
|(2)  Group effects only      377.47534        18.94145      .97771 |
|(3)  X - variables only      577.89766        11.58975      .98636 |
|(4)  X and group effects    1538.36348         1.10080      .99870 |
+----------------------------------------------------------------------+
```

```
+----------------------------------------------------------------------+
|                        Hypothesis Tests                              |
|         Likelihood Ratio Test             F Tests                    |
|         Chi-squared   d.f.   Prob          F    num    denom  P value |
|(2) vs (1)   3103.79    47  .0000     716.77    47      768   .00000 |
|(3) vs (1)   3504.63     5  .0000   11716.51     5      810   .00000 |
|(4) vs (1)   5425.56    52  .0000   11312.81    52      763   .00000 |
|(4) vs (2)   2321.78     5  .0000    2473.18     5      763   .00000 |
|(4) vs (3)   1920.93    47  .0000     154.69    47      763   .00000 |
+----------------------------------------------------------------------+
```

The individual effects are accessible as a matrix in the work area. The following replicates the computations that underlie the listing after the LSDV results above. The individual effects are computed as

$$a_i = \overline{y}_i - \mathbf{b}'_{LSDV}\overline{\mathbf{x}}_i$$

The appropriate estimator of the asymptotic variance of a_i is

$$\text{Est.Asy.Var}[a_i] = \frac{s^2}{T_i} + \overline{\mathbf{x}}_i'\left[s^2(\mathbf{X}'\mathbf{M}_D'\mathbf{M}_D\mathbf{X})^{-1}\right]\mathbf{x}_i,\ s^2 = \frac{\Sigma_i\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{b}'_{LSDV}\mathbf{x}_{it})^2}{(\Sigma_i T_i) - n - K}.$$

The matrix $\mathbf{XM_D'M_DX}$ is the moment matrix computed using deviations from group means;

$$\mathbf{X}'\mathbf{M}_\mathbf{D}'\mathbf{M}_\mathbf{D}\mathbf{X} = \sum_{i=1}^{n}\sum_{t=1}^{T_i}\left(\mathbf{x}_{it} - \overline{\mathbf{x}}_i\right)\left(\mathbf{x}_{it} - \overline{\mathbf{x}}_i\right)'.$$

```
MATRIX      ; mti = Gsiz(stateid)        $ Computes T_i
MATRIX      ; xbr = Gxbr(x,stateid)      $ Obtains group means of x
MATRIX      ; varai = ssqrd*Diri(mti) + Qrow(xbr,varb) $
CLIST       ; statenm = _group_          $ State labels next to data matrix
DISPLAY     ; Parameters = alphafe
            ; Covariance = varai         ? A vector of variances, not the whole matrix
            ; Labels = statenm           $
```

This program produces the following results

```
-----------------------------------------------------------------------------
User Specified Model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
 ALABAMA|    2.54918***      .20118    12.67  .0000     2.15488    2.94348
 ARIZONA|    2.67049***      .19329    13.82  .0000     2.29165    3.04934
 (Arkansas – West Virginia omitted)
WISCONSI|    2.51680***      .20445    12.31  .0000     2.11608    2.91751
 WYOMING|    3.23311***      .19324    16.73  .0000     2.85437    3.61185
--------+--------------------------------------------------------------------
```

E17.2.5 Robust Estimation of the Fixed Effects Covariance Matrix

Under the assumptions of the model made at the outset, the appropriate covariance matrix for the fixed effects coefficient is shown in the preceding example. If it is believed that there is residual correlation across observations in the groups even with the individual effects included, one can compute a 'cluster correction' for the asymptotic covariance matrix. The correction would be

$$Est.Asy.Var[b_{LSDV}] = (\mathbf{X'M_D'M_DX})^{-1}\left[A\sum_{i=1}^{n}\left\{\left(\mathbf{M}_D^{(i)}\mathbf{X}^{(i)}\right)\mathbf{e}_i\right\}\left\{\left(\mathbf{M}_D^{(i)}\mathbf{X}^{(i)}\right)\mathbf{e}_i\right\}'\right](\mathbf{X'M_D'M_DX})^{-1}$$

$$A = \frac{n}{n-1}\frac{(\Sigma_t T_i)-1}{(\Sigma_t T_i)-(n+K)}$$

This correction is requested by adding

; Robust

to the command. Note, this is the 'cluster estimator' used elsewhere; we use **; Robust** to distinguish it here because the calculations are already accommodating clustering. The impact on the LSDV results is shown below. The effect on the estimated standard errors is substantial. (The other statistical results are the same.)

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|      .56623***     .10400     5.44  .0000      .36240    .77007
  LOGHWY|    -.23193***      .06325    -3.67  .0002    -.35590   -.10796
  LOGH2O|      .05375***     .01871     2.87  .0041      .01707    .09043
 LOGUTIL|    -.33878***      .04340    -7.81  .0000    -.42385   -.25372
  LOGEMP|    1.00378***      .02010    49.94  .0000      .96439   1.04318
--------+-------------------------------------------------------------
(Cluster corrected estimates)
   LOGKP|      .56623**      .23126     2.45  .0143      .11297   1.01949
  LOGHWY|    -.23193         .16730    -1.39  .1657    -.55984    .09598
  LOGH2O|      .05375        .04468     1.20  .2290    -.03383    .14133
 LOGUTIL|    -.33878***      .07115    -4.76  .0000    -.47823   -.19934
  LOGEMP|    1.00378***      .05186    19.35  .0000      .90213   1.10544
--------+-------------------------------------------------------------
```

When the **REGRESS** command contains **; Robust**, the original pooled estimator and the subsequent random effects estimator, if it is computed, are also computed with this robust covariance matrix correction.

E17.2.6 Fixed Effects Models with Time Invariant Variables

The fixed effects model generally cannot be estimated when the data contain time invariant variables. The reason is that the within groups transformation used to fit the coefficients,

regression of (y_{it} - $\bar{y}_i$) on ($\mathbf{x}_{it}$ - $\bar{\mathbf{x}}_i$) with no constant term),

produces a column of zeros for every time invariant variable. This is a problem of perfect collinearity, since a time invariant variable is just a multiple of the individual specific dummy variable. *LIMDEP* does not halt the regression when this condition is detected. For reasons noted shortly, *LIMDEP always* uses a generalized inverse and computes the regression anyway. However, the condition is noted. When there are no time invariant variables, the G-inverse gives precisely the correct result. When there are time invariant variables, there will be superfluous coefficients, but the results are still useable and clearly identified in the results. An example will illustrate. We first create a variable that contains no within state variation

CREATE ; lempbar = Group Mean(logemp, Pds = 17) $

```
+-----------------------------------------------------------------+
| Variable = ____________ Variable Groups     Max     Min   Average |
| LEMPBAR    Group means  LOGEMP        48      17      17      17.0 |
+-----------------------------------------------------------------+
```

The fixed effects model is recomputed with

REGRESS ; Lhs = loggsp
; Rhs = x,lempbar,one
; Panel ; Fixed Effects $

```
-----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=LOGGSP   Mean                 =      10.50885
             Standard deviation   =       1.02113
             No. of observations  =           816  Degrees of freedom
Regression   Sum of Squares       =       848.708          53
Residual     Sum of Squares       =       1.10080         762
Total        Sum of Squares       =       849.809         815
             Standard error of e  =        .03801
Fit          R-squared            =        .99870  R-bar squared =   .99861
Model test   F[ 53,    762]       =   11084.81418  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    1538.36346  Akaike I.C.   = -6.47602
             Restricted (b=0)     =   -1174.41748  Bayes  I.C.   = -6.16470
             Chi squared [ 53]    =    5425.56189  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)   =       .725563
--------------------------------------------------
Panel:Groups Empty      0,     Valid data       48
             Smallest  17,     Largest          17
             Average group size in panel     17.00
Variances    Effects a(i)         Residuals e(i,t)
              .133481                    .001445
These  1 variables have no within group variation.
LEMPBAR
F.E. estimates are based on a generalized inverse.
```

```
--------+-------------------------------------------------------------
        |                  Standard             Prob.       95% Confidence
  LOGGSP|  Coefficient       Error        z     |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|     .56623***      .10407      5.44  .0000      .36226     .77020
  LOGHWY|    -.23193***      .06329     -3.66  .0002     -.35598    -.10788
  LOGH2O|     .05375***      .01873      2.87  .0041      .01704     .09045
 LOGUTIL|    -.33878***      .04343     -7.80  .0000     -.42391    -.25366
  LOGEMP|    1.00378***      .02011     49.90  .0000      .96436    1.04321
 LEMPBAR|       0.0     .....(Fixed Parameter).....
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
----------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
|              Test Statistics for the Regression Model                |
+----------------------------------------------------------------------+
|       Model             Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only    -1174.41747        849.80888      .00000 |
|(2)  Group effects only      377.47534         18.94145      .97771 |
|(3)  X - variables only      597.16828         11.05507      .98699 |
|(4)  X and group effects    1538.36348          1.10080      .99870 |
+----------------------------------------------------------------------+
|                         Hypothesis Tests                             |
|         Likelihood Ratio Test           F Tests                      |
|         Chi-squared   d.f.   Prob          F    num   denom  P value |
|(2) vs (1)   3103.79     47  .0000    716.77    47     768   .00000 |
|(3) vs (1)   3543.17      6  .0000  10229.87     6     809   .00000 |
|(4) vs (1)   5425.56     53  .0000  11084.81    53     762   .00000 |
|(4) vs (2)   2321.78      6  .0000   2058.28     6     762   .00000 |
|(4) vs (3)   1882.39     47  .0000    146.61    47     762   .00000 |
+----------------------------------------------------------------------+
```

Note that there is an extensive warning about the time invariant variable(s). However, the regression has been computed. Most importantly, however, notice that the sum of squared residuals and the coefficients on the time varying variables in the two regressions are identical. The coefficient on the time invariant variable is not useable. The small difference in the standard errors in the second model is due to the loss of one degree of freedom (for each time invariant variable) in the second model.

This set of outcomes is noted here for two reasons: You would normally not deliberately add a time invariant variable to a fixed effects model, as we did here. However, one might include one (such as gender or education) inadvertently.

1. *LIMDEP* will warn you of this occurrence. It does not halt estimation. But, in this event, you should reconsider the specification of the model. The reduced specification is not necessarily useful.

2. Although fixed effects models cannot have time invariant variables, random effects models can. That is the reason for the computation. If you request estimation of a random effects model, the sum of squared residuals for the fixed effects model is needed for estimation of the variance components. The presence of the time invariant variables in the model does not prevent this computation. This FE model is always computed, either explicitly if you request fixed effects, or in the background if you have requested random effects (or both fixed and random effects). This method of doing the estimation allows estimation of all models to proceed even in the event of this complication.

NOTE: The recent literature contains a thread of results on a 'Fixed Effects Vector Decomposition' (FEVD) estimator that claims to solve the 'problem' of time invariant variables in a fixed effects model. (See Plumper and Troeger (2007, 2011) and Greene (2011) for discussion.) The so called FEVD estimator is not a 'solution' to this multicollinearity problem. It does reformulate the model so that it is essentially a random effects model. FEVD is not provided as a separate estimator in *LIMDEP*. It can be easily programmed. We return to the computation in Chapter E18 on fitting the random effects model.

E17.2.7 Restricted Least Squares

The (one or two way) fixed effects model can be fit with linear restrictions. (This option does not apply to the random effects estimator.) Use the standard specification,

; CLS: ... linear restrictions ...

The full set of results will be presented for the unrestricted estimates and the restricted estimates, with an F statistic for testing the hypothesis of the restrictions.

For the preceding application, the restriction of constant returns to scale – all slope coefficients sum to one – is a natural one. It is imposed by adding

```
NAMELIST   ; x = logkp,loghwy,logh2o,logutil,logemp $
REGRESS    ; Lhs = loggsp ; Rhs = x,one
           ; Panel ; Fixed Effects
           ; CLS: logkp + loghwy + logh2o + logutil + logemp = 1 $
```

to the **REGRESS** command. (The initial estimated OLS and fixed effects results are omitted.)

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGGSP   Mean                   =      10.50885
             Standard deviation     =       1.02113
             No. of observations    =           816  Degrees of freedom
Regression   Sum of Squares         =       848.696           51
Residual     Sum of Squares         =       1.11300          764
Total        Sum of Squares         =       849.809          815
             Standard error of e    =        .03817
Fit          R-squared              =        .99869  R-bar squared =   .99861
Model test   F[ 51,    764]         =   11453.02586  Prob F > F*   =   .00000
Diagnostic   Log likelihood         =    1533.86733  Akaike I.C.   = -6.46990
             Restricted (b=0)       =   -1174.41748  Bayes  I.C.   = -6.17011
             Chi squared [ 51]      =    5416.56963  Prob C2 > C2* =   .00000
F[ 1,   763] for constraint =    6.4352, P = .0114 <--
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .61507***      .10220     6.02  .0000       .41476    .81538
  LOGHWY|   -.32404***       .05179    -6.26  .0000     -.42555   -.22253
  LOGH2O|     .05154***      .01869     2.76  .0058       .01490    .08818
 LOGUTIL|   -.34469***       .04334    -7.95  .0000     -.42964   -.25975
  LOGEMP|    1.00212***      .02009    49.88  .0000       .96274   1.04150
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

(Additional results for analysis of variance – which have changed – are omitted.)

NOTE: There is a side result which can occur with this computation. The Hausman test requires the covariance matrix of the fixed effects estimator. Normally, this will be larger than that for the REM. But, with restrictions applied to the FEM and not the REM, this need not be the case. In this case (as occurs in this application), the Hausman statistic cannot be computed. The Hausman statistic is described in Chapter E18 where the random effects estimator is documented.

The restrictions affect the estimated fixed effects as well. Fortunately, the simplicity of the LSDV estimator remains even when the restrictions are imposed. The fixed effects are still the group mean residuals, but now computed with the restricted least squares estimator. This is all handled internally. The fixed effects are adjusted for this result after the restrictions are imposed.

E17.2.8 Technical Details on Estimation of One Way Fixed Effects Models

The calculations for the balanced design case are exactly those described in Wooldridge (2011) or Greene (2011). Since these are fully documented, we will just sketch them here. We will then turn to special considerations of the model when groups have unequal numbers of observations. The notations for group and overall means used below are the standard ones. We will refer back to the descriptions and the application in the preceding two sections at various points below.

Computing the Fixed Effects Estimator

Model 1 is estimated simply as the grand mean of *y*, so the sum of squares is

$$TOTAL \quad = \sum_i \sum_{t=1}^{T_i} \left(y_{it} - \bar{\bar{y}}\right)^2.$$

R^2 for this model is zero by definition. The value of *TOTAL* appears in the ANOVA tables in the first set of output for the model.

Analysis of Model 2 is the familiar unconditional analysis of variance for *y* ignoring the regressors. The coefficients would simply be the group means. The total variation above may be decomposed into

$$WITHIN \quad = \sum_i \sum_t \left(y_{it} - \bar{y}_{i.}\right)^2$$

$$BETWEEN \quad = \sum_i T_i \left(\bar{y}_{i.} - \bar{\bar{y}}\right)^2$$

Since $TOTAL \quad = BETWEEN + WITHIN,$

we may define $R_o^2 \quad = BETWEEN \,/\, TOTAL.$

Note that this analysis is equivalent to the regression of *y* on a constant term and a set of *n*-1 group dummy variables or, equivalently, just the *N* group dummy variables with no overall constant. The values of *WITHIN* and this R_o^2 are given as the 'Sum of Squares' and '*R*-squared' in the second row of the Test Statistics for the Classical Model, so *BETWEEN* may be deduced as R_o^2 times *TOTAL*.

Model 3 is the linear regression model. Estimation is by ordinary least squares regression of y_{it} on a single constant and the set of **x**s. No new issues arise. For this model, the groupwise nature of the data set is ignored; the full set of observations is pooled. The analysis of variance for this model is the conventional one. The diagnostic statistics that precede the listing of the coefficient estimates contain the sums of squared residuals, mean and standard deviation of the Lhs variable, and so on.

Model 4 is the full 'dummy variable' model. Parameters are estimated as follows:

estimate $\boldsymbol{\beta}$ by regression of (y_{it} - $\bar{y}_i$) on ($\mathbf{x}_{it}$ - $\bar{\mathbf{x}}_i$) (with no constant term),

estimate α_i with $a_i = \bar{y}_i - \mathbf{b}'\bar{\mathbf{x}}_i$

These calculations follow from the algebra of least squares. The estimated covariance matrix of **b**, sum of squared residuals, and estimator of σ^2 from the first regression are all appropriate as they stand and need not be modified. Estimates of the standard errors of a_is are obtained by

$$\text{Est.Asy.Var}[a_i] = \frac{s^2}{T_i} + \bar{\mathbf{x}}_i'\left[s^2(\mathbf{X}'\mathbf{M}_D'\mathbf{M}_D\mathbf{X})^{-1}\right]\bar{\mathbf{x}}_i,\; s^2 = \frac{\Sigma_i\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{b}_{LSDV}'\mathbf{x}_{it})^2}{(\Sigma_i T_i) - n - K}.$$

None of the preceding relies on equal group sizes. If the group sizes are unequal, as T_i, then, the means are based on the respective group sizes.

As noted earlier, although the estimated fixed effects are retained as matrix *alphafe*, the estimates of the variances are not kept. But, these are easy to obtain, and you can even recover the rest of the estimates if you have more than 50,000 groups. If you have a large number of groups and regressors, you may have to do this in parts. The program shown at the end of Section E17.2.4 can be applied to the entire sample or in separate parts of it.

Restricted Least Squares

The regression model is $y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$. We fit the model subject to the linear restrictions

$$\mathbf{Rb} - \mathbf{q} = \mathbf{0}.$$

Let **y** and **X** denote the full data matrices, and let **D** denote the full matrix of group dummy variables. Let $\mathbf{y}^*$ and $\mathbf{X}^*$ denote the matrix of data in deviations from the group means,

$$(\mathbf{X}^*,\mathbf{y}^*) = [\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'](\mathbf{X},\mathbf{y}).$$

Since the slopes are obtained just by applying ordinary least squares, the restricted slope estimator is obtained by the familiar formula,

$$\mathbf{b}_c = \mathbf{b} - (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}).$$

where **b** is the LSDV estimator. We now seek the restricted estimator of the vector of fixed effects. Write the full coefficient vector as $[\boldsymbol{\alpha}',\boldsymbol{\beta}']'$ and the estimates as $\mathbf{c} = [\mathbf{a}',\mathbf{b}']'$. Also, let $\mathbf{R}_0 = [\mathbf{0},\mathbf{R}]$, where the parts are $J\times N$ and $J\times K$, and J is the number of restrictions. The zero block results from the fact that no restrictions are being imposed on the fixed effects. Thus, in terms of the full coefficient vector, we have $\mathbf{R}_0\mathbf{c} - \mathbf{q} = \mathbf{0}$. Then, in terms of the full coefficient vector, in partitioned form, we have

$$\begin{bmatrix}\mathbf{a}_c\\ \mathbf{b}_c\end{bmatrix} = \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} - \begin{bmatrix}\mathbf{D}'\mathbf{D} & \mathbf{D}'\mathbf{X}\\ \mathbf{X}'\mathbf{D} & \mathbf{X}'\mathbf{X}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{R}'\end{bmatrix}\left[\begin{bmatrix}\mathbf{0}' & \mathbf{R}\end{bmatrix}\begin{bmatrix}\mathbf{D}'\mathbf{D} & \mathbf{D}'\mathbf{X}\\ \mathbf{X}'\mathbf{D} & \mathbf{X}'\mathbf{X}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{R}'\end{bmatrix}\right]^{-1}\left[\begin{bmatrix}\mathbf{0}' & \mathbf{R}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} - \mathbf{q}\right].$$

Using the partitioned inverse formula (Greene (2011, eq. A-66)) produces the result for the fixed effects,

$$\mathbf{a}_c \;=\; \mathbf{a} + (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{X}[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{X}]^{-1}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})$$

$$= \mathbf{a} + \overline{\mathbf{X}}'(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})$$

$$= \mathbf{a} + \overline{\mathbf{X}}'(\mathbf{b} - \mathbf{b}_c)$$

but $\quad \mathbf{a} = \overline{\mathbf{y}} - \overline{\mathbf{X}}'\mathbf{b}$

so $\quad \mathbf{a}_c = \overline{\mathbf{y}} - \overline{\mathbf{X}}'\mathbf{b}_c$

$$= (1/T_i)\Sigma_t\, e_{it} \text{ for each element, } i = 1,...,N.$$

Therefore, the restricted least squares estimators of the fixed effects are the group mean residuals using the restricted least squares estimators of the slopes.

E17.3 Two Way Fixed and Random Effects Models

The panel data estimator also allows ‘two way’ fixed and random effects models. The fixed effects model for a two way design is

$$y_{it} = \alpha_0 + \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

Notice that this model has an overall constant as well as a ‘group’ effect for each group and a ‘time’ effect for each period. The problem of multicollinearity – the time and group dummy variables both sum to one – is avoided by imposing the restrictions $\Sigma_i\alpha_i = \Sigma_t\gamma_t = 0$. (In an unbalanced panel, the sums are weighted by $T_i/(\Sigma_iT_i)$ or $N_t/(\Sigma_tN_t)$.) A full set of estimates is produced for the two factor model in the same fashion as for the one factor model. The random effects model for a two way design is

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i + w_t.$$

The model is described in standard textbooks such as Judge, et. al. (1985) or Greene (2011).

In this model, neither the number of time periods observed for each group nor the number of individuals observed in each period need be fixed. Your data can consist of simply a sample of observations indexed by both individual and time. The data setup is exactly as described in Section R5.3. To request the two factor model, you simply add the specification

; Period = time variable

to the usual command. Unlike a group stratification variable, the time variable must use the integers $1,2,...,T_i$. As noted earlier, *it is not necessary for every group to have data in every period; there may be gaps*. But, if you do have a balanced panel, you can easily set up the time indicator with the Trn function in **CREATE**. For example, in the data set we have been using for our application, there are 17 observations for each state. We could use

```
CREATE      ; time = Trn(-17,0) $
```

(The variable *yr*-1969 in the data set would have the right values.) If the sample is not balanced, in either dimension, it will be necessary to provide the time variable by some other means. When you request the two factor model, the command will appear as

REGRESS ; Lhs = ... ; Rhs = ... ; Panel ; Period = time $

The model is

$$y_{it} = \mu + \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

The textbook formula for two way fixed effects regression, least squares regression of $\left(y_{it} - \overline{y}_i - \overline{y}_t + \overline{\overline{y}}\right)$ on the same transformation of $\mathbf{x}_{it}$ does not work when the panel is unbalanced. It is necessary to add the time dummy variables to a one way fixed effects model (as ordinary regressors). We compute the fixed effects estimator of $\boldsymbol{\beta}$ by making this transformation. The estimated covariance matrix and sum of squares from this least squares regression as computed in the usual manner are appropriate.

NOTE: The two way fixed effects estimator must be computed by literally computing the dummy variables for the time effects. You may have up to 1,000 periods in the data set. You do not need to compute the dummy variables; this is done internally.

E17.3.1 Program Output for Two Factor Models

This estimator produces the full set of output described earlier for the one factor model defined by the **; Pds** setup and an additional set of results for the two factor model. The additional results will be

1. Full set of two factor fixed effects results. Do note, in accordance with the description above, this model, unlike the one way model, will contain an overall constant term. This model is estimated by OLS including both the time and group dummy variables.

2. Full table of estimates of fixed effects (if requested with **; Output = 2**). Note, as well, that the fixed effects produced for the groups will differ from the earlier results, since by design, the time dummy variables are not orthogonal to the group dummy variables.

3. Test statistics for the two way fixed effects model. This consists of the log likelihood, sum of squared deviations, and R^2s for five models:

   a. overall constant term only, no regressors,
   b. group dummies, no regressors,
   c. regressors and overall constant term,
   d. full one way fixed effects model,
   e. full two way fixed effects model.

   You should observe rising log likelihoods and R^2s and falling sums of squares as you go down the table, but if your regressors do not have much explanatory power the reverse could happen between b and c.

4. Full set of results for the two way random effects model including the LM statistic, Hausman statistic, estimates of the variance components, and the usual coefficient estimates with standard errors.

E17.3.2 Application

The following continues the earlier example with the two factor models.

CREATE ; t = Trn(-17,0) $
REGRESS ; Lhs = loggsp ; Rhs = x,one
; Fixed Effects ; Period = t ; Panel $

The first set of results is the same as shown earlier. The results for the two factor models are shown below.

```
----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=LOGGSP   Mean                 =     10.50885
             Standard deviation   =      1.02113
             No. of observations  =          816  Degrees of freedom
Regression   Sum of Squares       =      848.953          69
Residual     Sum of Squares       =      .856334         746
Total        Sum of Squares       =      849.809         815
             Standard error of e  =       .03388
Fit          R-squared            =       .99899  R-bar squared =   .99890
Model test   F[ 69,    746]       =  10704.03809  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =   1640.82592  Akaike I.C.   = -6.68794
             Restricted (b=0)     =  -1174.41748  Bayes  I.C.   = -6.28438
             Chi squared [ 69]    =   5630.48680  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)   =      .751724
--------------------------------------------------
Panel:Groups Empty       0,      Valid data       48
             Smallest   17,      Largest          17
             Average group size in panel       17.00
Panel: Prds: Empty       0,      Valid data       17
             Smallest    0,      Largest          48
             Average group size in panel       48.00
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
   LOGKP|     .48498***      .09316     5.21  .0000      .30239    .66757
  LOGHWY|    -.21822***      .05733    -3.81  .0001    -.33058   -.10586
  LOGH2O|-.55828D-04         .01890      .00  .9976 -.37100D-01  .36988D-01
 LOGUTIL|    -.28424***      .03918    -7.26  .0000    -.36103   -.20745
  LOGEMP|     .92966***      .02114    43.97  .0000      .88822    .97111
Constant|    3.75284***      .25710    14.60  .0000     3.24895   4.25674
--------+-------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------

       Estimated Fixed Effects - Full sets of effects, normalized to sum to 0
       Group        Coefficient       Standard Error       t-ratio
            1          -.08276                .00939      -8.81674
            (Groups 2 – 47 omitted)
           48           .36739                .04277       8.59006
       Estimated Fixed Effects - Full sets of effects, normalized to sum to 0
       Period       Coefficient       Standard Error       t-ratio
            1          -.04440                .00716      -6.20374
            (Periods 2 – 16 omitted)
           17           .05882                .00641       9.17373
```

```
+----------------------------------------------------------------------+
|              Test Statistics for the Regression Model                 |
+----------------------------------------------------------------------+
|        Model              Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only    -1174.41747         849.80888      .00000 |
|(2)  Group effects only      377.47534          18.94145      .97771 |
|(3)  X - variables only      577.89766          11.58975      .98636 |
|(4)  X and group effects    1538.36348           1.10080      .99870 |
|(5)  X ind.&time effects    1640.27938            .85748      .99899 |
+----------------------------------------------------------------------+
|                         Hypothesis Tests                              |
|          Likelihood Ratio Test           F Tests                      |
|          Chi-squared   d.f.   Prob          F    num   denom  P value |
|(2) vs (1)   3103.79     47  .0000    716.77    47     768   .00000 |
|(3) vs (1)   3504.63      5  .0000  11716.51     5     810   .00000 |
|(4) vs (1)   5425.56     52  .0000  11312.81    52     763   .00000 |
|(4) vs (2)   2321.78      5  .0000   2473.18     5     763   .00000 |
|(4) vs (3)   1920.93     47  .0000    154.69    47     763   .00000 |
|(5) vs (4)    203.83     16  .0000     13.25    16     747   .00000 |
|(5) vs (3)   2124.76     64  .0000    146.09    64     747   .00000 |
+----------------------------------------------------------------------+
```

The LM statistic has been adjusted for the two types of effects – there is no Baltagi and Li counterpart for this. The Hausman statistic is also recomputed.

E17.4 Autocorrelation

The one factor fixed and random effects models may be estimated with an autocorrelated error structure. The structural equations would be as follows:

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \varepsilon_{it},$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \eta_{it}.$$

Estimation is done in two steps. In the first, the model is estimated ignoring the autocorrelation just for the purpose of obtaining an estimate of ρ. The second step is the generalized least squares procedure. The following describes the commands needed to estimate this model.

Estimation is essentially the same for both the fixed and random effects models. The first command is used to produce the estimate of ρ and is the basic model command. The estimate is saved automatically in the calculator scalar, *rho*. The second command will be the same as the first, with the addition of the specification

; AR1

You should also use

; Fixed or **; Random**

to specify which type of model you wish to estimate. If you omit both of these, as usual, both the fixed and random effects estimators will be computed, and the random effects model results will be saved.

For example,

REGRESS ; Lhs = y ; Rhs = xlist ; Panel ; Str = i ; Fixed $
REGRESS ; Lhs = y ; Rhs = xlist ; Panel ; Str = i ; Fixed ; AR1 $

will estimate the fixed effects model with the autocorrelated error structure. Changing **; Fixed** to **; Random** will estimate a random effects model. This model may also be combined with the two stage least squares procedure described in the next section. (It cannot be combined with the Hausman-Taylor or DPD estimators described later in this chapter.)

The value of ρ used when you specify **; AR1** is whatever value happens to be in the calculator scalar named *rho*. This is one of the named scalars which automatically contains the residual autocorrelation after you compute a least squares or two stage least squares linear regression. This is kept automatically by the first of the two regressions above. *If your command contains* **; AR1**, *the value of rho is left unchanged by that regression.* Thus, you can use the same value of ρ in several regressions. In addition, the value of *rho* need not be that produced by a fixed or random effects model. If you precede your panel data model with any other regression, it will leave a value of ρ behind to be used by this model. Alternatively, to set ρ, you can simply use the command

CALC ; rho = desired value $

(This scalar is not 'read-only,' as this command demonstrates, even though it appears to be in the project window.)

NOTE: Estimation with autocorrelated disturbances does not require that there be the same number of observations in each group (as usual).

The output produced by the **; AR1** model will differ from the usual output only in the display of the value of *rho* in use at the beginning of the first page. In addition, the output for the LSDV and FGLS estimators will contain estimates of the autocorrelation of the residuals. But, this value will not replace the value of *rho* being used in the calculations. Do note, however, that at every step, the entire analysis is based on transformed data (e.g., $y_{it} - \rho y_{i,t-1}$). As such, many statistics, such as group means, likelihood ratio statistics, analyses of variance, etc., will be meaningless.

In both random and fixed effects models, when **; AR1** is used, the full set of analyses is applied to the transformed data

$$z_{it} = Z_{it} - \rho Z_{i,t-1}$$

where Z_{it} is either y_{it}, $\mathbf{x}_{it}$, or the same transformation of the instruments. This is the Cochrane-Orcutt transformation. As such, the first observation in each group is lost. (The 'within' transformation, i.e., forming deviations from group means, will not remove the heterogeneity if the Prais-Winsten transformation is used for the first observation.) In the fixed effects model, the transformation produces

$$y_{it} - \rho y_{i,t-1} = \boldsymbol{\beta}'(\mathbf{x}_{it} - \rho\mathbf{x}_{i,t-1}) + \alpha_i(1-\rho) + \eta_{it}.$$

Thus, the same fixed effects model applies to the transformed data. The same set of procedures as usual is used to obtain the estimates. In the displayed output, the values of α_i, not $\alpha_i(1-\rho)$ are displayed and kept. But, all other results, including the various variance parameters are based on the transformed data. If predictions are computed, the correct values of the parameters are used to predict y_{it}, not the partial differences. The random effects model produces essentially the same set of complications with **; AR1**. The constant term, α, and the common effect, u_i, are transformed to $\alpha(1-\rho)$ and $u_i(1-\rho)$ during estimation. The constant term is adjusted back in the displayed output. The variance terms estimated using the transformed data are $\sigma_\varepsilon^2(1-\rho)^2$ and $\sigma_u^2(1-\rho)^2$. The final results for the model show these estimates as well as the original parameters, σ_ε^2 and σ_u^2. An application is shown below.

The estimate of ρ from the earlier model, based on the one way fixed effects model, is 0.725563. Using this estimate of ρ, the AR1 model is as shown below. The estimates computed without the autocorrelation correction are shown for both the OLS and LSDV results.

Results based on AR(1) correction

```
Estd. Autocorrelation of e(i,t)   =        .725563
--------------------------------------------------
Panel:Groups Empty      0,     Valid data       48
             Smallest  16,     Largest          16
             Average group size in panel     16.00
Variances    Effects a(i)         Residuals e(i,t)
              .001815                      .000448
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error        z    |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|      .19350          .17937     1.08  .2807     -.15806    .54505
  LOGHWY|    -.19371*          .10929    -1.77  .0763     -.40791    .02049
  LOGH2O|      .00928          .02874      .32  .7469     -.04706    .06561
 LOGUTIL|    -.15953**         .07316    -2.18  .0292     -.30292   -.01614
  LOGEMP|    1.14688***        .02486    46.14  .0000     1.09816   1.19560
--------+-------------------------------------------------------------
```

Original uncorrected results

```
--------------------------------------------------
Panel:Groups Empty      0,     Valid data       48
             Smallest  17,     Largest          17
             Average group size in panel     17.00
Variances    Effects a(i)         Residuals e(i,t)
              .022553                      .001443
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error        z    |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|     .56623***        .10400     5.44  .0000      .36240    .77007
  LOGHWY|    -.23193***        .06325    -3.67  .0002     -.35590   -.10796
  LOGH2O|     .05375***        .01871     2.87  .0041      .01707    .09043
 LOGUTIL|    -.33878***        .04340    -7.81  .0000     -.42385   -.25372
  LOGEMP|    1.00378***        .02010    49.94  .0000      .96439   1.04318
--------+-------------------------------------------------------------
```

The AR1 model for the fixed and random effects specifications is estimated by two step FGLS. In the first step, an estimator of ρ is automatically produced by whatever panel data estimator has been used. This will be any of the fixed or random effects models with one or two way specifications. The estimator is

$$r = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T_i} e_{it}e_{i,t-1} / \left[\sum_{i=1}^{N}(T_i - 1)\right]}{\sum_{i=1}^{N}\sum_{t=2}^{T_i} e_{it}^2 / \left[\sum_{i=1}^{N}(T_i - 1) - K\right]}, \text{ where } e_{it} = (y_{it} - \bar{y}_{i.}) - \hat{\boldsymbol{\beta}}'(\mathbf{x}_{is} - \bar{\mathbf{x}}_{i.}).$$

The estimator of $\boldsymbol{\beta}$ is whatever the most recent one happens to be at the time the calculation is made.

E17.5 Heteroscedasticity and Autocorrelation Robust Covariance Matrix

The cluster corrected robust variance estimator described in Section E17.2.5 accommodates correlation across observations within a group. In principle, the estimator accommodates both ‘autocorrelation,’ that is correlation across observations within the group, and heteroscedasticity, that is, different variances across groups. There are also more narrowly structured robust variance matrix estimators, the White and Newey-West estimators, that accommodate each of these effects alone.

E17.5.1 Heteroscedasticity

If the variances can be assumed to be the same for all observations in the *i*th group, then each group specific variance can be estimated by the group mean squared residual, and the result inserted directly into the textbook formulas for the variance of the OLS (LSDV) estimator. In this case, $\boldsymbol{\Omega}$ becomes a block diagonal matrix, in which the *i*th diagonal block is $\sigma_i^2\mathbf{I}$. (This resembles the time series/cross section model.) In practical terms, we simply replace e_{it}^2 with s_i^2 in the estimate of the asymptotic covariance matrix. To request these estimators, add

; Het or **; Het ; Hc1** or **; Het ; Hc2** or **; Het ; Hc3**
and **; Het = group**

respectively to the **REGRESS** command.

There is a counterpart to the White estimator for unspecified heteroscedasticity for the one way fixed effects model. The model is

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

Suppose that every ε_{it} has a different variance, σ_{it}^2 In the fashion of White’s estimator for the linear model, the natural approach is simply to replace ε_{it}^2 with e_{it}^2 in the preceding, and compute

$$\text{Est.Asy.Var}[\mathbf{b}_{lsdv}] = [\mathbf{X}^{*\prime}\mathbf{X}^*]^{-1}\, X^{*\prime}\, \hat{\boldsymbol{\Omega}}\, X^*\, [\mathbf{X}^{*\prime}\mathbf{X}^*]^{-1}$$

where '*' denotes deviations from group means. This produces the same results as if the White correction were applied to the OLS results in full model including both regressors and group dummy variables. You may also specify the three variations of the White estimator suggested by Davidson and MacKinnon,

$$\textbf{Hc1:}\ \ \text{Est.Var}[\mathbf{b}] = (\mathbf{X^{*\prime}X^*})^{-1} \times \frac{n}{n-K}\sum\nolimits_{i=1}^{n} e_i^2 \mathbf{x}_i{*}\mathbf{x}_i'{*} \times (\mathbf{X^{*\prime}X^*})^{-1}$$

$$\textbf{Hc2:}\ \ \text{Est.Var}[\mathbf{b}] = (\mathbf{X^{*\prime}X^*})^{-1} \times \sum\nolimits_{i=1}^{n} \frac{e_i^2}{\left(1-\mathbf{x}_i'{*}(\mathbf{X^{*\prime}X^*})^{-1}\mathbf{x}_i{*}\right)}\mathbf{x}_i{*}\mathbf{x}_i'{*} \times (\mathbf{X^{*\prime}X^*})^{-1}$$

$$\textbf{Hc3:}\ \ \text{Est.Var}[\mathbf{b}] = (\mathbf{X^{*\prime}X^*})^{-1} \times \sum\nolimits_{i=1}^{n} \frac{e_i^2}{\left(1-\mathbf{x}_i'{*}(\mathbf{X^{*\prime}X^*})^{-1}\mathbf{x}_i{*}\right)^2}\mathbf{x}_i{*}\mathbf{x}_i'{*} \times (\mathbf{X^{*\prime}X^*})^{-1}$$

(In these expressions, 'n' indicates the full sample, which would be $\Sigma_i T_i$ observations.)

LSDV results, uncorrected

```
--------------------------------------------------
Panel:Groups Empty       0,      Valid data       48
             Smallest   17,      Largest          17
             Average group size in panel       17.00
Variances    Effects a(i)          Residuals e(i,t)
              .022553                        .001443
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------------
   LOGKP|     .56623***     .10400     5.44  .0000      .36240     .77007
  LOGHWY|    -.23193***     .06325    -3.67  .0002     -.35590    -.10796
  LOGH2O|     .05375***     .01871     2.87  .0041      .01707     .09043
 LOGUTIL|    -.33878***     .04340    -7.81  .0000     -.42385    -.25372
  LOGEMP|    1.00378***     .02010    49.94  .0000      .96439    1.04318
--------+-------------------------------------------------------------------
```

Robust covariance matrix, unrestricted variances

```
--------------------------------------------------
White/Hetero. corrected covariance matrix was used
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------------
   LOGKP|     .56623***     .10884     5.20  .0000      .35291     .77956
  LOGHWY|    -.23193***     .07033    -3.30  .0010     -.36978    -.09408
  LOGH2O|     .05375***     .01928     2.79  .0053      .01596     .09154
 LOGUTIL|    -.33878***     .04339    -7.81  .0000     -.42383    -.25373
  LOGEMP|    1.00378***     .02546    39.42  .0000      .95388    1.05369
--------+-------------------------------------------------------------------
```

Robust covariance matrix, equal variances within groups

```
--------------------------------------------------
White/Hetero. corrected covariance matrix was used
Disturbance variances assumed equal within groups.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .56623***      .10937     5.18  .0000      .35187    .78060
  LOGHWY|     -.23193***      .06528    -3.55  .0004    -.35988   -.10398
  LOGH2O|      .05375***      .01927     2.79  .0053      .01598    .09152
 LOGUTIL|     -.33878***      .04576    -7.40  .0000    -.42848   -.24909
  LOGEMP|     1.00378***      .02077    48.32  .0000      .96307   1.04450
--------+--------------------------------------------------------------------
```

E17.5.2 Autocorrelation

The asymptotic covariance matrix for the fixed effects estimator may also be estimated with a Newey-West style correction for autocorrelation. Request this computation with

; Lags = the number of lags, up to 10.

Continuing the application from earlier, the Newey-West estimator of the covariance matrix for the LSDV coefficients produce the results below:

(Uncorrected)

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .56623***      .10400     5.44  .0000      .36240    .77007
  LOGHWY|     -.23193***      .06325    -3.67  .0002    -.35590   -.10796
  LOGH2O|      .05375***      .01871     2.87  .0041      .01707    .09043
 LOGUTIL|     -.33878***      .04340    -7.81  .0000    -.42385   -.25372
  LOGEMP|     1.00378***      .02010    49.94  .0000      .96439   1.04318
--------+--------------------------------------------------------------------
Used Newey-West robust VC estimator with   5 lags.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .56623***      .13441     4.21  .0000      .30280    .82966
  LOGHWY|     -.23193***      .08687    -2.67  .0076    -.40219   -.06167
  LOGH2O|      .05375**       .02440     2.20  .0276      .00594    .10156
 LOGUTIL|     -.33878***      .05218    -6.49  .0000    -.44105   -.23652
  LOGEMP|     1.00378***      .03190    31.46  .0000      .94126   1.06631
--------+--------------------------------------------------------------------
```

E18: Random Effects Linear Models for Panel Data

E18.1 Introduction

This chapter will detail estimation of random effects linear models for panel data. The essential structure for most of them is an 'effects' model,

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i + w_t$$

in which variation across groups (individuals) or time is captured in simple shifts of the regression function – i.e., changes in the intercepts. These models are the *random effects* (RE) models characterized by u and w being uncorrelated with $\mathbf{x}$. Under this assumption, the model can be estimated consistently by ordinary least squares. The focus here is on developing efficient estimators or constructing appropriate robust covariance matrices for least squares. Several variations on this structure can be analyzed with this estimator, including both one and two factor models, models of autocorrelation, and several involved, hierarchical models.

E18.2 One Way Random Effects Model

The fundamental part of the random effects model is a one way common effects specification,

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$$

where

$$\text{Cov}(u_i,\mathbf{x}_{it}) = 0 \text{ for all } t,$$

$$E[u_i|\, \mathbf{x}_{it}] = 0,\ \text{Var}[u_i|\, \mathbf{x}_{it}] = \sigma_u^2\ \ \text{Cov}[\varepsilon_{it},u_i|\, \mathbf{x}_{it}] = 0.$$

The random effects model is a generalized regression model. It is homoscedastic, as all disturbances have variance

$$\text{Var}[\varepsilon_{it} + u_i] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2 .$$

But, for a given i, the disturbances in different periods are correlated because of their common component, u_i,

$$\text{Corr}[\varepsilon_{it} + u_i,\varepsilon_{is} + u_i] = \rho = \sigma_u^2 / \sigma^2.$$

The efficient estimator is generalized least squares. *LIMDEP* provides a two step procedure and a maximum likelihood estimator under the additional assumption that ε is normally distributed. The variance components are first estimated by using the residuals from ordinary least squares regressions. Then, feasible GLS estimates are computed using the estimated variances.

An additional procedure is available to fit the model by maximum likelihood assuming normally distributed disturbances. The resulting estimator has the same properties as the FGLS estimator, so this is not a basis to choose it. But, two additional extensions of the model, exponential heteroscedasticity and nested random effects with unbalanced panels, are fairly easily handled by MLE, but are not feasible (logistically) using FGLS. These specifications trade the possibly narrow assumption of normality for the increased flexibility of the broader models. Obviously, the choice is up to the user in the context of a given application. There is no test for the specification.

E18.2.1 Command

The commands for estimation of these models are variants of the basic structure

SETPANEL ; Group = identifier ; Pds = variable name

Then, **REGRESS ; Lhs = y ; Rhs = x...**
; Panel
; Random Effects
; ... other options $

As always, you may use

; Panel
; Str = the name of a stratification variable
or **; Pds = specification of the number of periods, variable or fixed**

in the command to specify the panel instead.

The random effects model automatically includes a constant term, whether you have included *one* or not. If you want the random effects model to be fit *without a constant term*, include

; No constant

A crucial element of the computation of the random effects model is the estimation of the variance components. You may supply your own values for σ_ε^2 and σ_u^2. The specification is

; Var = s2e,s2u for the one factor model
; Var = s2e,s2u,s2w for the two factor model

This overrides all other computations. The values are checked for validity. A nonpositive value forces estimation to halt at that point.

NOTE: If you omit the **; Random Effects** part of the command, then *LIMDEP* reports full results for the pooled regression (as always) and the fixed effects (LSDV) regression in addition to the random effects results. By including **; Random Effects**, you will suppress the display of the LSDV results (though they are still computed internally).

E18.2.2 Output

After display of any previous results, including ordinary least squares and the fixed effects estimator, a display such as the following will be presented, followed by the standard form table of coefficient estimates, standard errors, etc. The results in the table are as follows:

REGRESS ; Lhs = loggsp
; Rhs = x,one
; Panel ; Random Effects $

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .001443
            Var[u]              =       .012866
            Corr[v(i,t),v(i,s)] =       .899169
            Sum of Squares            17.866362
            R-squared                   .978976
Fixed vs. Random Effects (Hausman)     =  35.24
( 5 degrees of freedom, prob. value =  .000001)
(High (low) values of H favor F.E.(R.E.) model).
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|   Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .61652***      .10156     6.07  .0000      .41747     .81558
  LOGHWY|    -.25482***      .05782    -4.41  .0000     -.36814    -.14150
  LOGH2O|      .05042***      .01838     2.74  .0061      .01439     .08644
 LOGUTIL|    -.35501***      .04262    -8.33  .0000     -.43854    -.27147
  LOGEMP|      .98293***      .01962    50.10  .0000      .94448    1.02138
Constant|     2.66833***      .16389    16.28  .0000     2.34711    2.98956
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

1. Estimates of σ_ε^2 and σ_u^2 based on the least squares dummy variable model residuals. These are used to estimate the variance components. The technical details in Section E18.2.5 describe the computations. Since there are some potential problems that can arise, the sequence of steps taken in this part is documented in the trace file. The application shows an example. This trace output may be quite lengthy, as several attempts may be made to fit the model with different variance components estimators.

2. The estimate of $\rho = \sigma_u^2 / (\sigma_\varepsilon^2 + \sigma_u^2)$ based on whatever first round estimator has been used.

3. The sum of squares is the sum of squared residuals based on the two step FGLS coefficient vector.

4. An R^2 measure is reported (by popular request)

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - \hat{\boldsymbol{\beta}}'_{RE}\mathbf{x}_{it})^2}{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - \bar{y})^2}.$$

   Users are warned, this measure can be negative. It is only guaranteed to be positive when OLS has been used to fit a model with a constant term. There are other measures that could be computed, such as the squared correlation between the actual and fitted values, but neither these, nor the one above, are fit measures in the same sense as in the linear model. It will always be less than the result for OLS (since OLS is LS).

5. The Hausman specification test for fixed vs. random effects if presented at the end of these results. See Section E18.2.3 for discussion.

NOTE: In computing the random effects model, the second step FGLS estimator generally relies on the first step OLS and LSDV (fixed effects) sums of squares. You may be suppressing the FE model, perhaps because of the presence of time invariant variables which preclude the FE model, but not the RE model. In previous versions of *LIMDEP*, and in some other programs, this will force the estimator to rely on another device to estimate the variance components, typically a group means estimator. In the current version of *LIMDEP*, the FE model is computed in the background, whether reported or not. The sums of squares needed are obtainable even in the presence of time invariant variables. Thus, you will get the same results for the RE model whether or not you have allowed *LIMDEP* to report the fixed effects results.

The standard table of coefficient results follows. The test statistic is denoted ‘z’ as the asymptotic normal distribution applies, rather than the finite sample t distribution.

E18.2.3 Specification Tests for Random vs. Fixed Effects

Hausman’s chi squared statistic for testing the REM against the FEM is

$$H = (\hat{\beta}_{LSDV} - \hat{\beta}_{RE})' \, [Est.Var(\hat{\beta}_{LSDV}) - Est.Var(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{LSDV} - \hat{\beta}_{RE})$$

The Hausman statistic for the specification test of fixed vs. random effects is also reported, as shown below:

REGRESS ; Lhs = loggsp ; Rhs = x,one ; Panel $

```
----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .001443
            Var[u]              =       .012866
            Corr[v(i,t),v(i,s)] =       .899169
            Sum of Squares            17.866362
            R-squared                   .978976
----------------------------------------------
Fixed vs. Random Effects (Hausman)     =  35.24
( 5 degrees of freedom, prob. value =  .000001)
(High (low) values of H favor F.E.(R.E.) model).
--------+-------------------------------------------------------------------
(Other results omitted)
```

The prob value and degrees of freedom for the Hausman statistic are reported.

HINT: Large values of the Hausman statistic argue in favor of the fixed effects model over the random effects model. Large values of the LM statistic argue in favor of one of the one factor models against the classical regression with no group specific effects. A large value of the LM statistic in the presence of a large Hausman statistic (as in our application) argues in favor of the fixed effects model.

NOTE: Sometimes it is not possible to compute the Hausman statistic. The difference matrix in the formula above may not be positive definite. The theory does not guarantee this. It is more likely to be so, but still not certain, if the same estimate of σ_ε^2 is used for both cases. As such, *LIMDEP* uses the FGLS estimator of this, however it has been obtained, for the computation. Still, the matrix may fail to be positive definite. (The program will issue an error message,

```
Error   425: REGR;PANEL. Could not invert VC matrix for Hausman test
```

when this occurs. In this case, a 0.00 is reported for the statistic and a diagnostic warning appears in the results. Users are warned, some other programs attempt to bypass this issue by using some other matrix or some other device to force a positive statistic. These ad hoc measures do not solve the problem – they merely mask it. At worst, the appropriate zero value can be replaced by a value that appears to be 'significant.' The better strategy in such a case is to take the difference between the two estimators to be random variation, which would favor the random effects estimator. The Wu variable addition test is also a useful alternative approach.

Wu's (1973) variable addition test is an alternative approach to computing the Hausman test for random vs. fixed effects. The test is carried out by adding the group means of the time varying variables to the random effects model then testing the joint hypothesis that the coefficients on the group means are all zero. (See Baltagi (2008) for details.) The test is not built in (since the program cannot tell from a variable list what variables are group means. But, it is straightforward to layer onto the random effects estimator. The following shows how to do so using our earlier example. The five group means are time invariant variables, which can be seen in the LSDV results. The results shown for these variables should be ignored.

```
CREATE    ; lpcbar  = Group Mean(logkp, Pds = 17) $
CREATE    ; lhwybar = Group Mean(loghwy, Pds = 17) $
CREATE    ; lh2obar = Group Mean(logh2o, Pds = 17) $
CREATE    ; lutilbar = Group Mean(logutil, Pds = 17) $
CREATE    ; lempbar = Group Mean(logemp, Pds = 17) $
REGRESS   ; Lhs = loggsp
          ; Rhs = x,lpcbar,lhwybar,lh2obar,lutilbar,lempbar
          ; Panel
          ; Test: lpcbar = 0, lhwybar = 0, lh2obar = 0, lutilbar = 0, lempbar = 0 $
```

The ordinary least squares regression results and the LSDV least squares with fixed effects results are omitted.

```
--------------------------------------------------
Panel:Groups Empty      0,      Valid data       48
             Smallest  17,      Largest          17
             Average group size in panel      17.00
Variances    Effects a(i)         Residuals e(i,t)
              .280528                   .001452
These  5 variables have no within group variation.
LPCBAR   LHWYBAR  LH2OBAR  LUTILBAR LEMPBAR
F.E. estimates are based on a generalized inverse.
--------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .56623***      .10434     5.43  .0000      .36173    .77074
  LOGHWY|    -.23193***      .06346    -3.65  .0003     -.35631   -.10755
  LOGH2O|     .05375***      .01878     2.86  .0042      .01695    .09055
 LOGUTIL|    -.33878***      .04354    -7.78  .0000     -.42413   -.25344
  LOGEMP|    1.00378***      .02017    49.77  .0000      .96426   1.04331
  LPCBAR|        0.0    .....(Fixed Parameter).....
 LHWYBAR|        0.0    .....(Fixed Parameter).....
 LH2OBAR|        0.0    .....(Fixed Parameter).....
LUTILBAR|        0.0    .....(Fixed Parameter).....
 LEMPBAR|        0.0    .....(Fixed Parameter).....
 --------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
(Results omitted)

Error   425: REGR;PANEL. Could not invert VC matrix for Hausman test.

-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .001452
            Var[u]              =       .011788
            Corr[v(i,t),v(i,s)] =       .890317
            Sum of Squares            10.658532
            R-squared                   .987458
---------------------------------------------
Wald test of  5 linear restrictions
Chi-squared =      38.17, P value   =    .00000
Test:      F ratio [ 5, 758]        =   7.63365  Prof F  > F*  =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .56623***      .10434     5.43  .0000      .36173    .77074
  LOGHWY|    -.23193***      .06346    -3.65  .0003     -.35631   -.10755
  LOGH2O|     .05375***      .01878     2.86  .0042      .01695    .09055
 LOGUTIL|    -.33878***      .04354    -7.78  .0000     -.42413   -.25344
  LOGEMP|    1.00378***      .02017    49.77  .0000      .96426   1.04331
  LPCBAR|    -.42125         .67847     -.62  .5347    -1.75103    .90853
 LHWYBAR|     .51236         .36396     1.41  .1592     -.20099   1.22572
 LH2OBAR|     .06305         .10529      .60  .5493     -.14330    .26941
LUTILBAR|     .30434         .29238     1.04  .2979     -.26872    .87740
 LEMPBAR|    -.44919***      .09559    -4.70  .0000     -.63654   -.26184
Constant|    2.15254***      .64061     3.36  .0008      .89698   3.40811
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E18.2.4 Saved Results

Results which are saved for later use are:

Matrices: *b* and *varb* These will be the FGLS estimates of the random effects model.

Scalars:

ssqrd	=	s^2 from least squares dummy variable (LSDV) estimator or from FGLS
rsqrd	=	R^2 from LSDV
s	=	$\sqrt{s^2}$ from LSDV
sumsqdev	=	sum of squared residuals from LSDV
rho	=	estimated disturbance autocorrelation from whatever model is fit last
degfrdm	=	$\Sigma_i T_i - K$
sy	=	standard deviation of Lhs variable
ybar	=	mean of Lhs variable
kreg	=	K
nreg	=	total number observations
logl	=	log likelihood from LSDV model
ssqrdu	=	estimate of σ_u^2 from FGLS
ssqrde	=	estimate of σ_ε^2 from FGLS
ssqrdw	=	estimate of σ_w^2 from GLS if two way random effects model is fit
exitcode	=	0.0 if the model was estimable
ngroup	=	number of groups
nperiod	=	number of periods. This will be 0.0 if you fit a one way model.

Last Model: *b_variable* constructed as usual.

Last Function: Conditional mean for the linear regression = a + **b'x**

Predicted values are based on the last model estimated, one or two way, fixed or random. Since the constant term is included in the function, **SIMULATE** will give appropriate predictions. **PARTIALS** operates as usual.

E18.2.5 Technical Details

The equal group sizes case is considered first. Special considerations for the unbalanced panel case are considered next. The random effects model is a generalized regression model. The specification has

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it},$$

$$v_{it} = \varepsilon_{it} + u_i,$$

$$E[v_{it}|\mathbf{X}] = 0,$$

$$E[v_{it}^2|\mathbf{X}] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,$$

$$E[v_{it}\, v_{is}|\mathbf{X}] = \sigma_u^2,$$

and $$E[v_{it}\, v_{js}|\mathbf{X}] = 0 \;\forall\; t,s \text{ if } i \neq j.$$

Estimation by feasible GLS (FGLS) is done by regressing $y_{it} - \theta\,\overline{y}_i$ on (1 - θ) and $(\mathbf{x}_{it} - \theta\,\overline{\mathbf{x}}_i)$. (1 - θ) is the constant term in this regression, where

$$\theta = 1 - \sigma_\varepsilon / \sigma_2$$

and

$$\sigma_2^2 = \sigma_\varepsilon^2 + T\sigma_u^2 .$$

Since the variances are unknown, they and θ must be estimated first. This is done as follows:

1. The residual variance from the LSDV estimator is a consistent estimator of σ_ε^2;

$$\hat{\sigma}_\varepsilon^2 = \Sigma_i\Sigma_t\, e_{LSDV,it}^{\;2} / (NT - N - K),$$

where $e_{LSDV,it}$ is the residual from the least squares dummy variable regression and '*NT*' denotes the entire sample size,

$$e_{it} = y_{it} - a_i - \mathbf{b}_{LSDV}'\mathbf{x}_{it} = (y_{it} - \overline{y}_i) - \mathbf{b}_{LSDV}'(\mathbf{x}_{it} - \overline{\mathbf{x}}_i).$$

(*The fixed effects estimator is always computed for this purpose, even if the results are not displayed.*)

2. The simple least squares estimator with no group effects can always be computed. The residual variance estimator from this procedure would be

$$s^2 = \Sigma_i\Sigma_t\, e_{OLS,it}^{\;2} / (NT - K - 1)$$

This is a consistent estimator of $\sigma_\varepsilon^2 + \sigma_u^2$, so a consistent estimator of σ_u^2 is

$$\hat{\sigma}_u^2 = s^2 - \hat{\sigma}_\varepsilon^2$$

3. This second estimate need not be positive, because of the differing degrees of freedom. In this event, a second attempt is made. If the degrees of freedom correction is not made, then by construction, both variance estimators must be positive, and estimation proceeds. The LSDV estimator must fit better than the model with only a single constant, so

$$\hat{\sigma}_u^2 = \Sigma_i\Sigma_t\, e_{OLS,it}^{\;2} / NT - \Sigma_i\Sigma_t\, e_{LSDV,it}^{\;2} / NT$$

must be positive, as will $\hat{\sigma}_\varepsilon^2 = \Sigma_i\Sigma_t\, e_{LSDV,it}^{\;2} / NT$.

4. If neither works – $\hat{\sigma}_u^2$ can be zero, though (at least in theory) not negative – then we try using the group means. (This should never happen, but we note this procedure to connect to the existing literature and to what is done in other software. The capability remains in *LIMDEP*.) In the REM,

$$\overline{y}_i = \alpha + \boldsymbol{\beta}'\overline{\mathbf{x}}_i + u_i + \overline{\varepsilon}_i$$

so

$$y_{it} - \overline{y}_i = \boldsymbol{\beta}'(\mathbf{x}_{it} - \overline{\mathbf{x}}_i) + \varepsilon_{it} - \overline{\varepsilon}_i .$$

Hence, the mean squared residual from this regression is a natural estimator of σ_ε^2. It can be shown that this one is unbiased. Of course, it might be based on a small number of observations. To estimate σ_u^2, we note that for the group means, the regression is a classical regression with disturbance variance

$$\text{Var}[u_i + \bar{\varepsilon}_{i\,.}] = \sigma_u^2 + \sigma_\varepsilon^2 / T.$$

Therefore, if we regress the group means of y on a constant and the group means of $\mathbf{x}$, the variance estimator in this regression is an unbiased estimator of

$$\sigma_1^2 = \sigma_u^2 + \sigma_\varepsilon^2 / T.$$

Since we have an estimator of σ_ε^2 in hand, we can use

$$\hat{\sigma}_u^2 = \mathbf{v}^{*\prime}\mathbf{v}^{*}/(N\text{-}K) - (1/T)\ \hat{\sigma}_\varepsilon^2$$

where $$v_i^* = \bar{y}_i - \mathbf{b}_{group\ means}{}'\,\bar{\mathbf{x}}_i$$

This is unbiased in the general case. Unfortunately, this estimator may also not be positive. An alternative can be based on the direct least squares estimates of Model 3 (or, for that matter, any other consistent estimator of $\boldsymbol{\beta}$). Using the same calculation otherwise, we would just compute

$$s_1^2 = \Sigma_\text{i}\,(\bar{y}_i - \mathbf{b}_{ols}{}'\,\bar{\mathbf{x}}_i)^2/(\text{N-K})$$

then $$\hat{\sigma}_u^2 = s_1^2 - (1/\text{T})\ \hat{\sigma}_\varepsilon^2\ .$$

If need be, *LIMDEP* tries all of these estimators. As noted, the second attempt, using the sums of squares without degrees of freedom corrections, will succeed in all but the most pathological cases. Still, it is possible that none of the procedures will produce a positive estimate of σ_u^2. In that instance, estimation is halted. The search is reported in the trace file when you fit this model, along with the decisions made at each point as the program seeks a valid estimate. The entry for the model fit in the previous section is as follows:

```
Regress ; Lhs = loggsp ; Rhs = x,one
;panel;pds=17;random effects$
Estimating variance components for random effects model.
FEM was computed. Using LSDV and OLS to get V[e] & V[u].
OLS & LSDV with d.f. provide both estimates > 0.
Exit status for this model command is   .0.
```

NOTE: The finding of a nonpositive estimate of σ_u^2 is quite common, and many programs do not use the same search we do. Notably, many do not use the second pass attempt (without degrees of freedom correction) that is used here. This leaves the negative estimate as a persistent possibility. Users should be aware of what the software does in this instance.

```
Regress ; Lhs = loggsp ; Rhs = x,one ;panel;pds=17 ; mle; hfu=ubari
Estimating variance components for random effects model.
FEM was computed. Using LSDV and OLS to get V[e] & V[u].
OLS & LSDV with d.f. provide both estimates > 0.
Entering iterative search for function optimizers.
Begin main iterations for optimization.
Maximum iterations reached. Exit iterations with status=1.
Exit from iterative procedure.  501 iterations completed.
Exit status appears above.
Exit status for this model command is  1.0.
```

Unequal Group Sizes

We consider the case of unequal group sizes. In the group means regression, the disturbance will now have variance

$$\mathrm{Var}[u_i + \Sigma_i \varepsilon_{it}/T_i] \;=\; \sigma_{1i}^2 \;=\; \sigma_u^2 + \sigma_\varepsilon^2/T_i.$$

Therefore, the group means regression is heteroscedastic. The unbiasedness result above does not hold any more. However, it can be shown that the ordinary least squares variance estimator in a heteroscedastic regression is a consistent estimator of

(*) $$\overline{\sigma}_1^2 \;=\; \mathrm{plim}(1/N)\Sigma_{i=1}^N \sigma_{1i}^2 ,$$

assuming that the probability limit exists. As will be useful later, the mean squared residual (using group means of y and $\mathbf{x}$) based on *any* consistent slope estimator is a consistent estimator of $\overline{\sigma}_1^2$. In this setting, we take the limit as applying to N increasing, *not* T or T_i. T or T_i is taken as fixed in this model and may not increase at all beyond a very small number. Consistency results depend on increasing N, not T or T_i. So, the variance estimator in the group means regression is a consistent estimator of

$$\begin{aligned}\overline{\sigma}_1^2 \;&=\; \sigma_u^2 \;+\; \sigma_\varepsilon^2\,\mathrm{plim}(1/N)\Sigma_i(1/T_i)\\ &=\; \sigma_u^2 \;+\; \sigma_\varepsilon^2\,\mathrm{plim}Q_N^*\\ &=\; \sigma_u^2 \;+\; \sigma_\varepsilon^2\,Q^*\end{aligned}$$

whatever that happens to be. Some assumption about the group sizes is obviously necessary. One possibility would be to assume that T_i is randomly distributed across individuals with $\mathrm{E}[T_i] = T$. Note that if $T_i = T$ for all i, then $Q_N^* = Q^* = Q = 1/T$. Suppose it is assumed that Q_N^* converges to some well defined Q^*. Then, in our sample, the statistic

$$Q \;=\; (1/N)\,(1/T_1 \;+\; 1/T_2 \;+\; \ldots \;+\; 1/T_N)$$

is a consistent estimator of Q^*, and the estimator

$$s_u^{2*} \;=\; \mathbf{v}^{*\prime}\mathbf{v}^* \,/\, (N\text{-}K) \text{ - } Q\,\hat{\sigma}_\varepsilon^{\,2}$$

is a consistent estimator as well. (If the group sizes are equal, equation (*) above emerges.) The degrees of freedom correction (N-K) instead of N is unnecessary, of course. We make it so the appropriate value will result in the equal sized groups case.

Heteroscedasticity in the Random Effects Model Due to Unequal Group Sizes

Estimation by FGLS is done by regressing

$$y_{it} - \theta_i\, \bar{y}_i \;\text{ on }\; (1 - \theta_i) \;\text{ and }\; (\mathbf{x}_{it} - \theta_i\, \bar{\mathbf{x}}_i).$$

($1 - \theta_i$ is the constant term in this regression), where

$$\theta_i = 1 - \sigma_\varepsilon / \sigma_{2i}$$

and

$$\sigma_{2i}^2 = \sigma_\varepsilon^2 + T_i\,\sigma_u^2.$$

(Note that the weights are already unequal if the group sizes vary, regardless of the heteroscedasticity.) Neglecting the heteroscedasticity and the unequal group sizes for the moment, the first step in the regression is computation of the variance components, which we do as follows:

1. Using simple OLS, use $\mathbf{e}_o'\mathbf{e}_o/NT$ to estimate $\sigma_\varepsilon^2 + \sigma_u^2$.
2. Using LSDV, use $\mathbf{e}'\mathbf{e}/NT$ to estimate σ_ε^2.
3. Estimate σ_u^2 with $\mathbf{e}_o'\mathbf{e}_o/NT - \mathbf{e}'\mathbf{e}/NT$.

Once again, 'NT' symbolizes the full sample size, ($\Sigma_i T_i$). Under homoscedasticity, both estimators are consistent.

Suppose, now, that T_i differs across groups and, as well, that so does Var[ε_{it}]. We consider estimation of σ_u^2. This estimator is, by construction

$$\hat{\boldsymbol{\sigma}}_u^2 = \sum_{i=1}^N \left(\frac{T_i}{\sum_{i=1}^N T_i}\right)\left[\frac{\mathbf{e}_{oi}'\mathbf{e}_{oi}}{T_i}\right] - \sum_{i=1}^N \left(\frac{T_i}{\sum_{i=1}^N T_i}\right)\left[\frac{\mathbf{e}_i'\mathbf{e}_i}{T_i}\right]$$

Note that this collects the sums of squares by groups, and multiplies and divides each contribution by the respective sample size.) We can write this as

$$\hat{\boldsymbol{\sigma}}_u^2 = \sum_{i=1}^N w_i\, \hat{\boldsymbol{\sigma}}_{ui}^2 \quad\text{where } w_i = T_i / (\Sigma_i T_i).$$

That is, it is an unequally weighted (unless T_i is fixed) average of N separate estimators of σ_u^2. To what does this estimator converge? The estimator can be written

$$\hat{\boldsymbol{\sigma}}_u^2 = \sum_{i=1}^N w_i \left(\widehat{\boldsymbol{\sigma}_{\varepsilon,i}^2 + \boldsymbol{\sigma}_u^2}\right) - \sum_{i=1}^N w_i\, \hat{\boldsymbol{\sigma}}_{\varepsilon,i}^2 .$$

Since each estimator within the brackets is based on the fixed and possibly small T_i observation, one cannot say that either is consistent. But, as N grows, the average in this dimension will be consistent under some fairly benign assumptions that would make an average of estimators of $\sigma_{\varepsilon,i}^2$ converge to some 'average variance,' $\bar{\sigma}_\varepsilon^2$. If so, then the first of these will converge to $\bar{\sigma}_\varepsilon^2 + \sigma_u^2$ while the second will converge to $\bar{\sigma}_\varepsilon^2$. (These are different estimators, but they should converge to the same thing.) If so, then the difference converges to σ_u^2. By this development, we use the original estimator of σ_u^2 for the FGLS estimator in this model. We then compute the heteroscedastic random effects estimator by recomputing the estimator of σ_ε^2 within each group. The group specific estimator, whether defined by the original data groups or by a higher level stratification, is obtained by the mean squared LSDV residual for that particular group. (This is the second term in square brackets in the earlier expression.)

E18.2.6 Robust Covariance Matrix

Since the random effects model is fit using two step GLS, it assumes a particular disturbance process. As such, a 'robust' covariance matrix would seem counterproductive, or at least contradictory. Nonetheless, just such an approach has been advocated in the recent literature. By adding

; Robust

to the command, you can request *LIMDEP* to abandon the FGLS covariance matrix, and use the cluster estimator shown in Section E17.2.5. For the random effects model, the estimator is

$$Est.Asy.Var[b_{FGLS}] = (\mathbf{X'X})^{-1}\left[A\sum\nolimits_{i=1}^{n}\left\{\left(\mathbf{X}^{(i)}\right)\mathbf{e}_i\right\}\left\{\left(\mathbf{X}^{(i)}\right)\mathbf{e}_i\right\}'\right](\mathbf{X'X})^{-1}$$

$$A = \frac{n}{n-1}\frac{(\Sigma_t T_i)-1}{(\Sigma_t T_i)-(n+K)},\; e_{it} = y_{it} - \hat{\boldsymbol{\beta}}'_{FGLS}\mathbf{x}_{it}$$

Results applied to our earlier example are shown below. The differences in the estimated standard errors are quite stark. Under the logic that the FGLS estimator is consistent regardless of the true underlying structure (as is OLS), we might conclude that the one way random effect model is misspecified – though this is not a formal test of that proposition.

```
--------------------------------------------------------------------------------
Random Effects Model: v(i,t)     = e(i,t) + u(i)
Estimates:  Var[e]               =        .001443
            Var[u]               =        .012866
            Corr[v(i,t),v(i,s)] =        .899169
            Sum of Squares              17.866362
            R-squared                     .978976
Robust cluster corrected covariance matrix used
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
   LOGKP|     .61652       .88256     .70  .4848    -1.11327    2.34631
  LOGHWY|    -.25482       .42253    -.60  .5465    -1.08296     .57332
  LOGH2O|     .05042       .10427     .48  .6287     -.15395     .25478
 LOGUTIL|    -.35501       .38459    -.92  .3560    -1.10880     .39879
  LOGEMP|     .98293***    .12082    8.14  .0000      .74612    1.21974
Constant|    2.66833***    .88133    3.03  .0025      .94096    4.39571
--------+-----------------------------------------------------------------------
(Uncorrected Covariance Matrix)
--------+-----------------------------------------------------------------------
   LOGKP|     .61652***    .10156    6.07  .0000      .41747     .81558
  LOGHWY|    -.25482***    .05782   -4.41  .0000     -.36814    -.14150
  LOGH2O|     .05042***    .01838    2.74  .0061      .01439     .08644
 LOGUTIL|    -.35501***    .04262   -8.33  .0000     -.43854    -.27147
  LOGEMP|     .98293***    .01962   50.10  .0000      .94448    1.02138
Constant|    2.66833***    .16389   16.28  .0000     2.34711    2.98956
--------+-----------------------------------------------------------------------
```

E18.3 ML Estimation of One Way Random Effects Models

The one way random effects linear model with normally distributed disturbances can be fit using maximum likelihood rather than two step FGLS. As always, the estimator allows unbalanced panels. The estimator is requested simply by adding

; MLE

to the basic command for the random effects model. The request has no impact on the fixed effects estimator or on the FGLS estimator. The full model is fit as usual, then an additional set of results are provided for the MLE.

Three other forms of random effects linear models can be fit by maximum likelihood. After extending the basic model to an MLE, we describe a formal model for heteroscedastic disturbances. A further extension of the model provides for nested random effects up to three levels. Finally, Section E18.8 presents a model with multiple way random effects that is estimated by simulation, rather than by FGLS.

E18.3.1 Application

To illustrate the estimator, we recompute the estimates for the random effects model shown in Section E18.2.2 using the maximum likelihood estimator. Results for this model are shown below. The FGLS estimates will always appear first. The maximum likelihood estimates will follow. We note, the ML results include estimates of the slope functions and of $\theta = 1/\sigma_\varepsilon^2$ and $\tau = \sigma_u^2/\sigma_\varepsilon^2$. *LIMDEP* reparameterizes the log likelihood function for purpose of estimation. The estimates of the two underlying variance parameters are derived from these, and appear in the box above the coefficient estimates. Further details appear in the technical notes below.

```
Normal exit:   6 iterations. Status=0, F=   -1380.630

-----------------------------------------------------------------------------
Random Effects Linear Regression Model
Dependent variable                LOGGSP
Log likelihood function       1380.62989
Restricted log likelihood      577.89766
Chi squared [   1 d.f.]       1605.46447
Significance level                .00000
McFadden Pseudo R-squared     -1.3890561
Estimation based on N =    816, K =   8
Inf.Cr.AIC  =-2745.260 AIC/N =   -3.364
Variance of e(i,t) =             .001435
Variance of u(i   ) =            .020935
Corr[v(i,t),v(i,s)]=             .935835
LR test is vs. null of no random effect
Panel contained      48 nonempty groups
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  LOGGSP|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .60013***      .10227     5.87  .0000       .39969    .80056
  LOGHWY|     -.24974***      .05918    -4.22  .0000     -.36573   -.13376
  LOGH2O|      .05145***      .01846     2.79  .0053       .01527    .08763
 LOGUTIL|     -.34962***      .04283    -8.16  .0000     -.43358   -.26567
  LOGEMP|      .99069***      .01993    49.70  .0000       .95162   1.02976
Constant|     2.67282***      .17388    15.37  .0000      2.33202   3.01362
        |Reparameterized Variance Components for ML Search
   1/s2e|     696.673***    35.59769    19.57  .0000      626.903   766.443
 s2u/s2e|     14.5847***     3.14783     4.63  .0000       8.4151   20.7544
--------+--------------------------------------------------------------------
(Two Step Feasible GLS Estimates)
Estimates:  Var[e]             =         .001443
            Var[u]             =         .012866
--------+--------------------------------------------------------------------
   LOGKP|      .61652***      .10156     6.07  .0000       .41747    .81558
  LOGHWY|     -.25482***      .05782    -4.41  .0000     -.36814   -.14150
  LOGH2O|      .05042***      .01838     2.74  .0061       .01439    .08644
 LOGUTIL|     -.35501***      .04262    -8.33  .0000     -.43854   -.27147
  LOGEMP|      .98293***      .01962    50.10  .0000       .94448   1.02138
Constant|     2.66833***      .16389    16.28  .0000      2.34711   2.98956
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E18.3.2 Technical Notes on ML Estimation of the Random Effects Model

The contribution of the ith individual to the log likelihood for the random effects model with normally distributed disturbances is

$$\log L_i(\boldsymbol{\beta},\sigma_\varepsilon^2,\sigma_u^2) = \frac{-1}{2}\left[T_i \log 2\pi + \log|\boldsymbol{\Omega}_i| + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]$$

$$= \frac{-1}{2}\left[T_i \log 2\pi + \log|\boldsymbol{\Omega}_i| + \boldsymbol{\varepsilon}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{\varepsilon}_i\right]$$

where $\boldsymbol{\Omega}_i = \sigma_\varepsilon^2 \mathrm{I}_{\mathrm{Ti}} + \sigma_u^2 \mathbf{ii}'$

Note that the $\boldsymbol{\Omega}_i$ varies over i because it is $T_i \times T_i$. By expanding $\boldsymbol{\Omega}_i$ in the expression, and with some straightforward algebra, we obtain

$$\log L = \Sigma_{i=1}^N \log L_i$$

$$= -\frac{-1}{2}[(\log 2\pi + \log\sigma_\varepsilon^2)\Sigma_{i=1}^N T_i + \Sigma_{i=1}^N \log(1+T_i\rho^2)] - \frac{1}{2\sigma_\varepsilon^2}\Sigma_{i=1}^N\left[\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i - \frac{\sigma_\varepsilon^2(T_i\bar{\varepsilon}_i)^2}{\sigma_\varepsilon^2 + T_i\sigma_u^2}\right]$$

where $\rho = \sigma_u^2 / (\sigma_\varepsilon^2 + \sigma_u^2)$.

With some further transformations,

$$\theta \;=\; 1/\sigma_\varepsilon^2,\ \gamma = \sigma_u^2/\sigma_\varepsilon^2,\ R_i = T_i\gamma + 1,\ Q_i = \gamma / R_i,$$

the individual contribution to the log likelihood becomes

$$\log L_i \;=\; -(1/2)[\theta(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i - Q_i(T_i\bar{\varepsilon}_i)^2) + \log R_i + T_i \log\theta + T_i \log 2\pi]\,.$$

We maximize the log likelihood function using *LIMDEP*'s general optimization program – BFGS is the default algorithm. The derivatives of the terms, which are summed over *i* to give the totals are

$$\frac{\partial \log L_i}{\partial \boldsymbol{\beta}} = \theta\left[\sum_{t=1}^{T_i} \mathbf{x}_{it}\varepsilon_{it}\right] - \theta\left[\frac{\gamma}{T_i\gamma+1}\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}\right)\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)\right],$$

$$\frac{\partial \log L_i}{\partial \theta} = -\frac{1}{2}\left[\left(\sum_{t=1}^{T_i}\varepsilon_{it}^2\right) - \frac{\gamma}{T_i\gamma+1}\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)^2 - \frac{T_i}{\theta}\right],$$

$$\frac{\partial \log L_i}{\partial \gamma} = \frac{1}{2}\left[\theta\left(\frac{1}{(T_i\gamma+1)^2}\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)^2\right) - \frac{T_i}{T_i\gamma+1}\right],$$

$$-\frac{\partial^2 \log L_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \left\{\left[\sum_{t=1}^{T}\theta\mathbf{x}_{it}\mathbf{x}_{it}'\right] - \frac{\theta\gamma}{T_i\gamma+1}\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}\right)\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}'\right)\right\},$$

$$-\frac{\partial^2 \log L_i}{\partial \boldsymbol{\beta}\partial \theta} = -\left\{\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}\varepsilon_{it}\right) - \frac{\gamma}{T_i\gamma+1}\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}\right)\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)\right\},$$

$$-\frac{\partial^2 \log L_i}{\partial \boldsymbol{\beta}\partial \gamma} = \left\{\frac{\theta}{(T_i\gamma+1)^2}\left(\sum_{t=1}^{T_i}\mathbf{x}_{it}\right)\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)\right\},$$

$$-\frac{\partial^2 \log L_i}{\partial \theta\partial \gamma} = -\frac{1}{2(T_i\gamma+1)^2}\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)^2,$$

$$-\frac{\partial^2 \log L_i}{\partial \theta^2} = \frac{T_i}{2\theta^2},$$

$$-\frac{\partial^2 \log L_i}{\partial \gamma^2} = \frac{\theta T_i}{(T_i\gamma+1)^3}\left(\sum_{t=1}^{T_i}\varepsilon_{it}\right)^2 - \frac{1}{2}\frac{T_i^2}{(T_i\gamma+1)^2}.$$

After estimation, we derive estimates of the underlying variances as $\sigma_\varepsilon^2 = 1/\theta$, $\sigma_u^2 = \gamma/\theta$. Estimated standard errors for these are not reported, but can be obtained easily using the delta method. Standard errors are reported for the estimates of θ and γ, though in general, one does not compute hypothesis tests about the variance parameters.

E18.4 Groupwise Heteroscedasticity in Random Effects

The random effects model can easily be extended to allow groupwise heteroscedasticity. We consider two forms. The model with the basic extension to groupwise heteroscedasticity is

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \\
\mathrm{E}[u_i | \mathbf{x}_{it}] &= \mathrm{E}[\varepsilon_{it}] = 0 \\
\mathrm{Cov}[u_i, \varepsilon_{is} | \mathbf{x}_{it}] &= 0 \text{ for all } i, s, t. \\
\mathrm{Var}[u_i | \mathbf{x}_{it}] &= \sigma_u^{\,2} \\
\mathrm{Var}[\varepsilon_{it} | \mathbf{x}_{it}] &= \sigma_{\varepsilon\, i}^{\;2}
\end{aligned}
$$

Thus, the variance of the unique component of the compound disturbance is allowed to vary across groups. (The variance of u_i could, in principle as well, but such a model would be inestimable, as there is only a single observation from the distribution of u_i in the sample.) This model is requested when you add

; Het = group

to the **REGRESS** command. *Note, this is not merely a correction to the asymptotic covariance matrix.* This computation produces a different set of weights and, therefore, a different set of estimates for the random effects model. The following illustrates for the data we have used in several earlier examples. The results based on homoscedasticity are shown first. The box of diagnostic statistics is identical to this case, save for the indication of the model used. Note, though, that the parameter estimates do change somewhat.

The number of groups in the sample may not exceed 50,000 when estimating this and the next model of heteroscedasticity.

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .001443
            Var[u]              =       .012866
            Corr[v(i,t),v(i,s)] =       .899169
            Sum of Squares            19.817672
            R-squared                   .978932
Fixed vs. Random Effects (Hausman)  =       .00
[ 5 degrees of freedom, prob. value = 1.000000]
[High (low) values of H favor F.E.(R.E.) model]
Var[e] above is an average. Groupwise
heteroscedasticity model was estimated.
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .57918***    .10137     5.71  .0000      .38050     .77787
  LOGHWY|    -.18700***    .05761    -3.25  .0012    -.29991    -.07409
  LOGH2O|     .05890***    .01840     3.20  .0014      .02284     .09497
 LOGUTIL|    -.35078***    .04253    -8.25  .0000    -.43414    -.26742
  LOGEMP|     .96707***    .01961    49.33  .0000      .92864    1.00549
Constant|    2.48380***    .16473    15.08  .0000     2.16094    2.80666
--------+--------------------------------------------------------------------
```

```
--------+-------------------------------------------------------------
 (Two Step Feasible GLS Estimates)
Estimates:  Var[e]              =       .001443
            Var[u]              =       .012866
--------+-------------------------------------------------------------
   LOGKP|      .61652***       .10156      6.07  .0000       .41747     .81558
  LOGHWY|     -.25482***       .05782     -4.41  .0000     -.36814    -.14150
  LOGH2O|      .05042***       .01838      2.74  .0061       .01439     .08644
 LOGUTIL|     -.35501***       .04262     -8.33  .0000     -.43854    -.27147
  LOGEMP|      .98293***       .01962     50.10  .0000       .94448    1.02138
Constant|     2.66833***       .16389     16.28  .0000      2.34711    2.98956
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

For the models with groupwise heteroscedasticity, the computation of the estimator is changed as follows: In all cases, an estimator of

$$\theta_i = 1 - \sigma_{\varepsilon i} / \sigma_{2i}$$

where $$\sigma_{2i}^2 = \sigma_{\varepsilon i}^2 + T_i\,\sigma_u^2 .$$

is needed for each i. We have to rely on some consistency results to have in hand an estimator of σ_u^2 regardless of what happens next. We use the one from the initial OLS and fixed effects regression. From the OLS regression, ignoring the degrees of freedom correction, which is now irrelevant, this would be

$$\begin{aligned}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} e_{it}^2}{\sum_{i=1}^{N} T_i} &= \sum_{i=1}^{N} \frac{T_i}{\sum_{i=1}^{N} T_i}\left(\frac{1}{T_i}\sum_{t=1}^{T_i} e_{it}^2\right)\\ &= \sum_{i=1}^{N} w_i \ \text{Est.}(\sigma_{\varepsilon,i}^2 + \sigma_u^2)\\ &= \text{Est.}(\overline{\sigma}_{\varepsilon}^2 + \sigma_u^2).\end{aligned}$$

From the fixed effects regression, we would have a variance estimator of $\overline{\sigma}_{\varepsilon}^2$ which, assuming that it converges to something, would, by subtraction, still be providing the estimator of σ_u^2 that we need. In computing the FGLS estimator, computation of θ_i will require this estimate and the estimate of $\sigma_{\varepsilon,i}$. The latter is computed by computing the group specific sum of squared residuals based on the consistent estimator of $\boldsymbol{\beta}$ (this may be FE or OLS, or, for that matter, any consistent estimator of $\boldsymbol{\beta}$), dividing by T_i or T_g for the stratification case, and subtracting the unchanging estimate of σ_u^2. The estimation of θ_i is changed for each group. Note that in the unbalanced panel case with stratification, the heteroscedasticity arises from two sources. Where $\sigma_{\varepsilon,g}$ differs by stratum, and T_i varies by group within the stratum, we will have to compute

$$\theta_{i,g} = 1 - \sigma_{\varepsilon,g} / \sigma_{2\,i,g}$$

where $$\sigma_{2\,i,g}^2 = \sigma_{\varepsilon,g}^2 + T_i\,\sigma_u^2 .$$

E18.4.1 A Model with Stratification and Grouping

Suppose that the groups in the data can be grouped at some higher level of stratification. For example, consider a panel in which city data are further grouped by state, so that there are several cities per state in the data. The groupwise heteroscedasticity might then be structured as

$$\text{Var}[\varepsilon_{i\,t}] = \sigma_j^2, j = 1,...,G, \quad G < N$$

$$N_j \quad = \text{ the number of groups contained in outer grouping } j.$$

To fit a model of this sort, use

; Hfn = the grouping variable

This is a simple change to the previous model in which the grouping variable supersedes the panel specification in the GLS computation. To continue our example, suppose we arbitrarily group our 48 states into eight groups of six states. (The states are in alphabetical order in the data file, so this is a meaningless grouping just for purpose of this example.) Results appear below.

```
CREATE      ; region = Trn((6*17), 0) $
REGRESS     ; Lhs = loggsp ; Rhs = x ; Panel ; Random Effects
            ; Hfn = region $
```

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .001443
            Var[u]              =       .012866
            Corr[v(i,t),v(i,s)] =       .899169
            Sum of Squares            18.148294
            R-squared                   .978645
Fixed vs. Random Effects (Hausman)  =     34.61
[ 5 degrees of freedom, prob. value =  .000002]
[High (low) values of H favor F.E.(R.E.) model]
Var[e] above is an average. Groupwise
heteroscedasticity model was estimated.
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .59098***     .10178     5.81  .0000      .39150     .79046
  LOGHWY|    -.22999***     .05803    -3.96  .0001    -.34372    -.11626
  LOGH2O|     .05220***     .01841     2.84  .0046      .01612     .08828
 LOGUTIL|    -.34696***     .04271    -8.12  .0000    -.43068    -.26324
  LOGEMP|     .98459***     .01967    50.05  .0000      .94603    1.02314
Constant|    2.59972***     .16511    15.74  .0000     2.27610    2.92334
--------+--------------------------------------------------------------------
```

```
--------+-------------------------------------------------------------
 (Groupwise Heteroscedastic)
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|      .57918***      .10137     5.71  .0000      .38050     .77787
  LOGHWY|     -.18700***      .05761    -3.25  .0012     -.29991    -.07409
  LOGH2O|      .05890***      .01840     3.20  .0014      .02284     .09497
 LOGUTIL|     -.35078***      .04253    -8.25  .0000     -.43414    -.26742
  LOGEMP|      .96707***      .01961    49.33  .0000      .92864    1.00549
Constant|     2.48380***      .16473    15.08  .0000     2.16094    2.80666
--------+-------------------------------------------------------------
(Two Step Feasible GLS Estimates)
--------+-------------------------------------------------------------
   LOGKP|      .61652***      .10156     6.07  .0000      .41747     .81558
  LOGHWY|     -.25482***      .05782    -4.41  .0000     -.36814    -.14150
  LOGH2O|      .05042***      .01838     2.74  .0061      .01439     .08644
 LOGUTIL|     -.35501***      .04262    -8.33  .0000     -.43854    -.27147
  LOGEMP|      .98293***      .01962    50.10  .0000      .94448    1.02138
Constant|     2.66833***      .16389    16.28  .0000     2.34711    2.98956
--------+-------------------------------------------------------------
```

E18.4.2 Exponential Heteroscedasticity in Random Effects

The one way random effects linear model,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it}$$

is extended to allow specific Harvey (1976) style heteroscedasticity in either component. The models are

$$\text{Var}[\varepsilon_{it}] = \sigma_\varepsilon^2 \, [\exp(\boldsymbol{\delta}_\varepsilon'\mathbf{h_i})]^2$$

$$\text{Var}[u_i] = \sigma_u^2 \, [\exp(\boldsymbol{\delta}_u'\mathbf{z}_i)]^2$$

Note that the variables in these variance functions are assumed to be time invariant. The program assumes this. The estimator is full information maximum likelihood. This model is estimated using the same FIML procedure as defined above. In the reparameterized model, we now have

$$\theta_i = \theta \times [\exp(-\boldsymbol{\delta}_\varepsilon'\mathbf{h_i})]^2$$

and

$$\gamma_i = \gamma \times [\exp(\boldsymbol{\delta}_u'\mathbf{z}_i)]^2 \times [\exp(-\boldsymbol{\delta}_\varepsilon'\mathbf{h_i})]^2$$

modification of the derivatives is a straightforward application of the chain rule. The Hessian is tedious, but, again, straightforward.

Request this estimator with

REGRESS ; Lhs = dependent variable
; Rhs = independent variables
; Panel ; Pds = specification or ; Str = specification
; MLE

and either or both of

; Hfe = variables in h, the variance of ε
; Hfu = variables in z, the variance of u $

To illustrate, we extend the model of the previous section by specifying

$$\text{Var}[\varepsilon_{it}] = \sigma_\varepsilon^2 \exp(\delta_\varepsilon \overline{unemp}_i)$$

$$\text{Var}[u_i] = \sigma_u^2 \exp(\delta_u \overline{unemp}_i)$$

The commands and model results follow:

```
CREATE    ; ubari = Group Mean(unemp, Pds = 17) $
REGRESS   ; Lhs = loggsp
          ; Rhs = x,one
          ; Panel ; MLE
          ; Hfu = ubari ; Hfe = ubari $
```

```
-----------------------------------------------------------------------------
Random Effects Linear Regression Model
Dependent variable                 LOGGSP
Log likelihood function        1368.13859
Restricted log likelihood       577.89766
Chi squared [   1 d.f.]        1580.48186
Significance level                 .00000
McFadden Pseudo R-squared      -1.3674410
Estimation based on N =    816, K =   8
Inf.Cr.AIC  =-2720.277 AIC/N =   -3.334
Variance of e(i,t) =              .002487
Variance of u(i  ) =              .261628
Corr[v(i,t),v(i,s)]=              .990585
LR test is vs. null of no random effect
Panel contained      48 nonempty groups
Exponential heteroscedasticity model for ui
Exponential heteroscedasticity model for eit
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .50530***      .08033     6.29  .0000      .34785    .66275
  LOGHWY|     -.18115***      .04492    -4.03  .0001     -.26919   -.09312
  LOGH2O|      .04780***      .01145     4.17  .0000      .02536    .07024
 LOGUTIL|     -.32485***      .03752    -8.66  .0000     -.39839   -.25130
  LOGEMP|     1.01343***      .01427    71.04  .0000      .98547   1.04139
Constant|     2.63598***      .09214    28.61  .0000     2.45539   2.81658
        |Reparameterized Variance Components for ML Search
   1/s2e|     402.139***    67.94438     5.92  .0000     268.970   535.307
  s2u/s2e|    105.211       93.84949     1.12  .2623     -78.731   289.152
        |Heteroscedasticity in unique term e(i,t)
   UBARI|     -.05172***      .01170    -4.42  .0000     -.07465   -.02880
        |Heteroscedasticity in common term u(t)
   UBARI|     -.22236***      .06136    -3.62  .0003     -.34261   -.10210
--------+--------------------------------------------------------------------
```

We note, the results above look reasonable enough. However, a closer look suggests that this is a poorly specified model in spite of this. The iterations ended with a diagnostic,

```
Maximum of   500 iterations. Exit iterations with status=1.
```

This suggests that one might want to look more closely at the specification.

E18.5 Autocorrelation

The model is

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i,$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \eta_{it}.$$

Note that the autocorrelation is embodied in the unique component, ε_{it}. It would not make sense in the context of this model to assume that u_i is autocorrelated, as it is assumed to be time invariant.

The AR1 model for the random effects model is estimated by two step FGLS. In the first step, an estimator of ρ is automatically produced by whatever panel data estimator has been used. This will be any of the fixed or random effects models with one or two way specifications. The estimator based on any of the panel data models is

$$r = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T_i} e_{it}e_{i,t-1} / \left[\sum_{i=1}^{N}(T_i - 1)\right]}{\sum_{i=1}^{N}\sum_{t=2}^{T_i} e_{it}^2 / \left[\sum_{i=1}^{N}(T_i - 1) - K\right]}, \text{ where } e_{it} = (y_{it} - \bar{y}_{i.}) - \hat{\boldsymbol{\beta}}'(\mathbf{x}_{is} - \bar{\mathbf{x}}_{i.}).$$

The estimator of **β** is whatever the most recent one happens to be at the time the calculation is made. The model fit above, now with a correction for first order autocorrelation, using the same estimate as before (0.725563), which is the value computed by the initial LSDV estimator – see the results in Section E17.2.4.

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
AR1 Model:  autocorrelation rho =       .725563
Estimates:  Var[e]*(1-rho^2)    =       .000448
            Var[u]*(1-rho^2)    =       .000908
            Corr[v(i,t),v(i,s)] =       .669637
            Sum of Squares          4438.689735
            R-squared                   .975254
Fixed vs. Random Effects (Hausman)  =     45.54
[ 5 degrees of freedom, prob. value =   .000000]
[High (low) values of H favor F.E.(R.E.) model]
Original:   Var[e]              =       .000946
            Var[u]              =       .001918
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient     Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
   LOGKP|     .33833**      .17076     1.98  .0476      .00365     .67301
  LOGHWY|    -.17402*       .09060    -1.92  .0548    -.35159     .00356
  LOGH2O|    -.00693        .02748     -.25  .8009    -.06079     .04693
 LOGUTIL|    -.23526***     .06962    -3.38  .0007    -.37171    -.09881
  LOGEMP|    1.10054***     .02341    47.00  .0000    1.05465    1.14643
Constant|    3.22127***     .24951    12.91  .0000    2.73225    3.71029
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E18.6 Two Way Random Effects Model

The panel data estimator also allows 'two way' random effects models. The random effects model for a two way design is

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i + w_t.$$

The model is described in standard textbooks such as Wooldridge (2010) or Greene (2011).

In this model, neither the number of time periods observed for each group nor the number of individuals observed in each period need be fixed. Your data can consist of simply a sample of observations indexed by both individual and time. The data setup is exactly as described in Section E17.3. To request the two factor model, you simply add the specification

; Period = time variable

to the usual command. Unlike a group stratification variable, the time variable must use the integers 1,2,...,T_i. As noted earlier, it is not necessary for every group to have data in every period; there may be gaps. But, if you do have a balanced panel, you can easily set up the time indicator with the Trn function in **CREATE**. For example, in the data set we have been using for our application, there are 17 observations for each state. We could use

```
CREATE      ; time = Trn(-17, 0) $
```

(The variable *yr*-1969 in the data set would have the right values.) If the sample is not balanced, in either dimension, it will be necessary to provide the time variable by some other means. When you request the two factor model, the command will appear as

```
REGRESS     ; Lhs = ... ; Rhs = ... ; Panel
            ; Period = time $
```

E18.6.1 Program Output for Two Factor RE Models

This estimator produces the full set of output described earlier for the one factor model and an additional set of results for the two factor model. The additional results will be

1. Full set of two factor fixed effects results. Do note, in accordance with the description above, this model, unlike the one way model, will contain an overall constant term. This model is estimated by OLS including both the time and group dummy variables. If your command contains **; Random effects**, the fixed effects results are not shown, though they are computed internally.

2. Full table of estimates of fixed effects (if requested with **; Output = 2**). Note, as well, that the fixed effects produced for the groups will differ from the earlier results, since by design, the time dummy variables are not orthogonal to the group dummy variables.

3. Test statistics for the two way fixed effects model. This consists of the log likelihood, sum of squared deviations, and R^2s for five models:

   a. overall constant term only, no regressors,
   b. group dummies, no regressors,
   c. regressors and overall constant term,
   d. full one way fixed effects model,
   e. full two way fixed effects model.

   You should observe rising log likelihoods and R^2s and falling sums of squares as you go down the table, but if your regressors do not have much explanatory power the reverse could happen between b and c.

5. Full set of results for the two way random effects model including the LM statistic, Hausman statistic, estimates of the variance components, and the usual coefficient estimates with standard errors.

E18.6.2 Application

The following continues the earlier example with the two factor models.

```
CREATE      ; t = year - 1979  $
REGRESS     ; Lhs = loggsp ; Rhs = x,one
            ; Period = t  ; Panel $
```

The first set of results is the same as shown earlier. The results for the two factor models are shown below.

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t) = e(i,t)+u(i)+w(t)
Estimates:  Var[e]              =       .001148
            Var[u]              =       .011624
            Corr[v(i,t),v(i,s)] =       .910126
            Var[w]              =       .001536
            Corr[v(i,t),v(j,t)] =       .572300
            Sum of Squares            17.866362
            R-squared                   .978976
Lagrange Multiplier Test vs. Model (3) =4486.28
[ 2 degrees of freedom, prob. value =  .000000]
[High values of LM favor FEM/REM over CR model]
Fixed vs. Random Effects (Hausman)  =     32.29
[ 5 degrees of freedom, prob. value =  .000005]
[High (low) values of H favor F.E.(R.E.) model]
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .48216***      .09189     5.25  .0000       .30206    .66227
  LOGHWY|    -.15038***      .05375    -2.80  .0051     -.25573   -.04503
  LOGH2O|     .01917         .01749     1.10  .2730     -.01510    .05344
 LOGUTIL|    -.28462***      .03871    -7.35  .0000     -.36049   -.20875
  LOGEMP|     .93432***      .01927    48.48  .0000      .89655    .97210
Constant|    3.00101***      .16042    18.71  .0000     2.68659   3.31543
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

The LM statistic has been adjusted for the two types of effects – there is no Baltagi and Li counterpart for this. The Hausman statistic is also recomputed.

E18.6.3 Technical Details

The model is

$$y_{it} = \mu + \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

The random effects model is estimated as follows: With estimates of the three variance components in hand, we compute the GLS estimator by computing the moments of the transformed variables

$$z_{it}^* = z_{it} - \theta_{1i}\,\bar{z}_{i.} - \theta_{2t}\,\bar{z}_{.t} + \theta_{3it}\,\bar{\bar{z}}_{..}$$

where z_{it} is either the vector $\mathbf{x}_{it}$, including the constant, or y_{it}, and

$$\sigma_{1i}^2 = \sigma_\varepsilon^2 + T_i\sigma_u^2,$$

$$\sigma_{2t}^2 = \sigma_\varepsilon^2 + N_t\sigma_w^2,$$

$$\sigma_{3it}^2 = \sigma_\varepsilon^2 + T_i\sigma_u^2 + N_t\sigma_w^2.$$

Then,

$$\theta_{1i} = 1 - \sigma_\varepsilon/\sigma_{1i},$$

$$\theta_{2t} = 1 - \sigma_\varepsilon/\sigma_{2t},$$

and

$$\theta_{3it} = \theta_{1i} + \theta_{2t} - 1 + \sigma_\varepsilon/\sigma_{3it}.$$

If N and T are fixed, these specialize to the familiar textbook formulas.

The LM statistic, which will now have two degrees of freedom, is the earlier one plus a term which looks the same as given in Section E16.6, but in which the roles of 'i' and 't' are reversed. Note, once again, the modification necessary if the panel is unbalanced.

Finally, we note the different method of moments estimators for the variance components. Consider the balanced panel case first. As before, the sums of squared residuals from OLS and the two way fixed effects estimators provide estimates of $\sigma_\varepsilon^2 + \sigma_u^2 + \sigma_w^2$ and σ_ε^2 respectively. A third moment estimator is provided by the one way fixed effects estimator computed previously, in which the mean squared residual estimates $\sigma_\varepsilon^2 + \sigma_w^2(1 + 1/T)$ in the balanced panel case. Call these estimators m_0, m_{FE2} and m_{FE1}, respectively. Without degrees of freedom corrections, we know that $m_0 \geq m_{FE1} \geq m_{FE2}$. The method of moments estimators are

$$\hat{\sigma}_\varepsilon^2 = m_{FE2}$$

$$\hat{\sigma}_w^2 = \frac{m_{FE1} - m_{FE2}}{(1+1/T)}$$

$$\hat{\sigma}_u^2 = m_0 - \frac{m_{FE1} - m_{FE2}}{(1+1/T)} - m_{FE2}$$

The first two are guaranteed to be nonnegative. The first obviously is. The numerator of the second must be nonnegative because the sum of squared residuals falls when the time dummy variables are added to the equation. The third estimator may be negative. With a bit of manipulation, it can be written $\hat{\sigma}_u^2 = m_0 - [T/(1+T)]m_{FE1} - [1/(1+T)]m_{FE2}$. There is no assurance that this is positive. These estimators can be replaced with degrees of freedom corrected estimators, and the attempt repeated. In the unbalanced panel case, the $1/T$ term is replaced with Q, defined below, the sample average of $1/T_i$. This does not change the possible problem, however.

The initial estimators may produce a full set of positive estimates. They do in our application above. If not, *LIMDEP* undertakes a search among some possible candidates for valid estimates of the variance components. To begin, the group means estimators, if computed, provide another possibility. For the one factor model,

$$s_{ols}^2 \text{ estimates } \sigma_\varepsilon^2 + \sigma_u^2$$

and the group means estimator,

$$s_{means}^2 \text{ estimates } Q\sigma_\varepsilon^2 + \sigma_u^2,$$

where $$Q = (1/N)\Sigma_i(1/T_i).$$

These two equations can be solved to provide alternative estimates of the variance components:

$$\hat{\sigma}_\varepsilon^2 = (s_{ols}^2 - s_{means}^2) / (1 - Q)$$

and $$\hat{\sigma}_u^2 = s_{means}^2 - Q\,\hat{\sigma}_\varepsilon^2.$$

In the two way random effects model, the moment equations and their solutions would be as follows:

$$s_{ols}^2 = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_u^2 + \hat{\sigma}_w^2 \quad \text{(from the OLS regression)},$$

$$s_{group\ means}^2 = \hat{\sigma}_u^2 + Q_u\,\hat{\sigma}_\varepsilon^2 \quad \text{(from the group means regression)},$$

and $$s_{period\ means}^2 = \hat{\sigma}_w^2 + Q_w\,\hat{\sigma}_\varepsilon^2 \quad \text{(from the period means regression)},$$

where N = the total number of individuals observed,

T = the total number of periods observed,

Q_u = $(1/N)\Sigma_i(1/T_i)$ (or $1/T$ if the sample is balanced),

and Q_w = $(1/T)\Sigma_t(1/N_t)$ (or $1/N$ if the sample is balanced).

Then, $$\hat{\sigma}_\varepsilon^2 = (s_{ols}^2 - s_{group\ means}^2 - s_{period\ means}^2)/(1 - Q_u - Q_w),$$

$$\hat{\sigma}_u^2 = s_{group\ means}^2 - Q_u\,\hat{\sigma}_\varepsilon^2$$

$$\hat{\sigma}_w^2 = s_{period\ means}^2 - Q_w\,\hat{\sigma}_\varepsilon^2.$$

The preceding may yet fail to produce a positive estimator for the variance of w_t or u_i. If so, a last ditch estimator is used. In the two way fixed effects model, we may take the estimates of the dummy variable coefficients as estimators of u_i and w_t. If so, then the sample variances of a_i and c_t are used as estimators of $\sigma_u^2$and σ_w^2. We should note, the need for a protracted search such as this might be taken as a suggestion that the data are not consistent with this model.

During estimation, a log is kept of the search for the estimates of the variance components. The following shows the entry for a model estimated with the Grunfeld data in which it takes several tries to find an estimator. We have added the boldface annotation to the text from the trace file.

```
REGR;Lhs=I;Rhs=F,C;Pds=20;Period=T;Panel$
Estimating variance components for random effects model.
Random Effects Model: v(i,t) = e(i,t) + u(i)
Q =       0.0500. (Note, in a balanced sample, Q=1/T)
Uses sum of squared deviations (ybar(i) - b*xbar(i))^2/(N-K-1)
2 Ests. of beta available are group means regression and OLS
Tries group means and OLS. These are sums of squares.
EE1 uses GROUP MEANS=  0.111911E+05, EE2 uses OLS=  0.277499E+05
       Trying LSDV residual variance to estimate Var[e].
       Trying to estimate Var[u] with EE1 - Q * Var[e]
First attempt is successful for the one factor model.
Current estimates: Var[e]=  0.477729E+04, Var[u]=  0.109522E+05
Now search for estimates for two way model.
Estimating variance components for 2 way REM.
Trying LSDV residual variance to estimate Var[e].
Variance estimate for unique term is OK.
This estimate of Var[e]=   4917.47408
Try Var[u]=EE1-Qu*Var[e], Var[w]=EE2-Qu*Var[e]
Negative estimate for variance of w.
Current estimates:Var[u]=  10945.19678 Var[w]=   -169.64266
Attempting to use LSDV to fix nonpositive estimates
Var. est. < 0. If Var[w]<0 use FIXED EFFECTS.  Found    4917.474
Final estimate is based on the variance of the fixed effects.
Estimated Var[u] using ai as ui
Estimated Var[w] using ct as wt
Reports the estimates actually used at this point.
Current estimates:Var[u]=  13190.45414 Var[w]=    946.35364
```

E18.7 Two and Three Way Nested Random Effects

The linear random effects model for panel data is extended to a three level nested structure,`

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \varepsilon_{ijt} + v_{ij} + u_i$$

This is not a ‘three way’ random effects model – the effects are strictly nested. An example might be a regression of test scores by school which includes a school effect (u_i), a teacher within school effect (v_{ij}) and the period or student of observation (ε_{ijt}). The model is fit by full information maximum likelihood. (Note, the random parameters model with multiple effects described below allows for up to 10 levels in this same fashion. The estimator there is maximum simulated likelihood.)

E18.7.1 Command

The command for this estimator is

REGRESS ; Lhs = dependent variable
; Rhs = independent variables
; MLE
; Stratum = identifier for broader grouping
; Cluster = identifier for narrower grouping

This is a random effects linear regression model, and thus provides the usual optional features, including residuals, fitted values, hypothesis tests, and so on. A few options are unavailable, the linear restricted estimator, AR(1) correction and the White and Newey-West robust covariance matrices.

E18.7.2 Results

Estimates produced by this program include an initial ordinary least squares regression with all statistics usually produced. Since the command has specified a **; Cluster** and **; Stratum** correction, the OLS covariance matrix will be corrected for the clustering and stratification (appropriately so). The application below shows this result. The OLS results will be followed by the maximum likelihood estimates of the model parameters. A reparameterized form of the model is estimated – see the technical details for discussion. Results generated by this estimator are shown and annotated in the application below. In addition to the fitted values and residuals that may be requested, the following results are saved:

Matrices: *b* and *varb* These include only the regression parameters, not the variance parameters.

Scalars: *ssqrd* = σ_ε^2

$R^2 = 1 - \dfrac{\Sigma_{i,j,t}\left(y_{ijt} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{ijt}\right)^2}{\Sigma_{i,j,t}(y_{ijt} - \overline{y})^2}$

s = σ_ε

sumsqdev = $\Sigma_{i,j,t}(y_{ijt} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{ijt})^2$

ρ = 0

degfrdm = *NOBS* - *K*

ybar = $\Sigma_{i,j,t}y_{ijt}$ / *NOBS*

s*y* = (1/(*NOBS*-1)) $\Sigma_{i,j,t}(y_{ijt} - \overline{y})^2$

kreg = *K*,

nreg = *NOBS*

logl = log likelihood

s2u = σ_u^2

s2v = σ_w^2

E18.7.3 Application

To continue our earlier application, we have arbitrarily divided our 48 states into 12 'regions' with four states in each (actually just contiguously grouped sets of four states in the data set – just for purpose of a numerical illustration of the computation). The commands used are

CREATE ; region = Trn(68,0) $
REGRESS ; Lhs = loggsp
; Rhs = x,one
; MLE
; Cluster = 17
; Stratum = region $

```
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of    816 observations contained     48 clusters defined by  |
|     17 observations (fixed number) in each cluster.                 |
| Sample of    816 observations contained     12 strata defined by    |
| variable REGION   which identifies by a value a stratum ID.         |
+---------------------------------------------------------------------+

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGGSP   Mean                 =      10.50885
             Standard deviation   =       1.02113
             No. of observations  =           816  Degrees of freedom
Regression   Sum of Squares       =       838.219           5
Residual     Sum of Squares       =       11.5898         810
Total        Sum of Squares       =       849.809         815
             Standard error of e  =        .11962
Fit          R-squared            =        .98636  R-bar squared =   .98628
Model test   F[  5,    810]       =   11716.50948  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     577.89755  Akaike I.C.   = -4.23959
             Restricted (b=0)     =   -1174.41748  Bayes  I.C.   = -4.20500
             Chi squared [  5]    =    3504.63006  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|      .45392       1.39449      .33  .7448    -2.27922   3.18707
  LOGHWY|      .08572        .77979      .11  .9125    -1.44264   1.61408
  LOGH2O|      .08663        .19609      .44  .6586     -.29769    .47096
 LOGUTIL|     -.18742        .59165     -.32  .7514    -1.34704    .97220
  LOGEMP|      .61908***     .18018     3.44  .0006      .26594    .97223
Constant|     2.01100*      1.12515     1.79  .0739     -.19425   4.21625
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  32 iterations. Status=0, F=   -1380.755
```

```
--------------------------------------------------------------------------------
Nested Random Effects Linear Regression Model
Total Sample of    816 Observations
Number of Strata in Sample:      12
Number of Clusters in Strata:
Average   Std.Dev.  Minimum   Maximum
    4.0        .00        4         4
Number of Observations in Clusters
   17.0        .00       17        17
Variance Components Decomposition
t = within cluster = period or observation,
i = cluster within stratum, j = stratum
Proportion is Var[.] / [Var(e)+Var(v)+Var(u)]
Source      Variance      Std.Dev. Proportion
e(t,i,j)        .001        .03789      .0642
v(  i,j)        .020        .14001      .8772
u(    j)        .001        .03617      .0585
Log likelihood for nested model =  1380.75501
Log likelihood for no effects    =  577.89756
Chi-squared[2] = 1605.7149,  Prob =      .0000
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient      Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
   LOGKP|     .60205***      .10209     5.90  .0000       .40196    .80214
  LOGHWY|    -.25292***      .05912    -4.28  .0000     -.36878   -.13705
  LOGH2O|     .05138***      .01845     2.78  .0054       .01521    .08754
 LOGUTIL|    -.34997***      .04278    -8.18  .0000     -.43383   -.26611
  LOGEMP|     .99074***      .01974    50.18  .0000       .95204   1.02943
Constant|    2.68569***      .17360    15.47  .0000     2.34543   3.02595
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E18.7.4 Technical Details

The analysis is based on Antweiler (2001). Antweiler analyzes a four level model, though the recursive pattern of his results suggests it would not be difficult to extend it to five or more levels. The *LIMDEP* adaptation restricts it to three levels. (The random parameters model discussed later allows higher numbers of levels.) These results are taken from the article: The four level model is

$$y_{ijkt} = \boldsymbol{\beta}'\mathbf{x}_{ijkt} + \varepsilon_{ijkt} + w_{ijk} + v_{ij} + u_i.$$

The group sizes in the nested structure are M_i, N_{ij} and T_{ijk}. The total sample size is

$$NOBS = \sum_{i=1}^{L} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} T_{ijk}$$

The following are used to parameterize the log likelihood:

$$\rho_u = \frac{\sigma_u^2}{\sigma_\varepsilon^2},\ \rho_v = \frac{\sigma_v^2}{\sigma_\varepsilon^2},\ \rho_w = \frac{\sigma_w^2}{\sigma_\varepsilon^2}.$$

$$\theta_{ijk} = 1 + T_{ijk}\rho_u,\ \phi_{ij} = \Sigma_{k=1}^{N_{ij}}\frac{T_{ijk}}{\theta_{ijk}},\ \theta_{ij} = 1 + \phi_{ij}\rho_v,\ \phi_i = \Sigma_{j=1}^{M_i}\frac{\phi_{ij}}{\theta_{ij}},\ \ \theta_i = 1 + \rho_w\phi_i$$

$$e_{ijkt} = y_{ijkt} - \mathbf{x}_{ijkt}'\boldsymbol{\beta}$$

$$A_{ijk} = \Sigma_{t=1}^{T_{ijk}} e_{ijkt}^2,\ B_{ijk} = \Sigma_{t=1}^{T_{ijk}} e_{ijkt},\ B_{ij} = \Sigma_{k=1}^{N_{ij}}\frac{B_{ijk}}{\theta_{ijk}},\ \ B_i = \Sigma_{j=1}^{M_i}\frac{B_{ij}}{\theta_{ij}}$$

Then,

$$\begin{aligned}\log L = \frac{-1}{2}[NOBS\log(2\pi\sigma_\varepsilon^2) + \Sigma_{i=1}^{L}\{ & \\ & \log\theta_i + \Sigma_{j=1}^{M_i}\{ \\ & \log\theta_{ij} + \Sigma_{k=1}^{N_{ij}}\{ \\ & \log\theta_{ijk} + \frac{A_{ijk}}{\sigma_\varepsilon^2} - \frac{\rho_u}{\theta_{ijk}}\frac{B_{ijk}^2}{\sigma_\varepsilon^2}\} - \frac{\rho_v}{\theta_{ij}}\frac{B_{ij}^2}{\sigma_\varepsilon^2}\} - \frac{\rho_w}{\theta_i}\frac{B_i^2}{\sigma_\varepsilon^2}\}].\end{aligned}$$

For the three level model, we set $L = 1$ and $\rho_w = 0$. We use the BFGS method with numerical derivatives to maximize this log likelihood. Antweiler suggests that the second derivatives needed for the estimator for the asymptotic covariance matrix of the maximum likelihood estimator are intractable – and proposes numerical derivatives. However, by taking advantage of the results that for the generalized regression model, the expected Hessian of the log likelihood will be block diagonal, we do obtain a tractable form for the asymptotic covariance matrix of the slope estimator – it is the counterpart to the moment matrix that would be used for the GLS estimator:

$$\begin{aligned}-\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = & \frac{1}{\sigma_\varepsilon^2}\Sigma_{i=1}^{L}\Sigma_{j=1}^{M_i}\Sigma_{k=1}^{N_{ij}}\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}\mathbf{x}_{ijkt}' \\ & - \frac{\rho_w}{\sigma_\varepsilon^2}\Sigma_{i=1}^{L}\Sigma_{j=1}^{M_i}\Sigma_{k=1}^{N_{ij}}\frac{1}{\theta_{ijk}}\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}\right)\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}'\right) \\ & - \frac{\rho_v}{\sigma_\varepsilon^2}\Sigma_{i=1}^{L}\Sigma_{j=1}^{M_i}\frac{1}{\theta_{ij}}\left(\Sigma_{k=1}^{N_{ij}}\frac{1}{\theta_{ijk}}\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}\right)\right)\left(\Sigma_{k=1}^{N_{ij}}\frac{1}{\theta_{ijk}}\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}'\right)\right) \\ & - \frac{\rho_u}{\sigma_\varepsilon^2}\Sigma_{i=1}^{L}\left(\Sigma_{j=1}^{M_i}\frac{1}{\theta_{ij}}\left(\Sigma_{k=1}^{N_{ij}}\frac{1}{\theta_{ijk}}\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}\right)\right)\right)\left(\Sigma_{j=1}^{M_i}\frac{1}{\theta_{ij}}\left(\Sigma_{k=1}^{N_{ij}}\frac{1}{\theta_{ijk}}\left(\Sigma_{t=1}^{T_{ijk}}\mathbf{x}_{ijkt}'\right)\right)\right)\end{aligned}$$

We use this matrix, specialized to three levels, for estimating the asymptotic covariance matrix of the slope estimator. As is usual in the linear regression case, we do not report asymptotic standard errors for the variance coefficients. We note, for purposes of testing for the nesting structure, one can use the likelihood ratio test.

E18.8 Multilevel and Multiple Effects in the RP Model

The following applies to all random parameters models in *LIMDEP* – the entire class of models estimable with the **; RPM** specification with only the exception of the two equation models, bivariate probit and sample selection. In this section, we document the use of the model in the linear regression case.

The model is based on an index function

$$Index_{it} = \boldsymbol{\beta}'\mathbf{x}_{it}$$

such as the linear regression model, $y_{it} = Index_{it} + \varepsilon_{it}$ or the probit model, where $y_{it} = 1(Index_{it}+\varepsilon_{it} > 0)$. We add to this M = up to 10 'effects.'

$$Index_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + c_{j1}\,\omega_{j1,i} + c_{j2}\,\omega_{j2,i} + \dots\, c_{jM}\,\omega_{jM,i}.$$

The c_{jm} are ones and zeros simply used to select the effects in the model. The effects are up to 10 normally distributed random terms associated with discrete group indicators such as strata, clusters, etc. Effects may appear singly or as products, and may be nested or simply be associated with any desired groupings of the data. The associated variables can be any desired discrete indicator that associates a unique value with a group. Consider an example based on test scores. Suppose we have nationwide data, arranged by *region*, *state*, *county*, *district*, *school*. These are individual test scores observed in five decreasing levels of aggregation. Then, in addition to the data on test scores (presumably individual students) and the covariates in **x**, we have variables with distinct codes for the five levels of aggregation – the only restriction is that codes must be integers from 1,2,...,9999. The specification is

; REM = name1, name2, ..., nameM

For our example, this would thus be

; REM = region,state,county,district,school

This estimator does not require that these 'effects' be nested. The effects can be defined at any level of aggregation, and could be a mixture of nested and nonnested groupings. Suppose, for example, you also had indicators of grouping by *type of program*, which might be one of, say, 10, which varies all over the range of observations, without respect to the other five groupings listed. For another example, one might also have a *party* effect in that list, for whether the state in question had a *Democratic*, *Republican*, or *Other Party* governor at the time This could also be included.

Effects may also be main or secondary (products). You can specify secondary effects by writing the effects as products, as in

; REM = name1, name2, name3*name4,
name2*name3*name4, name1*name4

You may define up to 10 effects or combinations of effects in total, using up to 10 basic effects. To continue the example, you might specify an interaction between state and district with

; REM = region,state,county,district,school,state*district

The **; REM** specification can be added to a random parameter model (RPM) or may appear by itself instead of **RPM ; Fcn = …**

E18.8.1 Command

This estimator uses *LIMDEP*'s package of random parameter model estimators, and thus is in a different class of estimators from those we have considered in this manual up to this point. These models are discussed in greater generality in Chapter R24. We will omit some of the detail in the specification here, as it is given in full in the broader chapter. For this application (the linear model), the essential part of the command is

REGRESS ; Lhs = the dependent variable
; Rhs = the independent variables
; RPM
; Pds = the correct specification for your panel (see below)
; REM = the specification of your random effects $

Typically, the panel specification in **; Pds = ...** would correspond to the structure of one of your effects variables. But, this is not required. Indeed, you could have **; Pds = 1**. But, if you are analyzing a panel, you should specify it as usual. Note that the command does not contain **; Panel**. This must be omitted from this command. The effects are set up as described above. There is one other specification that you should use. The estimator for this model is maximum simulated likelihood (described in the technical notes below). You may want to control the number of random draws used in the simulations. This is an extremely computation intensive estimator. The number of random draws is specified with

; Pts = the desired number

The default value is 100. For generating final results in a study, you will probably use several hundred. But, for exploratory work, as in our example below, you might want to choose a small value, such as 10 or 25.

E18.8.2 Application

In Section E18.6, we fit a two way random effects model of the form

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i + w_t.$$

For present purposes, we rewrite this as

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\varepsilon\varepsilon_{it} + \sigma_u u_i + \sigma_v w_t$$

where now, all random effects have variance one. This model is identical to the previous one. The extension we consider here can be written

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\varepsilon\varepsilon_{it} + \sigma_u u_i + \sigma_v w_t + \gamma(\sigma_u u_i)(\sigma_v w_t)$$

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\varepsilon\varepsilon_{it} + \sigma_u u_i + \sigma_v w_t + \theta u_i w_t.$$

That is, we add a product term which has a freely estimated additional 'effect' on the dependent variable. The commands are

```
CALC      ; Ran(1234579) $
CREATE    ; t = yr - 1969 $
REGRESS   ; Lhs = loggsp
          ; Rhs = x,one
          ; Pds = 17
          ; RPM
          ; Pts = 50
          ; REM = state,t,state*t $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGGSP   Mean                 =      10.50885
             Standard deviation   =       1.02113
             No. of observations  =           816  Degrees of freedom
Regression   Sum of Squares       =       838.219          5
Residual     Sum of Squares       =       11.5898        810
Total        Sum of Squares       =       849.809        815
             Standard error of e  =        .11962
Fit          R-squared            =        .98636  R-bar squared =   .98628
Model test   F[  5,   810]        =   11716.50948  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     577.89755  Akaike I.C.   = -4.23959
             Restricted (b=0)     =   -1174.41748  Bayes  I.C.   = -4.20500
             Chi squared [  5]    =    3504.63006  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LOGKP|     .45392***      .15355     2.96  .0031      .15298    .75487
  LOGHWY|     .08572         .08184     1.05  .2949     -.07468    .24612
  LOGH2O|     .08663***      .02479     3.50  .0005      .03805    .13521
 LOGUTIL|    -.18742***      .06580    -2.85  .0044     -.31639   -.05845
  LOGEMP|     .61908***      .02281    27.14  .0000      .57437    .66380
Constant|    2.01100***      .15245    13.19  .0000     1.71221   2.30979
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  32 iterations. Status=0, F=   -1346.382

-----------------------------------------------------------------------------
Random Coefficients  LinearRg Model
Dependent variable              LOGGSP
Log likelihood function     1346.38369
Restricted log likelihood       .00000
Chi squared [   3 d.f.]     2692.76738
Significance level              .00000
Estimation based on N =    816, K =  10
Inf.Cr.AIC  =-2672.767 AIC/N =   -3.275
Sample is 17 pds and     48 individuals
LINEAR regression model
Simulation based on  50 random draws
Model contained  3 random effects.
```

```
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Nonrandom parameters
   LOGKP|     .82974***     .17833     4.65  .0000      .48021   1.17926
  LOGHWY|    -.32864***     .09112    -3.61  .0003    -.50722   -.15005
  LOGH2O|     .02490        .02877      .87  .3868    -.03149    .08128
 LOGUTIL|    -.49038***     .07484    -6.55  .0000    -.63708   -.34369
  LOGEMP|     .99220***     .02879    34.47  .0000      .93578   1.04863
Constant|    2.70616***     .20479    13.21  .0000     2.30477   3.10755
        |Standard Deviations of Random Effects
R.E.(01)|     .18379***     .00175   105.06  .0000      .18036    .18722
R.E.(02)|     .00077        .00154      .50  .6173    -.00224    .00378
R.E.(03)|     .00479**      .00191     2.51  .0119      .00106    .00853
        |Variance parameter given is sigma
Std.Dev.|     .03899***     .00051    76.96  .0000      .03800    .03999
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------

+--------------------------------------------+--------------+
| Random effects in the model are based on   |Random Effect |
| these expanded qualitative variables.      |     Variance |
| R.E.(01) = STATEID                         |      .033779 |
| R.E.(02) = T                               |      .000001 |
| R.E.(03) = STATEID  T                      |      .000023 |
+--------------------------------------------+--------------+

----------------------------------------------------------------------
Random Effects Model: v(i,t) = e(i,t)+u(i)+w(t)
Estimates:  Var[e]              =       .001148
            Var[u]              =       .011624
            Corr[v(i,t),v(i,s)] =       .910126
            Var[w]              =       .001536
            Corr[v(i,t),v(j,t)] =       .572300
            Sum of Squares            17.866362
            R-squared                   .978976
(Some results omitted)
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGGSP|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
   LOGKP|     .48216***     .09189     5.25  .0000      .30206    .66227
  LOGHWY|    -.15038***     .05375    -2.80  .0051    -.25573   -.04503
  LOGH2O|     .01917        .01749     1.10  .2730    -.01510    .05344
 LOGUTIL|    -.28462***     .03871    -7.35  .0000    -.36049   -.20875
  LOGEMP|     .93432***     .01927    48.48  .0000      .89655    .97210
Constant|    3.00101***     .16042    18.71  .0000     2.68659   3.31543
--------+-------------------------------------------------------------
```

E18.8.3 Technical Details

Conditioned on the unobserved effects, the contribution of each observation to the log likelihood for the linear regression model is

$$\log L_i|\boldsymbol{\omega}_i = -\log\sigma_\varepsilon - ½ \log 2\pi - ½ (\varepsilon_i / \sigma_\varepsilon)^2$$

where 'i' is used generically to denote a single observation. Conditioned on the effects, the observations are independent. In the conditional form above, $\boldsymbol{\omega}_i$ is the set of up to 10 random effects. There are assumed to be T_i 'observations' for individual i. The conditional likelihood is

$$L_i \mid \omega = \prod_{t=1}^{T_i} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_{it}}{\sigma_\varepsilon}\right)^2\right]$$

where $\varepsilon_{it} = y_{it}$ - $\boldsymbol{\beta}'\mathbf{x}_{it}$ – *all common effects*. The unconditional likelihood function is obtained by integrating out the common effects:

$$L_i = \int_{\omega_i} \prod_{t=1}^{T_i} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_{it}}{\sigma_\varepsilon}\right)^2\right] f(\boldsymbol{\omega}_i)d\boldsymbol{\omega}_i.$$

This integral is approximated by simulation. The function that we maximize with respect to $(\boldsymbol{\beta},\sigma_\varepsilon,\boldsymbol{\gamma})$ is

$$\log L = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_i} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_{itr}}{\sigma_\varepsilon}\right)^2\right]\right)$$

Further details on the maximization appears in Chapter R24. We note one important aspect of the simulation/integration. Where the common effect is of the form $\sigma_\omega u_i$ – that is, the subscript on the effect matches the index of the product operation, as in the familiar random effects model – then the preceding is exactly equivalent to that RE model. In other cases, however, the effect may be varying over a different range than the index in the product. Consider the time effects in our example. There are 17 of them in each i, since each state is observed in each period. Thus, for our example,

$$\varepsilon_{it,r} = y_{it} - \boldsymbol{\beta}'\mathbf{x}_{it} - \gamma_1 v_{i,r} - \gamma_2 w_{it,r} - \gamma_3 v_{i,r} w_{it,r}.$$

That is, the integral over periods is recomputed for each i, while the integral over v_i is only computed once. Moreover, in principle, though w_t is a 'time' effect, we are treating it as if it were a state specific time effect when we integrate it out. (There is a separate random variable w_t for each period, however.) This means that although state observations are correlated across states because of the common time effect, we are treating them as uncorrelated by this procedure. Thus, it must be considered approximate.

E19: Random Parameters Linear Models

E19.1 Introduction

LIMDEP provides two approaches to fitting linear regression models with random parameters:

- Mixed or random parameters models – parameters are distributed continuously
- Latent class or finite mixture models – parameters have a discrete distribution

The models are built around the structural equations

$$y_{it} \;=\; \boldsymbol{\alpha}'\mathbf{w}_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i,\; i = 1,...,N,\; t = 1,\ldots,T_i,$$

$$\varepsilon_{\text{it}} \;\sim\; \text{N}[0,\sigma_{\text{i}}^2]$$

For the mixed model, the general form is

$$\boldsymbol{\beta}_I \;=\; \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i,$$

$$\text{E}[\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i] \;=\; \mathbf{0},\;\; \text{Var}[\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i] \;=\; \mathbf{I},$$

$$\text{Var}[\boldsymbol{\beta}_i|\,\mathbf{x}_i,\mathbf{z}_i] \;=\; \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}',$$

$$\sigma_i^2 \;=\; \sigma^2 \text{ (constant).}$$

The familiar linear regression model, a random effects linear model, and a hierarchical linear model are all particular cases. In the latent class model,

$$\boldsymbol{\beta}_i,\sigma_i^2 \;\in\; [(\boldsymbol{\beta}_1,\sigma_1^2), (\boldsymbol{\beta}_2,\sigma_2^2), \ldots, (\boldsymbol{\beta}_j,\sigma_j^2), \ldots (\boldsymbol{\beta}_J,\sigma_J^2)],$$

$$\text{Prob[class=}j|\mathbf{z}_i] \;=\; \pi_j(\mathbf{z}_i,\boldsymbol{\theta}_j),\; j = 1,\ldots,J.$$

Estimation of the random parameters (RP) model is described in this chapter. The latent class model is documented in Chapter E20.

The randomness of the parameters is interpreted simply as latent heterogeneity. The linear regression model with random coefficients is normally associated with panel data settings, but we allow this formulation with a cross section as well. The model is identified in a cross section, though results are generally better when this is applied to a panel.

E19.2 Random Parameters Linear Models

The structure of the basic model is

$$y_{it} = \boldsymbol{\alpha}'\mathbf{w}_{it} + \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it},\ i = 1,...,n,\ t = 1,\ldots,T_i.$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{v}_i.$$

The conditional mean function is

$$E[y_{it}|\, \mathbf{x}_{it}, \boldsymbol{\beta}_i] = \boldsymbol{\alpha}'\mathbf{w}_{it} + \boldsymbol{\beta}_i'\mathbf{x}_{it},\ i = 1,...,n,\ t = 1,...,T_i.$$

The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) distribution,

$$E[\boldsymbol{\beta}_i|\, \mathbf{X}_i] = \boldsymbol{\beta},\ \ \mathbf{X}_i = [\mathbf{x}_1,\ldots,\mathbf{x}_{Ti}]$$

$$\text{Var}[\boldsymbol{\beta}_i|\, \mathbf{X}_i] = \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'.$$

The full parameter vector is partitioned into a nonrandom part, $\boldsymbol{\alpha}$, which multiplies a set of K_0 regressors and the random part $\boldsymbol{\beta}_i$ which multiples the remaining K_1 of the total of K regressors. The random coefficient vector, $\boldsymbol{\beta}_i$ is assumed to be distributed with mean provided by the deterministic component, $\boldsymbol{\beta}$ and stochastic component, $\boldsymbol{\Gamma}\mathbf{v}_i$. The random vector, $\mathbf{v}_i$, is assumed to have mean zero (with no loss of generality, given $\boldsymbol{\beta}$) and covariance matrix equal to an identity matrix, $\mathbf{I}$. The coefficient matrix, $\boldsymbol{\Gamma}$, provides the variances and cross parameter correlations in the distribution of $\boldsymbol{\beta}_i$. For estimation purposes, $\boldsymbol{\Gamma}$ is taken to be a free lower triangular matrix, so the covariance matrix of the random parameter vector is $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$. The base case assumes that $\boldsymbol{\Gamma}$ is a diagonal matrix with diagonal element γ_k.

Estimation is based on the following approach: (It is only sketched here. A fuller discussion of the method and the underlying theory is presented in Chapter R24.) The reduced form of this model is

$$y_{it}|\boldsymbol{\beta}_i = \boldsymbol{\alpha}'\mathbf{w}_{it} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'[\mathbf{v}_i \otimes \mathbf{x}_{it}] + \varepsilon_{it}$$

$$= \boldsymbol{\theta}'\mathbf{h}_{it}(\mathbf{v}_i) + \varepsilon_{it}.$$

We assume that ε_{it} is normally distributed with mean zero and variance σ^2 and is uncorrelated with all other observations, ε_{js}, $j \neq i$. Conditioned on realizations $\mathbf{v}_i$, that is, on the sample of draws, $\mathbf{V} = [\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n]$, the maximum likelihood estimator of the set of structural parameters, $\boldsymbol{\theta}$, is the least squares estimator. But, this is only conditional on a particular realization of the parameter vector. In order to obtain the unconditional estimator, it would be necessary to take the expectation of this estimator, over the distribution of the random parameter vector. This would be the integral of the conditional density over the range of $\boldsymbol{\beta}_i$ (induced by $\mathbf{v}_i$). Since this integral is unlikely to have a closed form in general, we use simulation to approximate the distribution, instead. A total of R draws of $\boldsymbol{\beta}_i$ are obtained for each i. The results are averaged over the draws. Thus, the full set of structural parameters is obtained by minimizing the sum of squares. The procedure is iterated over the estimated disturbance variance, until convergence or a maximum of 20 iterations.

NOTE: If only the constant term is considered random – the model may specify that some parameters are nonrandom – then this model is equivalent to the random effects model presented in Chapter E18.

E19.3 Command for the Random Parameters Models

The basic command for the random parameters regression estimators is

REGRESS ; Lhs = dependent variable
; Rhs = full list of independent variables
; RPM
; Fcn = random parameters specification
; other specifications $

Use

SETPANEL ; … $

and **; Panel** in the command to specify a panel. If this is omitted, the data are assumed to be a cross section.

NOTE: For this model, your Rhs list *should* include a constant term.

E19.3.1 Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one,x1,x2,x3,x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be nonrandom (i.e.,) constant. For those that you wish to specify as random, use

; Fcn = variable name (distribution),
variable name (distribution), ...

Three distributions may be specified All random variables have mean 0.

n = standard normal distribution, variance = 1,
t = triangular (tent shaped) distribution in [-1,+1], variance = 1/6,
u = standard uniform distribution [-1,1], variance = 1/3
or *c* = variance = 0. (The parameter is not random.)

For an example, to specify that the constant term and the coefficient on *x*1 are normally distributed with fixed mean and variance, use

; Fcn = one(n), x1(n).

This specifies that the first and second coefficients are random while the remainder are not. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

Each random parameter is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. For example, if you specify that a parameter is normally distributed, then that parameter is $\beta_{k,i} = \beta_k + \sigma_k v_{i,k} \sim N[\beta,\sigma_k^2]$. For a variable with the triangular or uniform distribution, the variance of $\beta_{k,i}$ is $\sigma_k^2/6$ or $\sigma_k^2/3$, respectively. (See Chapter R24 for discussion of this computation. There are several additional specifications for random parameters discussed there as well. Some are provided to reduce the amount of variation in the tails of the distribution of the parameters across individuals and to limit the range of variation. (See Train (2009) for discussion.)

E19.3.2 Constraining the Sign of a Parameter – Lognormal and Triangular

Two methods are provided for constraining the sign of a parameter.

1. **Lognormal distribution:** $\beta_i = \exp(\beta + \gamma v_i) = \exp(\beta)\times[\exp(v_i)]^\gamma$.

The parameter thus specified is constrained to be positive. Use

; Fcn = variable (l) (type ‘el’ not ‘one’)

The lognormal distribution is effective, but can cause problems in estimation. If your theory specifies a positive parameter for all *i*, but the model is not well specified, then the estimator may be improperly attempting to force the parameter to be positive. The situation can be visualized by considering a model in which the simple least squares estimate of β is large and negative. If you then try to force the parameter to be positive by specifying a lognormal distribution, the end result will be that the mean will gravitate toward $-\infty$ and σ will tend toward zero. The second complication with lognormal parameters is that the distribution has a long thick tail and can allow large a probability of implausible values even if they do have the right sign. Given these two results, we find that the lognormal specification can be a difficult model to work with.

2. **One sided triangular distribution:** $\beta_i = \beta + \beta v_i$.

This specification is obtained with

; Fcn = variable (o).

The distribution is produced by forcing equality of the mean parameter and the scaling parameter. The result ranges between 0 and 2β, which is always positive if β is positive and always negative if β is negative. This specification has the advantage that it will accommodate the underlying data and not force a sign on the coefficient.

E19.3.3 Correlated Random Parameters

The preceding defines an estimator for a model in which the covariance matrix of the random parameters is diagonal. To extend it to a model in which the parameters are freely correlated, add

; Correlation (or just ; Cor)

to the command. The uncorrelated case is characterized by the diagonal matrix, Γ, so that the full set of random parameters is

$$\beta_{k,i} = \beta_k + \gamma_k v_{k,i},\ k = 1,\ldots,K_1.$$

In order to induce the correlation across parameters, the below diagonal elements of Γ are allowed to be nonzero. In this case, we have

$$\beta_{1,i} = \beta_1 + \gamma_{11} v_{1,i},$$

$$\beta_{2,i} = \beta_2 + \gamma_{21} v_{1,i} + \gamma_{22} v_{2,i},$$

$$\beta_{3,i} = \beta_3 + \gamma_{31} v_{1,i} + \gamma_{32} v_{2,i} + \gamma_{33} v_{3,i},$$

and so on. The implied covariance and correlation matrices are reported with the final results of the model. We note one caution with this specification. If all random parameters are assumed to be normally distributed, the mixing of the distributions shown above will preserve the normality. In all other cases, the mixed distribution will not retain the specification of the model. For example, if you specify that parameter 1 is normally distributed and parameter 2 is uniformly distributed in your model specification, then parameter 1 will retain the specification, but parameter 2 will be distributed as the sum of a normal and a uniform random variable, which is complicated. Thus, while free correlations are estimable, it must be understood that the mixed distributions that give rise to the correlations may not have the expected shapes.

E19.3.4 Autocorrelation

The latent heterogeneity may evolve over time, rather than remain constant. To accommodate this, you may specify that

$$v_{ikt} = \rho_k v_{ik,t-1} + u_{ikt}$$

that is, the familiar AR(1) kind of model. For only a nonrandom constant term, this is similar to the autocorrelation setup for the random effects model, but note that a crucial difference here is that it is the common term that evolves over time (comparable to u_i) rather than the unique term, ε_{it}. The specification is

; AR1

E19.4 Hierarchical Model – Heterogeneity in the Means

We obtain a hierarchical model by allowing the mean of the parameters to vary with a set of covariates,

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it},\ i = 1,...,n,\ t = 1,\ldots,T_i.$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i.$$

(We are allowing some of the parameters to be nonrandom. For convenience, the term $\boldsymbol{\alpha}'\mathbf{w}_{it}$ is absorbed in $\boldsymbol{\beta}_i'\mathbf{x}_{it}$.) The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) distribution,

$$E[\boldsymbol{\beta}_i|\ \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

This expanded formulation produces two useful special cases:

$\boldsymbol{\Delta} = \mathbf{0}$ is the familiar random parameters model, as in Section E19.3.
$\boldsymbol{\Gamma} = \mathbf{0}$ produces a hierarchical model with some interaction terms.

The implied form of the coefficients and the regression function are

$$E[\beta_{ki}] = \beta_k + \Sigma_m\, \delta_{km}\, z_{mi}$$

$$y_{it}|\boldsymbol{\beta}_i = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'[\mathbf{z}_i \otimes \mathbf{x}_{it}] + \boldsymbol{\gamma}'[\mathbf{v}_i \otimes \mathbf{x}_{it}] + \varepsilon_{it}$$

$$= \boldsymbol{\theta}'\mathbf{h}_{it}(\mathbf{v}_i) + \varepsilon_{it}$$

where z_m is a variable that is measured for each individual. The command is be modified to

; RPM = list of variables in z (must not include one).

In a panel data set, these variables must be repeated for each observation in the group. They are assumed not to vary over time. (Typically, they would be sociodemographic variables such as gender or education.)

A device is provided to allow the list of variables z_i to differ across coefficients. The general format is as follows: The specification **; RPM = list of variables** provides the full list of variables to be used. For example, if **; RPM = z1,z2,z3,z4**, then

; Fcn = x1(n)

specifies that the coefficient is $\beta_{1i} = \beta_1 + \delta_{11}z1_i + \delta_{12}z_{2i} + \delta_{13}z_{3i} + \delta_{14}z_{4i} + \gamma_1 v_{1i}$. To remove $z2$ and z_3 from this mean, use

; Fcn = x1(n | # 1001).

The vertical bar is followed by a # sign followed by a string of 0s and 1s to indicate exclude or include the respective zs from the mean of the coefficient. This device can be used with any specific coefficient, or all of them. Note that the number of digits after the # must exactly match the number of variables in the RPM list.

E19.5 Saved Results

Results saved automatically by this estimator are:

Matrices:	*b*	=	estimate of θ
	varb	=	asymptotic covariance matrix for estimate of θ.
	gammarpm	=	maximum likelihood estimate of Γ
	sdrpm	=	vector of estimated standard deviations from $\Gamma\Gamma'$ computed as the square roots of the diagonals of $\Gamma\Gamma'$.
Scalars:	*kreg*	=	number of variables in Rhs
	nreg	=	number of observations
	logl	=	log likelihood function

Last Model: *b_variables*

Last Function: None

There is no *Last Function* saved for the **PARTIALS** or **SIMULATE** command by the random parameters models, because of the need to simulate the parameters to do the computations. Partial effects and predicted values are computed locally within the estimator, and can be requested with

; Partial Effects

and **; Keep = variable** and/or **; Res = variable**.

E19.6 Controlling the Simulation

There are three parameters of the simulations that you can change. The number of points in the simulation is *R*. Authors differ in the appropriate value. Train recommends several hundred. Bhat suggests 1,000 as an appropriate value. The program default is 100. You can choose the value with

; Pts = number of draws, *R*

The values of 25 or 50 that we set in our experiments are chosen purely to produce an example that you can replicate without spending an inordinate amount of waiting for the results. (Simulation based estimation is unavoidably a time intensive computation.)

In order to replicate a simulation based estimation, you must use the same random draws. One implication of this is that if you give the identical model command twice in sequence, you will not get the identical set of results because the random draws in the sequences will be different. To obtain the same results, you must reset the seed of the random number generator with a command such as

CALC ; Ran(seed value) $

The specific value you use for the seed is not of consequence; any odd number will do. (That value is modified internally to produce a seed for each individual in the sample. But, it remains that you can replicate a set of results by using the same global seed.)

In this connection, we note a consideration which is crucial in this sort of estimation. The random sequence used for the model estimation must be the same in order to obtain replicability. In addition, during estimation of a particular model, the same set of random draws must be used for each person every time. That is, the sequence $\mathbf{v}_{i1}$, $\mathbf{v}_{i2}$, ..., $\mathbf{v}_{iR}$ used for individual i must be the same every time it is used to calculate a probability, derivative, or likelihood function. (If this is not the case, the likelihood function will be discontinuous in the parameters, and successful estimation becomes unlikely.) One way to achieve this that has been suggested in the literature is to store the random numbers in advance, and simply draw from this reservoir of values as needed. Because *LIMDEP* is able to use very large samples, this is not a practical solution, especially if the number of draws is large as well. We achieve the same result by assigning to each individual, i, in the sample, their own random generator seed which is a unique function of the global random number seed, S, and their group number, i;

$$\text{Seed}(S,i) = S + 123.0 \times i, \text{ then minus 1.0 if the result is even.}$$

Since the global seed, S, is a positive odd number, this seed value is unique, at least within the several million observation range of *LIMDEP*.

The standard approach to simulation estimation is to use random draws from the specified distribution. As suggested above, good performance in this connection requires very large numbers of draws. The drawback to this approach is that with large samples and large models, this entails a huge amount of computation and can be very time consuming. Some authors have documented dramatic speed gains with no degradation in simulation performance through the use of a small number of Halton draws instead of a large number of random draws. Halton sequences are discussed in Chapter R24. Authors (e.g., Bhat (2001)) have found that a Halton sequence of draws with only one tenth the number of draws as a random sequence is equally effective. (The efficiency does fall as the number of parameters rises, but large gains persist.) To use this approach, add

; Halton

to your model command.

E19.7 Other Options

Other optional features for the random parameters regression model are the usual, including

; Keep = name	to retain predictions
; Prob = name	to retain fitted probabilities
; Res = name	to retain residuals
; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**
; List	to display predicted values
; Table= name	to retain the model results for constructing tables
; Test:	to test hypotheses about $\boldsymbol{\beta}$.

In spite of the fact that this is a linear regression model, the estimator is nonlinear. The default (and best) algorithm for estimation is BFGS. But, all other algorithms are available as are other settings for the optimization process:

; Alg = DFP, BFGS, Newton	
; Maxit = n	to set maximum iterations
; Tlg[= value]	to set tolerance for convergence on gradient
; Tlb[= value]	to set tolerance for convergence on change in parameters
; Tlf [= value]	to set tolerance for convergence on change in log likelihood
; Output = value	to control output during iterations

This estimator can accommodate restrictions, so

; Rst = list

and **; CML: specification**

are both available. Do note that forcing the ancillary parameter, in this case, the variance parameters, to equal a slope parameter will almost surely produce unsatisfactory results, and may impede or even prevent convergence of the iterations.

E19.8 Individual Specific Estimates

Individual specific estimates of $E[\boldsymbol{\beta}_i|\text{data}_i]$ can be obtained by the method described in Chapter R24, by adding

; Parameters

to your command. This requests computation of matrices *beta_i* and *sdbeta_i* that contain the estimated means and standard deviations of the conditional distributions of β_i. Some discussion appears below in the applications.

E19.9 Applications

We provide two illustrations to demonstrate the linear RP model.

E19.9.1 Random Parameters Linear Regression Model

This application shows a straightforward application of the RP model in an unbalanced panel. There are three random parameters, which are assumed to be correlated. The simulation uses Halton sequences rather than random draws, so there is no need to set the seed for the random number generator.

```
SAMPLE      ; All $
CREATE      ; income = hhninc $
REJECT      ; income = 0 $
CREATE      ; loginc = Log(income) $
SETPANEL    ; Group = id ; Pds = ti $
REGRESS     ; Lhs = loginc ; Rhs = one,age,age*age,educ,female
            ; Panel
            ; RPM ; Fcn = one(n),educ(n),female(n)
            ; Correlated ; Halton ; Pts = 25
            ; Parameters $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGINC   Mean                    =      -1.15746
             Standard deviation      =        .49149
             No. of observations     =         27322  Degrees of freedom
Regression   Sum of Squares          =       718.053            4
Residual     Sum of Squares          =       5881.56        27317
Total        Sum of Squares          =       6599.61        27321
             Standard error of e     =        .46401
Fit          R-squared               =        .10880  R-bar squared =   .10867
Model test   F[  4, 27317]           =     833.75219  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =   -17786.71322  Akaike I.C.   = -1.53550
             Restricted (b=0)        =   -19360.30987  Bayes  I.C.   = -1.53400
             Chi squared [  4]       =    3147.19331  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGINC|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|      .07744***      .00211    36.64  .0000       .07330    .08159
 AGE*AGE|     -.00087***   .2386D-04   -36.41  .0000      -.00092   -.00082
Constant|    -3.35440***      .04689   -71.54  .0000    -3.44630  -3.26251
    EDUC|      .05229***      .00125    41.98  .0000       .04985    .05473
  FEMALE|     -.01690***      .00572    -2.95  .0032      -.02812   -.00568
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Random Coefficients  LinearRg Model
Dependent variable               LOGINC
Log likelihood function    -12047.25811
Estimation based on N =  27322, K =  12
Inf.Cr.AIC  =24118.516 AIC/N =     .883
Unbalanced panel has   7293 individuals
LINEAR regression model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGINC|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        | Nonrandom parameters
     AGE|      .06799***      .00120    56.86  .0000       .06564    .07033
 AGE*AGE|     -.00067***   .1358D-04   -49.13  .0000      -.00069   -.00064
        | Means for random parameters
Constant|    -3.40320***      .02661  -127.91  .0000    -3.45535  -3.35106
    EDUC|      .05824***      .00078    75.04  .0000       .05672    .05976
  FEMALE|     -.05416***      .00355   -15.25  .0000      -.06111   -.04720
        | Diagonal elements of Cholesky matrix
Constant|      .50105***      .00862    58.11  .0000       .48415    .51795
    EDUC|      .00471***      .00020    23.78  .0000       .00432    .00510
  FEMALE|      .01135***      .00252     4.50  .0000       .00640    .01629
        | Below diagonal elements of Cholesky matrix
lEDU_ONE|      .01066***      .00071    15.09  .0000       .00928    .01204
lFEM_ONE|      .07436***      .00335    22.21  .0000       .06780    .08092
lFEM_EDU|      .04889***      .00324    15.09  .0000       .04254    .05524
        | Variance parameter given is sigma
Std.Dev.|      .30580***      .00078   390.99  .0000       .30427    .30733
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

```
Implied covariance matrix of random parameters
Var_Beta|             1             2             3
--------+------------------------------------------
       1|       .251053      .00534136      .0372579
       2|     .00534136   .135804E-03     .00102286
       3|      .0372579      .00102286     .00804854
Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       .501052
       2|      .0116535
       3|      .0897137
Implied correlation matrix of random parameters
Cor_Beta|             1             2             3
--------+------------------------------------------
       1|       1.00000       .914770       .828851
       2|       .914770       1.00000       .978367
       3|       .828851       .978367       1.00000
```

E19.9.2 Conditional Estimates of Means of Random Parameters

Data file dairy.dat contains six years of observations on 247 dairy farms in northern Spain, drawn from 1993-1998. The raw data consist of the *farm* and *year* identification, plus measurements on one output, *milk*, and four inputs, *cows*, *land*, *labor* and *feed*. Figure E19.1 displays several observations.

A1 FARM

	A	B	C	D	E	F	G	H	I	J	K	L
1	FARM	YEAR	COWS	LAND	MILK	LABOR	FEED	YIT	X1	X2	X3	X4
2	1	93	15.3	8	73647	2	33435.74	11.20704	-0.25085	-0.38127	0.236075	-0.26802
3	1	94	18.1	8	91260	2	36869.04	11.42147	-0.08279	-0.38127	0.236074	-0.17027
4	1	97	17.1	7	110419	2	51013.58	11.61204	-0.13962	-0.5148	0.236075	0.154448
5	1	96	17.3	8	111454	2	50711.57	11.62137	-0.12799	-0.38127	0.236075	0.14851
6	1	95	17.8	8	118498	2	54153.59	11.68265	-0.0995	-0.38127	0.236075	0.21418
7	1	98	19.5	7.2	131197	2	59038.65	11.78446	-0.00829	-0.48663	0.236075	0.300549
8	2	93	20.3	9	118149	2	53875.94	11.6797	0.031921	-0.26349	0.236075	0.20904
9	2	94	20.3	10.4	127742	2	51990.99	11.75777	0.031921	-0.11891	0.236075	0.173427
10	2	95	22	10.7	146490	2	61379.31	11.89471	0.112342	-0.09047	0.236074	0.339429
11	2	96	23.3	10.7	163434	2	71093.79	12.00416	0.169753	-0.09047	0.236074	0.486356
12	2	97	23.3	10.6	163603	2	69204.07	12.0052	0.169753	-0.09986	0.236075	0.459416
13	2	98	25	9.4	169540	3	73580.36	12.04084	0.240176	-0.22	0.64154	0.520734
14	3	93	19.6	11	102445	2.5	42412.23	11.53708	-0.00317	-0.06282	0.459218	-0.03021
15	3	94	22.2	11	129938	2.5	63149.87	11.77481	0.121392	-0.06282	0.459218	0.367867
16	3	96	24.7	11	132594	2.5	54893.92	11.79505	0.228103	-0.06282	0.459219	0.227759
17	3	95	25.4	12	134282	2.5	58681.23	11.8077	0.256049	0.024196	0.459218	0.294476
18	3	97	25.3	13.5	140581	2.5	55810.66	11.85354	0.252104	0.141979	0.459218	0.244321
19	3	98	26.1	14.5	182037	2.5	93567.02	12.11197	0.283235	0.213438	0.459218	0.761034
20	4	93	55.4	22	405042	2.5	196445.4	12.91175	1.03588	0.630332	0.459218	1.502741
21	4	94	63.5	22	489134	2.5	212773.3	13.10039	1.17234	0.630332	0.459218	1.582583
22	4	95	63.4	22	526054	2.5	267761.5	13.17316	1.170764	0.630332	0.459218	1.812453
23	4	98	68.8	23	543844	2.5	263764.3	13.20642	1.252504	0.674783	0.459218	1.797412
24	4	96	63.7	23	545834	2.5	296933.7	13.21007	1.175485	0.674783	0.459218	1.915865
25	4	97	68	23	570261	2.5	285700.8	13.25385	1.240808	0.674783	0.459218	1.877301
26	5	94	18.7	11	103142	1.5	49301.88	11.54386	-0.05018	-0.06282	-0.05161	0.120318
27	5	93	17.5	11	111642	1.5	49680.69	11.62305	-0.1165	-0.06282	-0.05161	0.127973
28	5	98	19.3	11.5	125715	1.5	55440.32	11.74177	-0.01859	-0.01836	-0.05161	0.237663
29	5	95	18.2	11	126711	1.5	61834.97	11.74966	-0.07728	-0.06282	-0.05161	0.346825
30	5	97	19.9	11.5	144745	1.5	77293.83	11.88273	0.01202	-0.01836	-0.05161	0.56997

dairy

Figure E19.1 ***Excel*** **Display of Dairy Farm Data**

To illustrate the random parameters (*RP*) estimator, we will fit a Cobb-Douglas production function,

$$\log y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{v}_i$$

where $\mathbf{x}_{it}$ = (1, log $cows_{it}$, log $land_{it}$, log $labor_{it}$, log $feed_{it}$). The data have been normalized so that the logs of the inputs sum to zero over the 1,482 observations. In the first application, we assume $\boldsymbol{\Gamma}$ is diagonal. To illustrate the difference across farms in the coefficients, we produce a centipede plot of the farm specific expected values of the coefficient on log *feed*. We also plot a kernel density estimator of the 247 observations on $E[\beta_{feed,i}|\text{data}_i]$. The second set of results is for a model in which $\boldsymbol{\Gamma}$ is unrestricted, so the parameters are freely correlated.

First, compute the random parameters regression.

```
REGRESS     ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4
            ; RPM
            ; Parameters
            ; Fcn = one(n),x1(n),x2(n),x3(n),x4(n)
            ; Pds = 6 ; Pts = 50 ; Halton $
```

We call the fifth coefficient *b*4, as it is the coefficient on *x*4 (*feed*). This picks up the estimates of $E[\beta_i 4 \mid \text{data}_i]$ and the conditional standard deviations.

```
MATRIX      ; b4 = beta_i(1:247,5:5)
            ; sb4 = sdbeta_i(1:247,5:5)
```

This forms an interval of the conditional mean plus/minus two standard deviations.

```
            ; lower = b4 - 2*sb4 ; upper = b4 + 2*sb4 $
```

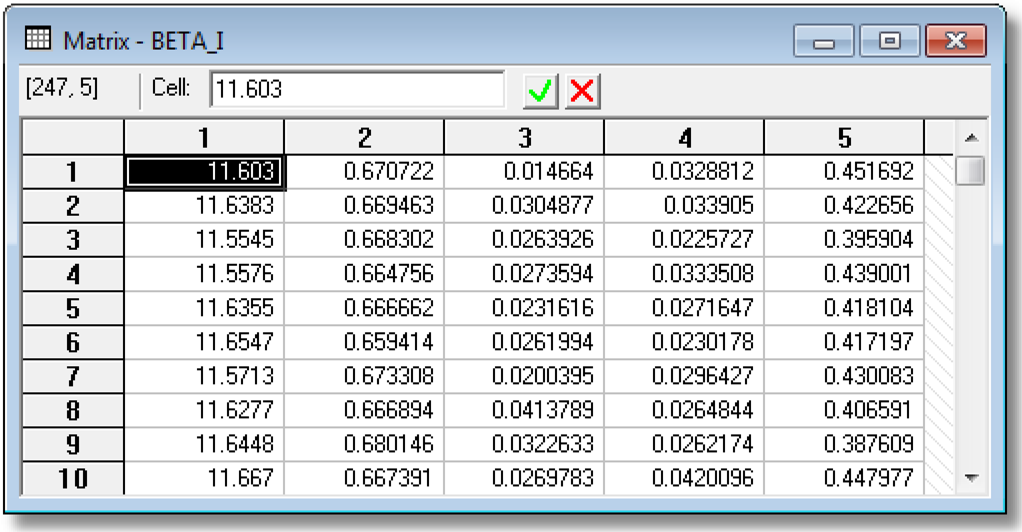
Now prepare a centipede plot.

```
CREATE      ; i = Trn(1,1)$
SAMPLE      ; 1-247 $
MATRIX      ; farmid = i $
MPLOT       ; Lhs = farmid
            ; Rhs = lower,upper
            ; Centipede ; Endpoints = 0,250
            ; Yaxis = BetaFeed
            ; Title = Farm Specific E[b_feed|data] $
```

Display a kernel density estimator.

```
CREATE      ; b4i = b4 $
KERNEL      ; Rhs = b4i
            ; Title = Kernel Density Estimator for Conditional Means of Beta(4) $
```

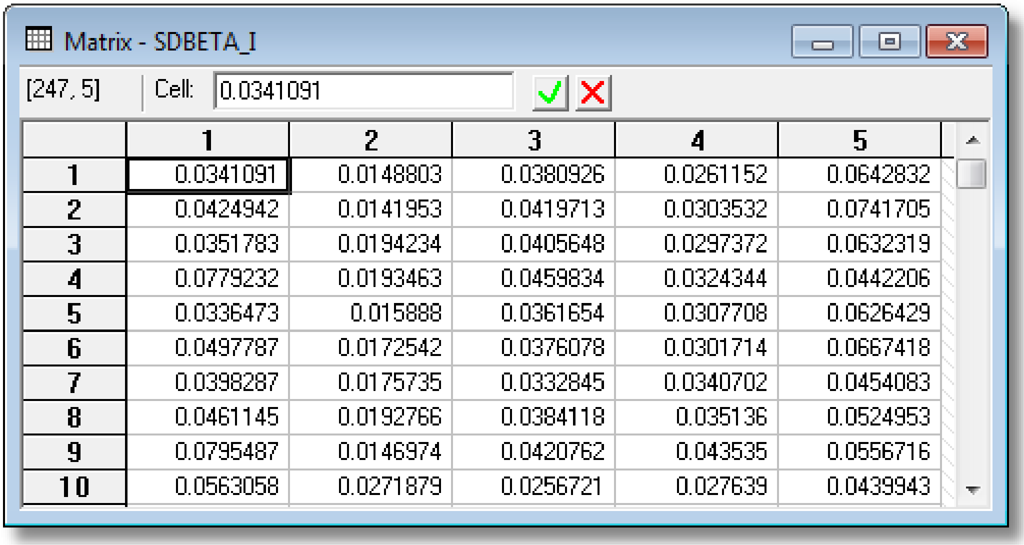
```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=YIT      Mean                 =      11.57749
             Standard deviation   =        .64344
             No. of observations  =          1482  Degrees of freedom
Regression   Sum of Squares       =       584.056           4
Residual     Sum of Squares       =       29.0957        1477
Total        Sum of Squares       =       613.152        1481
             Standard error of e  =        .14035
Fit          R-squared            =        .95255  R-bar squared =   .95242
Model test   F[  4,  1477]        =    7412.18528  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     809.67608  Akaike I.C.   = -3.92381
             Restricted (b=0)     =   -1448.90834  Bayes  I.C.   = -3.90592
             Chi squared [  4]    =    4517.16884  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    11.5775***      .00365  3175.52  .0000     11.5703   11.5846
      X1|     .59518***      .01958    30.39  .0000      .55679    .63356
      X2|     .02305**       .01122     2.05  .0400      .00105    .04505
      X3|     .02319*        .01303     1.78  .0751     -.00235    .04873
      X4|     .45176***      .01078    41.89  .0000      .43062    .47290
--------+--------------------------------------------------------------------
Random Coefficients  LinearRg Model
Dependent variable                  YIT
Log likelihood function      1330.37176
Restricted log likelihood        .00000
Chi squared [   5 d.f.]      2660.74352
Significance level               .00000
Estimation based on N =   1482, K =  11
Inf.Cr.AIC  =-2638.744 AIC/N =   -1.781
Sample is  6 pds and    247 individuals
LINEAR regression model
Simulation based on  50 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|    11.5629***      .00201  5753.98  .0000     11.5590   11.5669
      X1|     .66634***      .01067    62.45  .0000      .64543    .68725
      X2|     .02700***      .00633     4.26  .0000      .01458    .03941
      X3|     .02689***      .00720     3.74  .0002      .01278    .04100
      X4|     .38640***      .00573    67.42  .0000      .37516    .39763
        |Scale parameters for dists. of random parameters
Constant|     .10443***      .00190    54.99  .0000      .10071    .10815
      X1|     .01754***      .00420     4.18  .0000      .00931    .02578
      X2|     .04048***      .00468     8.65  .0000      .03131    .04965
      X3|     .03400***      .00590     5.76  .0000      .02243    .04557
      X4|     .07652***      .00298    25.70  .0000      .07069    .08236
        |Variance parameter given is sigma
Std.Dev.|     .07750***      .00102    76.04  .0000      .07550    .07950
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```



Matrix - BETA_I

[247, 5] Cell: 11.603

	1	2	3	4	5
1	11.603	0.670722	0.014664	0.0328812	0.451692
2	11.6383	0.669463	0.0304877	0.033905	0.422656
3	11.5545	0.668302	0.0263926	0.0225727	0.395904
4	11.5576	0.664756	0.0273594	0.0333508	0.439001
5	11.6355	0.666662	0.0231616	0.0271647	0.418104
6	11.6547	0.659414	0.0261994	0.0230178	0.417197
7	11.5713	0.673308	0.0200395	0.0296427	0.430083
8	11.6277	0.666894	0.0413789	0.0264844	0.406591
9	11.6448	0.680146	0.0322633	0.0262174	0.387609
10	11.667	0.667391	0.0269783	0.0420096	0.447977

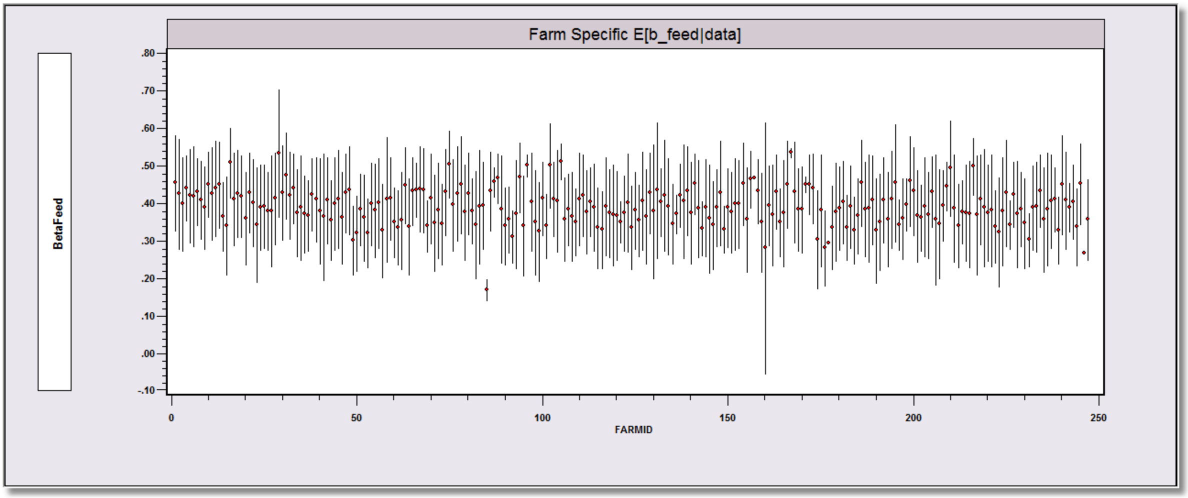


Matrix - SDBETA_I

[247, 5] Cell: 0.0341091

	1	2	3	4	5
1	0.0341091	0.0148803	0.0380926	0.0261152	0.0642832
2	0.0424942	0.0141953	0.0419713	0.0303532	0.0741705
3	0.0351783	0.0194234	0.0405648	0.0297372	0.0632319
4	0.0779232	0.0193463	0.0459834	0.0324344	0.0442206
5	0.0336473	0.015888	0.0361654	0.0307708	0.0626429
6	0.0497787	0.0172542	0.0376078	0.0301714	0.0667418
7	0.0398287	0.0175735	0.0332845	0.0340702	0.0454083
8	0.0461145	0.0192766	0.0384118	0.035136	0.0524953
9	0.0795487	0.0146974	0.0420762	0.043535	0.0556716
10	0.0563058	0.0271879	0.0256721	0.027639	0.0439943

Figure E19.2 Matrices Created by Random Parameters Regression



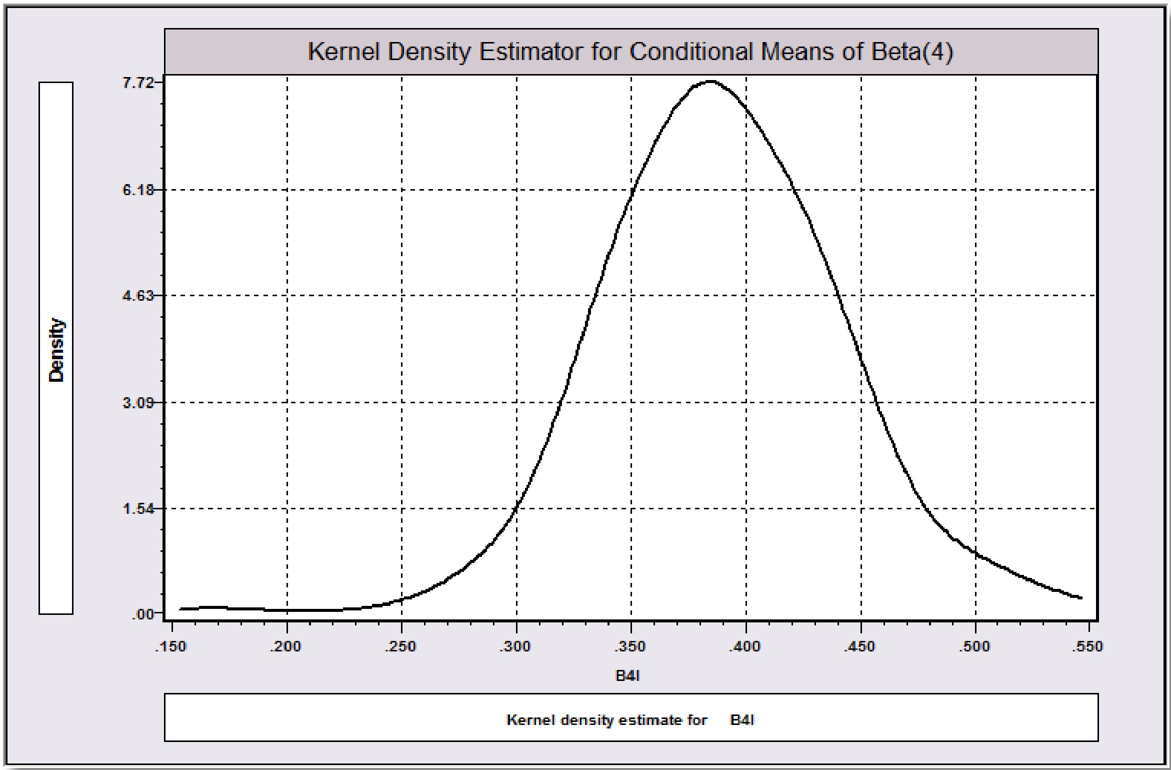


Figure E19.3 Distribution of Individual Specific Means

E19.10 The Parameter Vector and Starting Values

Starting values for the iterations are obtained by fitting the basic model without random parameters by least squares. Other parameters are set to zero. Thus, the initial results in the output for these models will be the simple linear regression model. You may provide your own starting values for the parameters with

; Start = ... the list of values for θ.

The parameter vector is laid out as follows, in this order:

$\alpha_1, ..., \alpha_K$	are the K nonrandom parameters.
$\beta_1,...,\beta_M$	are the M means of the distributions of the random parameters.
$\sigma_1,\sigma_2,...,\sigma_M$	are the M scale parameters for the distributions of the random Parameters.

These are the essential parameters. If you have specified that parameters are to be correlated, then the σs are followed by the below diagonal elements of **Γ**. (The σs are the diagonal elements.) If you have specified heterogeneity variables, **z**, then the preceding are followed by the rows of **Δ**. The autocorrelation model adds yet another vector of parameters. Consider an example: The model specifies:

; RPM = z1,z2
; Rhs = one,x1,x2,x3,x4 ? base parameters β_1, β_2, β_3, β_4, β_5
; Fcn = one(n),x2(n),x4(n)
; Cor

Then, after rearranging, the model becomes

Variable	**Parameter**
$x1$	α_1
$x3$	α_2
one	$\beta_1 + \sigma_1 v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
$x2$	$\beta_2 + \sigma_2 v_{i2} + \gamma_{21} v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
$x4$	$\beta_3 + \sigma_3 v_{i3} + \gamma_{31} v_{i1} + \gamma_{32} v_{i2} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$

and the parameter vector would be

$$\boldsymbol{\theta} = \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2, \sigma_3, \gamma_{21}, \gamma_{31}, \gamma_{32}, \delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}, \delta_{31}, \delta_{32}.$$

You may use **; Rst** and **; CML:** to impose restrictions on the parameters. Use the preceding as a guide to the arrangement of the parameter vector. (We concede the complexity of this. In point of fact, this is a complex model, unavoidably so.)

The variances of the underlying random variables are 1 for the normal distribution, 1/3 for the uniform, and 1/6 for the tent distribution. The σ parameters are only the standard deviations for the normal distribution. For the other two distributions, σ_k is a scale parameter. The standard deviation is obtained as $\sigma_k/\sqrt{3}$ for the uniform distribution and $\sigma_k/\sqrt{6}$ for the triangular distribution. When the parameters are correlated, the implied covariance matrix is adjusted accordingly. The correlation matrix is unchanged by this. All variance parameters are labeled 'scale parameter' in the model results.

E19.11 Technical Details on the RP Model

The structure of the random parameters model is based on the conditional density

$f[y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i]$ = Normal with mean $(\boldsymbol{\beta}_i'\mathbf{x}_{it})$, $i = 1,...,N$, $t = 1,...,T_i$, and variance σ^2.

NOTE: The force of the conditional normality assumption is only that the parameters are estimated by least squares.

The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) mean

$E[\boldsymbol{\beta}_i \mid \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i$, (the second term is optional – the mean may be constant)

$\text{Var}[\boldsymbol{\beta}_i \mid \mathbf{z}_i] = \boldsymbol{\Sigma}$

so that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$.

As noted earlier, the heterogeneity term, $\boldsymbol{\Delta}\mathbf{z}_i$, is optional. In addition, it may be assumed that some of the parameters are nonrandom. It is convenient to analyze the model in this fully general form here. One can easily accommodate nonrandom parameters just by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$. The actual treatment is discussed in the preceding sections.

The log likelihood function is

$$\log L = \Sigma_i \log L_i$$

where $\log L_i$ is the contribution of the ith individual (group) to the total. Conditioned on $\mathbf{v}_i$, the joint density for the ith group is

$$f[y_{i1},...,y_{iT_i} \mid \mathbf{x}_{i1,...,}\mathbf{z}_{i,}\mathbf{v}_i] = \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i)$$

$$= \left(2\pi\sigma^2\right)^{-T_i/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{t=1}^{T_i} \left(y_{it} - \boldsymbol{\beta}_i'\mathbf{x}_{it}\right)^2\right]$$

Since $\mathbf{v}_i$ is unobserved, it is necessary to obtain the unconditional log likelihood by taking the expectation of this over the distribution of $\mathbf{v}_i$. For convenience, write the tth term in the density above as $f(y_{it}, \boldsymbol{\beta}_i'\mathbf{x}_{it})$, so that

$$L_i \mid \mathbf{v}_i \;=\; \prod_{t=1}^{T_i} f(y_{it}, \boldsymbol{\beta}_i'\mathbf{x}_{it}).$$

Then,

$$L_i \;=\; \mathbf{E}_{\mathbf{v}i}\,[L_i \mid \mathbf{v}_i] \;=\; \int_{\text{Range of } \mathbf{v}_i} g(\mathbf{v}_i) \prod_{t=1}^{T_i} f(y_{it}, \boldsymbol{\beta}_i'\mathbf{x}_{it})\, d\mathbf{v}_i$$

(Note that this is a multivariate integral.) Then, finally,

$$\log \mathrm{L} \;=\; \sum_{i=1}^{N} \log L_i.$$

For convenience in what follows, let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma})$. The likelihood function is maximized by solving the likelihood equations:

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} \;=\; \sum_{i=1}^{N} \frac{\partial \log L_i}{\partial \boldsymbol{\theta}} \;=\; \mathbf{0}.$$

Note that $\boldsymbol{\Gamma}$ is a lower triangular matrix; $\boldsymbol{\theta}$ is understood to contain only the nonzero elements, moving rowwise through the matrix (one element in row one, two in row two, and so on). Estimation is done conditionally on an estimate of σ^2. This is described below.

The integration is done by Monte Carlo simulation. In general, we use the approximation strategy:

$$\mathbf{E}_{\mathbf{v}i}\,[L_i \mid \mathbf{v}_i] \;\approx\; \frac{1}{R} \sum_{r=1}^{R} L \mid \mathrm{v}_{ir},$$

where $\mathbf{v}_{ir}$ is a random draw from the distribution of $\mathbf{v}_i$. See Brownstone and Train (1999), Train (1998), and Revelt and Train (1998) for discussion. The approximation improves with increased R (this is under your control) and with increases in N, though the simulation variance which decreases with increases in R does not decrease with N.

Collecting terms, then, the log likelihood is computed with

$$\log L \;\approx\; \sum_{i=1}^{N} \log\left\{\frac{1}{R}\sum_{r=1}^{R}\left[\prod_{t=1}^{T_i} f(y_{it}, \boldsymbol{\beta}_{ir}'\mathbf{x}_{it})\right]\right\},$$

where

$$\boldsymbol{\beta}_{ir} \;=\; \boldsymbol{\beta} \;+\; \boldsymbol{\Delta}\mathbf{z}_i \;+\; \boldsymbol{\Gamma}\mathbf{v}_{ir}.$$

The derivatives of the log likelihood function are approximated as well.

$$\frac{\partial \log L_i}{\partial \boldsymbol{\theta}} = \frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}$$

$$\frac{\partial L_i}{\partial \boldsymbol{\theta}} = \int_{\text{Range of } \mathbf{v}_i} g(\mathrm{v}_i)\frac{\partial}{\partial \boldsymbol{\theta}}\prod_{t=1}^{T_i} f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{it})\, d\mathrm{v}_i$$

$$\frac{\partial}{\partial \boldsymbol{\theta}}\prod_{t=1}^{T_i} f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{it}) = \sum_{t=1}^{T_i}\left[\frac{\partial f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{it})}{\partial \boldsymbol{\theta}}\right]\prod_{s\neq t} f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{is})$$

$$= \prod_{t=1}^{T_i} f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{it})\sum_{t=1}^{T_i}\left[\frac{\partial \log f(y_{it},\boldsymbol{\beta}_i'\mathbf{x}_{it})}{\partial \boldsymbol{\theta}}\right]$$

Collecting terms once again, we obtain the approximation,

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N} \frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}$$

$$\approx \sum_{i=1}^{N} \frac{\left\{\frac{1}{R}\sum_{r=1}^{R}\left[\prod_{t=1}^{T_i} f(y_{it},\boldsymbol{\beta}_{ir}'\mathbf{x}_{it})\right]\left[\sum_{t=1}^{T_i}\frac{\partial \log f(y_{it},\boldsymbol{\beta}_{ir}'\mathbf{x}_{it})}{\partial \boldsymbol{\theta}}\right]\right\}}{\left\{\frac{1}{R}\sum_{h=1}^{H}\left[\prod_{t=1}^{T_i} f(y_{it},\boldsymbol{\beta}_{ir}'\mathbf{x}_{it})\right]\right\}}$$

Note that L_i and its derivatives are approximated separately. The index is

$$\begin{aligned} w_{irt} &= \boldsymbol{\beta}_{ir}'\mathbf{x}_{it} \\ &= \boldsymbol{\beta}'\mathbf{x}_{it} + \mathbf{z}_i'\boldsymbol{\Delta}'\mathbf{x}_{it} + \mathbf{v}_{ir}'\boldsymbol{\Gamma}'\mathbf{x}_{it} \end{aligned}$$

We will need
$$\frac{\partial w_{irt}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{x}_{it} \\ \mathbf{z}_i \otimes \mathbf{x}_{it} \\ \mathbf{v}_{ir} \otimes \mathbf{x}_{it} \end{bmatrix} = \mathbf{h}_{irt}$$

Then,
$$\frac{\partial \log f(y_{it}, w_{irt})}{\partial \boldsymbol{\theta}} = [(y_{it} - w_{irt})/\sigma^2]\mathbf{h}_{irt} = \mathbf{g}_{irt}.$$

In the vector at the end of the expression, the lower term is the result of the term $\mathbf{x}_{it}'\boldsymbol{\Gamma}\mathbf{v}_{ir}$. Since $\boldsymbol{\Gamma}$ is a lower triangular matrix, this term actually involves the $K(K+1)/2$ terms that are nonzero in the matrix $\boldsymbol{\Gamma}$.

The estimate of σ^2 is obtained residually while the estimates of the other parameters are obtained by maximizing the likelihood. The initial estimator of σ^2 is the ordinary least squares estimator. The likelihood function above is then maximized conditionally on this estimate of σ^2. After convergence, σ^2 is reestimated with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \frac{1}{R}\sum_{r=1}^{R} \sum_{t=1}^{T_i} \left(y_{it} - \hat{\boldsymbol{\beta}}'_{ir}\mathbf{x}_{it}\right)^2}{\sum_{i=1}^{N} T_i}$$

The parameters of the model are then reestimated using this estimate of σ^2. After convergence, σ^2 is recomputed again and the iterations are entered a third time. This process continues until σ^2 stabilizes, which will usually occur in only a few passes. After this last estimation, σ^2 is recomputed, and this is the value reported in the results.

The Hessian is fairly complicated, so we will only sketch the necessary components. Let

$$f_{ir} = \prod_{t=1}^{T_i} f_{irt}$$

$$\mathbf{g}_{ir} = \sum_{t=1}^{T_i} \mathbf{g}_{irt}$$

Then,

$$\frac{\partial \log L_i}{\partial \boldsymbol{\theta}} = \frac{1}{L_i}\frac{1}{R}\sum_{r=1}^{R} f_{ir}\mathbf{g}_{ir} = \frac{1}{L_i}\mathbf{g}_i$$

Since each of the three parts is a function of $\boldsymbol{\theta}$, the Hessian will have three parts. The end result is

$$\frac{\partial^2 \log L_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -\left(\frac{1}{L_i}\mathbf{g}_i\right)\left(\frac{1}{L_i}\mathbf{g}_i\right)' + \frac{1}{L_i}\frac{1}{R}\sum_{r=1}^{R} f_{ir}\mathbf{H}_{ir} + \frac{1}{L_i}\frac{1}{R}\sum_{r=1}^{R} f_{ir}\mathbf{g}_{ir}\mathbf{g}_{ir}'$$

where

$$\mathbf{H}_{ir} = \sum_{t=1}^{T_i} \frac{\partial^2 \log f_{irt}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$$

The asymptotic covariance matrix may be estimated by the BHHH estimator,

$$\textbf{BHHH} = \left[\sum_{i=1}^{N} \left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)\left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)'\right]^{-1}$$

or with the actual second derivatives.

The remaining detail concerns the random draws, $\mathbf{v}_i$, which are discussed in Chapter R24.

E20: Latent Class Linear Models

E20.1 Introduction

LIMDEP provides two approaches to fitting linear regression models with random parameters:

- Mixed, or random parameters model – parameters are distributed continuously;
- Latent class, or finite mixture model – parameters have a discrete distribution.

The models are built around the structural equations

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it},\ i = 1,...,N,\ t = 1,\ldots,T_i,$$
$$\varepsilon_{\text{it}} \sim N[0,\sigma_i^2]$$

For the mixed, or random parameters model,

$$\begin{aligned}
\boldsymbol{\beta}_i &= \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i,\\
\text{E}[\mathbf{v}_i|\mathbf{z}_i,\mathbf{x}_{it}] &= \mathbf{0},\ \text{Var}[\mathbf{v}_i] = \mathbf{I},\\
\text{Var}[\boldsymbol{\beta}_i|\mathbf{z}_i,\mathbf{x}_{it}] &= \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}',\\
\sigma_i^2 &= \sigma^2 \text{ (constant)}.
\end{aligned}$$

In the latent class model

$$\begin{aligned}
(\boldsymbol{\beta}_j,\sigma_j^2) &\in [(\boldsymbol{\beta}_1,\sigma_1^2), (\boldsymbol{\beta}_2,\sigma_2^2), \ldots, (\boldsymbol{\beta}_J,\sigma_J^2)],\\
\text{Prob}[\text{class}=j|\mathbf{z}_i] &= \pi_j(\mathbf{z}_i,\boldsymbol{\theta}_j),\ j = 1,\ldots,J.
\end{aligned}$$

The models apply naturally to panel data but can be used (somewhat less effectively) with cross sections as well. The mixed model is estimated by maximum simulated likelihood. The latent class model is estimated by maximum likelihood. The random parameters linear model is developed in Chapter E19. This chapter documents how to fit a latent class linear model.

E20.2 Latent Class Linear Regression Model

A linear regression model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is specified in terms of the density,

$$f(y_{it} \mid \mathbf{x}_{it}) = f(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it}) = \phi(i,t).$$

(We allow for the cross section case of $T_i = 1$.) For this special case of the linear regression model, we assume that the underlying distribution is normal with mean $\boldsymbol{\beta}'\mathbf{x}_{it}$ and variance σ^2. Henceforth, we use the term ‘group’ to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. Unobserved heterogeneity in the distribution of y_{it} is assumed to impact the density in the form of a random effect. The continuous distribution of the heterogeneity such as discussed in the preceding chapter is approximated by using a finite number of ‘points of support.’ The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J 'classes' with a model which allows for heterogeneity as follows: The probability of observing y_{it} given that regime j applies is

$$\phi(i,t|j) = \phi(y_{it}|\, \mathbf{x}_{it}, j),$$

where the normal density is now specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$\phi(i,t|j) = \phi[y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta_j], \text{Prob[class} = j] = F_j$$

We formulate this approximation more generally as,

$$\phi(i,t|j) = \phi[y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}_j'\mathbf{x}_{it}], F_j = \exp(\theta_j) / \Sigma_j \exp(\theta_j), \text{ with } \theta_J = 0.$$

In this formulation, each group has its own parameter vector, $\boldsymbol{\beta}_j' = \boldsymbol{\beta} + \boldsymbol{\delta}_j$, though the variables that enter the mean are assumed to be the same. (We show how to modify this assumption in Section E20.4.) In sum, then, for this application, the model is

$$f(y_{it} \,|\, \text{class} = \text{j}) = \text{N}[\boldsymbol{\beta}_j'\mathbf{x}_{it}, \sigma_j^2], \text{Prob(class} = \text{j}) = F_j = \exp(\theta_j) / \Sigma_j \exp(\theta_j).$$

Thus, the within class model is the linear regression model with normally distributed disturbances. Thus far, it is assumed that the prior class probabilities are constants, π_j. In Section E20.5, we detail how to introduce covariates into the class probabilities.

E20.3 Command for Latent Class Regression

The estimation command for this model is

REGRESS ; Lhs = ...
; Rhs = independent variables
; LCM (for latent class model)
; Panel
; Pts = the number of classes $

As noted, this model can be (and often is) applied to cross section data. Thus, you may omit the **; Panel** in the command, in which case it is assumed that $T_i = 1$. The default number of support points is five. But, this is fairly high. You may set J to 2, 3, 4, 5, 6, 7, 8, or 9 with

; Pts = the value you wish

Other options and further details on the model appear in Chapter R25. The latent class model provides estimates of the J class member parameter vectors for the model and the class probabilities.

Estimates retained by this model include:

Matrices: *b* = full parameter vector, [$\boldsymbol{\beta}_1'$, $\boldsymbol{\beta}_2'$,... F_1,...,F_J]
varb = full covariance matrix
Note that *b* and *varb* involve $J \times (K+1)$ estimates.

Three additional matrices are created,

b_class = a $J \times K$ matrix with each row equal to the corresponding $\boldsymbol{\beta}_j$
class_pr = a $J \times 1$ vector containing the estimated class probabilities
beta_i = individual conditional (posterior) expectations of β_i

Scalars: *kreg* = number of variables in Rhs list
nreg = total number of observations used for estimation
logl = maximized value of the log likelihood function
exitcode = exit status of the estimation procedure.

Last Function: None

E20.4 Restricted Models

There are several interesting special cases of the latent class linear regression – these will be extended to latent class models generally in the development of other models. Restrictions can be imposed on the coefficients of the LC model, both within class and across classes. The **; Rst = list** specification is used for this purpose. The parameters of the LC linear regression model, in the order in which they appear in the program, are

$$\Theta = \beta_1,\sigma_1,\ \beta_2,\sigma_2,\ \ldots,\ \beta_J,\sigma_J,\ \theta_1,\ \theta_2,\ \ldots,\ \theta_J$$

Depending on the model, β_j may have 1, 2, up to K elements. Note that the variance parameter is σ_j, not σ_j^2. The last *J* parameters are the structural parameters in the class probabilities. There are *J* of these, though the last one equals zero. The list of items in **; Rst = list** provides either symbols or values for the elements in Θ. Equality restrictions are imposed by using the same name. Fixed values are imposed by placing the fixed value in the list. For an example, consider a three class model with four regressors, so that the command is

REGRESS ; LCM ; Lhs = y ; Rhs = one,x1,x2,x3,x4 ; Pts = 3 $

The unrestricted model would be specified (redundantly) by

; Rst = a1,b11,b12,b13,b14,sg1, a2,b21,b22,b23,b24,sg2, a3,b31,b32,b33,b34,sg3, t1,t2,t3.

(Note, we chose the symbols *aj* and *bjk* purely for convenience and clarity. You may use any symbols you like.) This **; Rst** specification does not impose any constraints. Suppose it were desired to force the coefficient on x4 to be the same in all three classes. The list is changed to

; Rst = a1,b11,b12,b13,b4,sg1, a2,b21,b22,b23,b4,sg2, a3,b31,b32,b33,b4,sg3, t1,t2,t3.

Note that *b*14, *b*24 and *b*34 have all been changed to *b*4. Since the name is the same in all three sets of symbols, this will impose the constraint that the parameter is equal in all three places. Second, suppose, in addition to the equality constraint on *bj*4, we wished to fix *b*11 at zero and *b*21 at one. We would use

; Rst = a1,0,b12,b13,b4,sg1, a2,1,b22,b23,b4,sg2, a3,b31,b32,b33,b4,sg3, t1,t2,t3.

You may impose any number of equality and fixed value constraints with this device. Note, however,

- Although you provide a place holder for σ_j, you should generally not constrain these parameters. (It is allowed by the specification, but it will likely lead to very poor results.)

- You must provide place holders for the θ parameters, but you should never constrain these. Note, as well that although the third will be constrained to equal zero, the program will do this. You should treat this parameter as unconstrained.

Heckman and Singer

Heckman and Singer's specification is obtained by forcing all classes to have the same coefficients save for the constant terms. By way of the preceding example, this is specified as follows:

; Rst = a1,b1,b2,b3,b4,sg, a2,b1,b2,b3,b4,sg, a3,b1,b2,b3,b4,sg, t1,t2,t3.

This produces a random effects regression in which the effect has a discrete distribution. The implied random effects model is

$$y_{it|j} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + \delta_j,\ j = 1,2,3$$

$$P(\delta = \delta_j) = \pi_j.$$

$$E[\delta] = 0.$$

More generally, the Heckman and Singer formulation is obtained by forcing all coefficients in the classes to be equal save for a class specific constant term.

Exclusions

There are cases in which the analyst believes a priori that different models apply to the different classes. We will examine a number of such cases in applications in later chapters. One of these cases relates to the prior information (belief) that certain variables do not appear in the model in certain classes. (Note that this implies that the classes are not completely latent.) This type of specification can be obtained by imposing zero restrictions. Zeros and other fixed values may be placed wherever it is desired in the list, though we emphasize once again, fixed value restrictions on the disturbance standard deviations generally produce undesirable results.

E20.5 Modeling Class Probabilities

The prior probabilities of class membership, π_1, π_2,…, π_J are estimated with the model parameters. In order to impose the constraints $\pi_j > 0$ and $\Sigma_j\pi_j = 1$, the probabilities are parameterized with a multinomial logit form,

$$\pi_j = \frac{\exp(\theta_j)}{\Sigma_{j=1}^J \exp(\theta_j)},\ \theta_J = 0.$$

(The constraint on the last θ is imposed because only J-1 parameters are needed to specify the J probabilities. The last probability is one minus the sum of the first J-1.)

The prior probabilities may be extended to depend on variables in the data set. For example, in a typical application, the prior probabilities are often made a function of demographics such as age or gender. The expanded model is

$$\pi_j(\mathbf{z}_i) = \frac{\exp(\boldsymbol{\theta}'_j\mathbf{z}_i)}{\Sigma_{j=1}^J \exp(\boldsymbol{\theta}'_j\mathbf{z}_i)},\ \boldsymbol{\theta}_J = \mathbf{0}.$$

Variables $\mathbf{z}_i$ are added to the model by specifying

; LCM = the list of variables (must not include one)

For example,

; LCM = age,sex

specifies a model in which age and sex enter the class probabilities. An application appears in Section E20.7.

E20.6 Posterior Class Probabilities and Predicting Class Membership

After estimation of the model parameters, a secondary exercise is estimation of the posterior probabilities,

$$\text{Prob}[\text{class}=j|\{(y_{it},\mathbf{x}_{it}),t=1,\ldots,T_i\},\mathbf{z}_i],$$

which we denote $P(j|i)$. To derive $P_j(j|i)$, use Bayes theorem as follows: The probability that individual is a member of class j given the information in the sample about them is denoted $P(j|i)$. The joint density of the class membership and the observed outcome is denoted $P(i,j)$. By definition,

$$\text{P}(j|i) = P(i,j)/P(i).$$

The joint density of the outcome and the class membership is the product of the conditional times the marginal, and the marginal has already been defined as the prior probability, π_j;

$$P(i,j) = P(i|j)\pi_j.$$

$P(i|j)$ is the density for individual i given they are in class j, which is the contribution of individual i to the likelihood function given class j, $f(i|j) = \Pi_t f(y_{it}|j)$. By definition, the marginal density is the sum of the joint densities, so that the unconditional density for individual i is

$$P(i) = \Sigma_j P(i,j) = \Sigma_j P(i|j)\pi_j.$$

Collecting terms, we find the posterior probabilities,

$$P(j\,|\,i) = \frac{\pi_j\left(\Pi_{t=1}^{T_i} f(y_{it}\,|\,j)\right)}{\Sigma_{j=1}^{J}\pi_j\left(\Pi_{t=1}^{T_i} f(y_{it}\,|\,j)\right)}.$$

As noted earlier, the prior probabilities may involve covariates in $\pi_j(\mathbf{z}_i)$. The marginal densities in the products are the normal densities with mean $\boldsymbol{\beta}_j'\mathbf{x}_{it}$ and standard deviations σ_j. Assembling all the parts, then,

$$P(j\,|\,i) = \frac{\pi_j(\mathbf{z}_i)\prod_{t=1}^{T_i}\frac{1}{\sigma_j}\phi\left(\frac{y_{it}-\boldsymbol{\beta}_j'\mathbf{x}_{it}}{\sigma_j}\right)}{\sum_{j=1}^{J}\pi_j(\mathbf{z}_i)\prod_{t=1}^{T_i}\frac{1}{\sigma_j}\phi\left(\frac{y_{it}-\boldsymbol{\beta}_j'\mathbf{x}_{it}}{\sigma_j}\right)}.$$

The posterior probabilities will embody the model estimates and the sample information about the individual. A natural next step is to use the posterior class probabilities to predict the class membership. We predict (albeit imperfectly) that individual i is a member of class j if $P(j|i) > P(m|i)$ for all other m – i.e., we predict the class with the highest posterior probability.

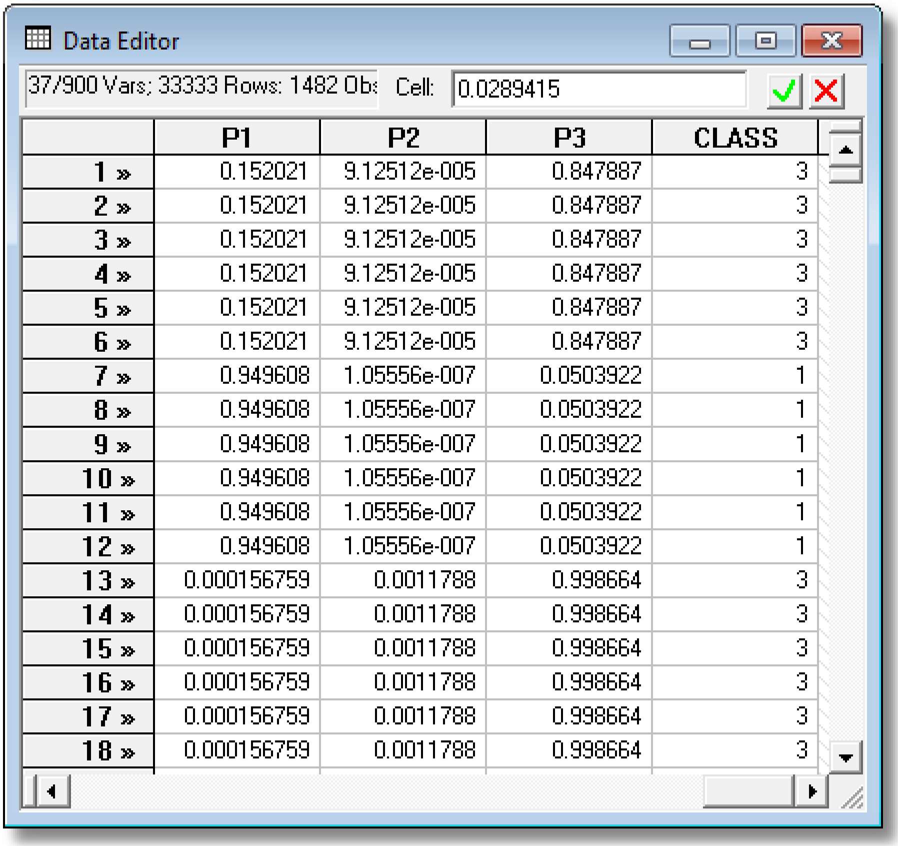
Posterior probabilities and the class predictions can be retained in the data set as follows: For the probabilities, it is necessary to create (or provide) a set of J existing variables, in a namelist, in the REGRESS command, using

; Classp = the namelist.

For an example, to extend the earlier example, we used

```
CREATE      ; p1 = 0 ; p2 = 0 ; p3 = 0 $
NAMELIST    ; cp = p1,p2,p3$
REGRESS     ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4
            ; LCM
            ; Pts = 3
            ; Parameters
            ; Pds = 6
            ; Classp = cp $
```

This produces new variables in the data area, as shown in Figure E20.1.



Data Editor

37/900 Vars; 33333 Rows: 1482 Obs Cell: 0.0289415

	P1	P2	P3	CLASS
1 »	0.152021	9.12512e-005	0.847887	3
2 »	0.152021	9.12512e-005	0.847887	3
3 »	0.152021	9.12512e-005	0.847887	3
4 »	0.152021	9.12512e-005	0.847887	3
5 »	0.152021	9.12512e-005	0.847887	3
6 »	0.152021	9.12512e-005	0.847887	3
7 »	0.949608	1.05556e-007	0.0503922	1
8 »	0.949608	1.05556e-007	0.0503922	1
9 »	0.949608	1.05556e-007	0.0503922	1
10 »	0.949608	1.05556e-007	0.0503922	1
11 »	0.949608	1.05556e-007	0.0503922	1
12 »	0.949608	1.05556e-007	0.0503922	1
13 »	0.000156759	0.0011788	0.998664	3
14 »	0.000156759	0.0011788	0.998664	3
15 »	0.000156759	0.0011788	0.998664	3
16 »	0.000156759	0.0011788	0.998664	3
17 »	0.000156759	0.0011788	0.998664	3
18 »	0.000156759	0.0011788	0.998664	3

Figure E20.1 Estimated Class Probabilities

Note that the posterior probabilities will differ substantially from the priors. In the model estimated above, the prior probabilities are 0.323, 0.185 and 0.492. Second, note that the probabilities are repeated for each observation in the group in a panel.

Finally, you may request the class assignments to be saved as a variable by adding

; Group = variable name

to create the new variable. The variable will contain the index of the class with the largest posterior probability. The results in the last column in Figure E20.1 are obtained with

; Group = class

E20.7 Applications

We examine two applications. The first purely for illustration, considers the classic 'mixture of normals' application. The second applies the methods to a regression model using the panel data on dairy farm production that was used in Chapter E19 for the mixed regression model.

E20.7.1 Finite Mixture of Normals

The mixture of normals is simply a latent class application to the marginal distribution of a variable. We treat it here by treating the marginal normal distribution as a regression model in which there are no regressors, only a constant term. The first result below is the base case. The second specifies that the prior class probabilities depend on gender. The very large increase in the log likelihood suggests that gender is, indeed, relevant in the class probabilities.

```
REJECT      ; hhninc = 0 $
CREATE      ; loginc = log(hhninc) $
REGRESS     ; Lhs = loginc ; Rhs = one ; Pts = 2 ;  LCM $
REGRESS     ; Lhs = loginc ; Rhs = one ; Pts = 2 ;  LCM = female $
```

```
-----------------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Dependent variable               LOGINC
Log likelihood function    -18604.87698
Sample is  1 pds and  27322 individuals
Model fit with  2 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGINC| Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Model parameters for latent class 1
Constant|   -1.43704***      .02249   -63.90  .0000    -1.48112  -1.39297
   Sigma|     .76606***      .00808    94.76  .0000      .75022    .78191
        | Model parameters for latent class 2
Constant|   -1.10721***      .00349  -317.55  .0000    -1.11404  -1.10037
   Sigma|     .40362***      .00344   117.22  .0000      .39688    .41037
        | Estimated prior probabilities for class membership
Class1Pr|     .15235***      .00929    16.40  .0000      .13414    .17056
Class2Pr|     .84765***      .00929    91.22  .0000      .82944    .86586
--------+--------------------------------------------------------------------
Log likelihood function    -18545.67813
--------+--------------------------------------------------------------------
        | Model parameters for latent class 1
Constant|   -1.43965***      .02117   -67.99  .0000    -1.48115  -1.39814
   Sigma|     .74502***      .00730   102.04  .0000      .73071    .75933
        | Model parameters for latent class 2
Constant|   -1.10150***      .00352  -313.15  .0000    -1.10839  -1.09461
   Sigma|     .40057***      .00341   117.43  .0000      .39389    .40726
        | Estimated prior probabilities for class membership
   ONE_1|   -2.06919***      .08969   -23.07  .0000    -2.24498  -1.89340
FEMALE_1|     .82416***      .07820    10.54  .0000      .67090    .97743
   ONE_2|        0.0    .....(Fixed Parameter).....
FEMALE_2|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
```

E20.7.2 Latent Class Linear Model

```
REGRESS   ; Lhs = yit
          ; Rhs = one,x1,x2,x3,x4
          ; LCM
          ; Pts = 3
          ; Parameters
          ; Pds = 6  $
```

```
-----------------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Log likelihood function      1243.78697
Restricted log likelihood        .00000
Chi squared [  15 d.f.]      2487.57394
(Some results omitted)
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    11.7014***      .00373  3140.56  .0000     11.6941   11.7087
      X1|     .57307***      .01464    39.14  .0000      .54438    .60177
      X2|     .08533***      .01076     7.93  .0000      .06425    .10642
      X3|     .03306**       .01455     2.27  .0231      .00455    .06157
      X4|     .42682***      .00666    64.06  .0000      .41376    .43988
   Sigma|     .08729***      .00210    41.58  .0000      .08317    .09140
        |Model parameters for latent class 2
Constant|    11.3944***      .00740  1538.86  .0000     11.3799   11.4089
      X1|     .78593***      .03271    24.03  .0000      .72182    .85005
      X2|    -.06285***      .01633    -3.85  .0001     -.09486   -.03085
      X3|     .06089**       .02829     2.15  .0314      .00544    .11635
      X4|     .35185***      .01699    20.71  .0000      .31855    .38515
   Sigma|     .11316***      .00350    32.32  .0000      .10630    .12003
        |Model parameters for latent class 3
Constant|    11.5622***      .00312  3706.84  .0000     11.5561   11.5683
      X1|     .65425***      .01704    38.39  .0000      .62084    .68765
      X2|     .05083***      .00909     5.59  .0000      .03302    .06863
      X3|     .05779***      .00953     6.06  .0000      .03911    .07648
      X4|     .40208***      .00913    44.05  .0000      .38419    .41997
   Sigma|     .08492***      .00240    35.34  .0000      .08021    .08963
        |Estimated prior probabilities for class membership
Class1Pr|     .32285***      .03337     9.67  .0000      .25744    .38826
Class2Pr|     .18473***      .02723     6.78  .0000      .13135    .23810
Class3Pr|     .49243***      .03657    13.47  .0000      .42075    .56410
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The Heckman and Singer model is obtaining by a set of restrictions – all parameters are the same across classes save for the constant terms.

```
REGRESS    ; Lhs = yit
           ; Rhs = one,x1,x2,x3,x4
           ; LCM
           ; Pts = 3
           ; Parameters
           ; Pds = 6
           ; Rst = a1, b1, b2, b3, b4, sg,
                   a2, b1, b2, b3, b4, sg,
                   a3, b1, b2, b3, b4, sg,
                   t1, t2, t3 $
```

```
-----------------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Dependent variable                  YIT
Log likelihood function      1217.07146
(Some results omitted)
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    11.7047***      .00393  2981.71  .0000     11.6970   11.7124
      X1|     .64523***      .00998    64.63  .0000      .62566    .66479
      X2|     .03432***      .00654     5.25  .0000      .02151    .04714
      X3|     .05467***      .00868     6.30  .0000      .03765    .07168
      X4|     .40610***      .00502    80.93  .0000      .39626    .41593
   Sigma|     .09245***      .00131    70.81  .0000      .08989    .09500
        |Model parameters for latent class 2
Constant|    11.3728***      .00395  2878.66  .0000     11.3651   11.3806
      X1|     .64523***      .00998    64.63  .0000      .62566    .66479
      X2|     .03432***      .00654     5.25  .0000      .02151    .04714
      X3|     .05467***      .00868     6.30  .0000      .03765    .07168
      X4|     .40610***      .00502    80.93  .0000      .39626    .41593
   Sigma|     .09245***      .00131    70.81  .0000      .08989    .09500
        |Model parameters for latent class 3
Constant|    11.5587***      .00297  3894.50  .0000     11.5529   11.5645
      X1|     .64523***      .00998    64.63  .0000      .62566    .66479
      X2|     .03432***      .00654     5.25  .0000      .02151    .04714
      X3|     .05467***      .00868     6.30  .0000      .03765    .07168
      X4|     .40610***      .00502    80.93  .0000      .39626    .41593
   Sigma|     .09245***      .00131    70.81  .0000      .08989    .09500
        |Estimated prior probabilities for class membership
Class1Pr|     .31497***      .03404     9.25  .0000      .24825    .38169
Class2Pr|     .14634***      .02459     5.95  .0000      .09813    .19454
Class3Pr|     .53869***      .03574    15.07  .0000      .46865    .60874
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This application imposes some zero restrictions on the model within the classes.

```
REGRESS ; Lhs = yit
        ; Rhs = one,x1,x2,x3,x4
        ; LCM
        ; Pts = 3
        ; Parameters
        ; Pds = 6
        ; Rst = a1, b11, b12, b13, 0, sg1,
                a2, b21, b22, 0,   0, sg2,
                a3, 0, b32, b33, b34, sg3,
                t1,t2,t3 $
```

```
-----------------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Log likelihood function        860.37424
(Some results omitted)
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
     YIT|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    11.6030***      .00401  2893.08  .0000     11.5952   11.6109
      X1|    1.17538***      .00967   121.52  .0000     1.15642   1.19433
      X2|     .04902***      .01023     4.79  .0000      .02897    .06907
      X3|     .10212***      .01469     6.95  .0000      .07333    .13091
      X4|        0.0    .....(Fixed Parameter).....
   Sigma|     .11270***      .00369    30.57  .0000      .10548    .11993
        |Model parameters for latent class 2
Constant|    11.3406***      .00875  1296.45  .0000     11.3235   11.3578
      X1|    1.27066***      .02381    53.36  .0000     1.22398   1.31733
      X2|    -.00187         .02193     -.09  .9322     -.04484    .04111
      X3|        0.0    .....(Fixed Parameter).....
      X4|        0.0    .....(Fixed Parameter).....
   Sigma|     .14905***      .00590    25.26  .0000      .13749    .16062
        |Model parameters for latent class 3
Constant|    11.6208***      .00318  3656.13  .0000     11.6146   11.6271
      X1|        0.0    .....(Fixed Parameter).....
      X2|     .19726***      .01007    19.59  .0000      .17752    .21699
      X3|     .14742***      .01130    13.05  .0000      .12527    .16957
      X4|     .70111***      .00548   127.87  .0000      .69036    .71185
   Sigma|     .11208***      .00257    43.65  .0000      .10705    .11711
        |Estimated prior probabilities for class membership
Class1Pr|     .37368***      .03523    10.61  .0000      .30462    .44273
Class2Pr|     .18052***      .02705     6.67  .0000      .12750    .23354
Class3Pr|     .44580***      .03701    12.04  .0000      .37326    .51835
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

E20.8 Technical Details and the EM Algorithm

Details on estimation of the latent class model are provided in Chapter E26. The estimates are computed by directly maximizing the log likelihood function. For the latent class linear regression model, the log likelihood is

$$\log L(\Theta) = \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{J} \frac{\exp(\theta_j' \mathbf{z}_i)}{\Sigma_{m=1}^{J} \exp(\theta_m' \mathbf{z}_i)} \prod_{t=1}^{T_i} \frac{1}{\sigma_j} \phi\left(\frac{y_{it} - \boldsymbol{\beta}_j' \mathbf{x}_{it}}{\sigma_j} \right) \right\}.$$

We note, the EM algorithm has gained some attention in the recent literature. The algorithm provides a method of maximizing the log likelihood via the following iteration:

Step 0. Enter with initial values of $\boldsymbol{\theta}_j$ and $(\boldsymbol{\beta}_j,\sigma_j)$.

Step 1. Compute new estimates of the posterior probabilities derived in Section E20.6. This provides a different n set of weights, $P(j|i)$ for each of the j classes.

Step 2. Using the weights, in each class separately, maximize the log likelihood function. For the linear regression model, this means compute weighted least squares estimates of β_j, followed by a weighted sum of squared residuals estimate of σ_j. This step will involve J such weighted least squares regressions to produce the set of J vectors $(\mathbf{b}_j,s_j)$. Note that given the weights that are constant within the groups, this regression pools the panel data.

Step 3. Return to Step 1 or exit if the estimates have stopped changing.

The EM algorithm, which is not used here, has advantages and disadvantages. In its favor, it is very stable; each step goes uphill. One disadvantage is that it usually takes many iterations – it is slow. In addition, unlike the direct MLE, the EM method does not produce an estimator of the covariance matrix. That must be obtained ex post, after estimation is completed. Contrary to impressions suggested elsewhere, EM is not a model; it is an algorithm. It does not produce different results from direct maximization of the log likelihood. For our purposes, a significant disadvantage is that the EM method does not allow the sort of restricted model construction developed in Section E20.4, for example, the Heckman and Singer model.

E21: Single Equation Instrumental Variables Estimation

E21.1 Introduction

This chapter will present several IV estimators for linear and nonlinear single equation regression models. The models considered are

$$g(y_i) = f(\mathbf{x}_i,\boldsymbol{\beta},\boldsymbol{\theta}) + \varepsilon_i$$

in which $g(\bullet)$ and $f(\bullet,\bullet)$ are continuous functions and ε is a disturbance with zero mean. The extension here is estimators that are consistent when $\text{Cov}(\mathbf{x},\varepsilon) \neq \mathbf{0}$, so that linear and nonlinear least squares will be inconsistent. This chapter is concerned with estimation of slope parameters, $\boldsymbol{\beta}$, ancillary parameters, $\boldsymbol{\theta}$, and σ^2, the variance of ε, in cases in which linear and nonlinear least squares are not useable because of the correlation between $\mathbf{x}$ and ε. The essential estimation method is instrumental variables in several forms.

E21.2 Two Stage Least Squares

The standard case is the linear equation with endogenous right hand side variables,

$$\begin{aligned} y_i &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \\ \text{E}[\varepsilon_i|\mathbf{x}_i] &= g(\mathbf{x}_i) \neq \mathbf{0}, \\ \text{Var}[\varepsilon_i|\mathbf{x}_i] &= \sigma^2. \end{aligned}$$

The 2SLS, or IV estimator is based on a set of instruments, $\mathbf{z}_i$ which satisfy the two necessary conditions for an instrumental variable,

$$\begin{aligned} &\text{(orthogonality)} \quad \text{E}[\mathbf{z}_i\varepsilon_i] = \mathbf{0}, \\ &\text{(relevance)} \quad \text{E}[\mathbf{x}_i\mathrm{z}_i'] \neq \mathbf{0}. \end{aligned}$$

The 2SLS estimator is

$$\hat{\boldsymbol{\beta}} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

$$\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$$

The name of the estimator derives from the underlying result that the estimator can be computed by (1) regressing $\mathbf{X}$ on $\mathbf{Z}$ column by column and computing predicted values then (2) regressing $\mathbf{y}$ on the predicted values of $\mathbf{X}$ rather than the actual. The estimator is an instrumental variable estimator in that

$$\hat{\boldsymbol{\beta}} = \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

$$\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$$

because of the idempotency of $\mathbf{P}_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$; thus, $\hat{\mathbf{X}}$ is the set of instrumental variables. The method of instrumental variables is treated at length in standard texts such as Greene (2011) or Wooldridge (2010).

E21.2.1 Command

The essential command fitting linear models by instrumental variables is

2SLS ; Lhs = dependent variable
; Rhs = list of right hand side variables (all)
; Inst = list of all instrumental variables, including one $

The command for computing instrumental variables or two stage least squares estimates differs from that for ordinary least squares (**REGRESS**) in the list of instrumental variables. All options are the same as for the linear regression model – see Chapter E7 for details. This includes the specifications of **; AR1** disturbances, **; Plot** for residuals, etc. Chapters E7, E8, E10 and E11 give full details on these options. The list of instruments may include any variables existing in the data set.

HINT: If your equation (Rhs) includes a constant term, *one*, then you should also include *one* in the list of instrumental variables. Indeed, it might be the case that Inst should include *one* even if the Rhs does not. Note that the instrument list includes all exogenous variables that are in the Rhs list, plus the additional instrumental variables. The order condition for identification (and estimation here) requires that there be at least one instrumental variable in the Inst list for each endogenous variable in the Rhs list.

Computations use the standard results for two stage least squares. (See, e.g., Greene (2011).) There are no degrees of freedom corrections for variance estimators when this estimator is used. All results are asymptotic, and degrees of freedom corrections do not produce unbiased estimators in this context. Thus,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\boldsymbol{\beta}}'\mathbf{x}_i\right)^2 .$$

This is consistent with most published sources, but (curiously enough) inconsistent with most other commercially available computer programs. It will show up as a proportional difference in all estimated standard errors. If you would prefer that the degrees of freedom correction be made, add the specification

; Dfc

to your **2SLS** command. The estimator of the covariance matrix for the 2SLS estimator is

$$\text{Est.Var}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2\,[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}.$$

The command builder is essentially the same as that for the linear regression with the addition of the instrumental variables list. It can be opened by selecting Model:Linear Models/2SLS.

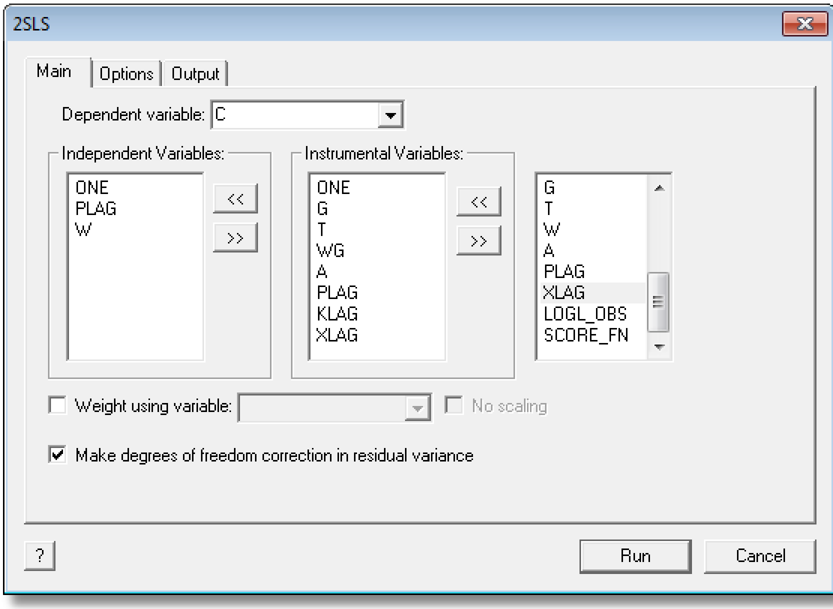


Figure E21.1 Command Builder for Two Stage Least Squares

E21.2.2 Model Output for the 2SLS Command

The output for the **2SLS** command is identical to that for **REGRESS**. The only indication that 2SLS, rather than OLS, was used in estimating the model will be a line at the top of the model results indicating that two stage least squares was used in the computations and a listing of the instrumental variables that will appear above the coefficient estimates. All retrievable results and methods for testing hypotheses are likewise identical.

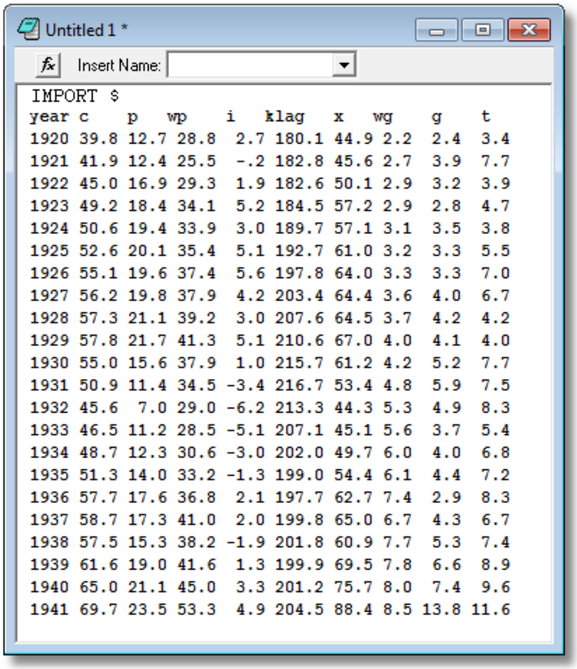
E21.2.3 Robust Estimation of the 2SLS Covariance Matrix

The White and Newey-West robust estimators of the covariance matrix of the least squares estimator described in Sections E7.10.1 and E7.10.2 can also be obtained for 2SLS by requesting them in the same fashion. All necessary corrections for the use of the instrumental variables are made in the computation. The calculation is otherwise the same as described in Section E7.10. The only difference here is that some of the columns of **x** are replaced by fitted values in the calculation.

E21.2.4 Application

The data listed below are Klein's data for estimation of his 'Model I.' These are used for testing simultaneous equations estimators and for demonstrating the techniques in most textbooks (in spite of the relative antiquity of the data). The model that is estimated by 2SLS is

$$\begin{array}{llllllll}
\text{(Consumption)} & c_t & = \alpha_0 & + \alpha_1 p_t & + \alpha_2 p_{t\text{-}1} & + \alpha_3(wp_t + wg_t) & + \varepsilon_{1t}, \\
\text{(Investment)} & i_t & = \beta_0 & + \beta_1 p_t & + \beta_2 p_{t\text{-}1} & + \beta_3 k_{t\text{-}1} & + \varepsilon_{2t}, \\
\text{(Private Wages)} & wp_t & = \gamma_0 & + \gamma_1 x_t & + \gamma_2 x_{t\text{-}1} & + \gamma_3(year\text{-}1931) & + \varepsilon_{3t}, \\
\text{(Equilibrium Demand)} & v_t & = c_t & + i_t & + g_t. & &
\end{array}$$



```
IMPORT $
year c    p    wp    i    klag  x    wg  g    t
1920 39.8 12.7 28.8  2.7 180.1 44.9 2.2  2.4  3.4
1921 41.9 12.4 25.5  -.2 182.8 45.6 2.7  3.9  7.7
1922 45.0 16.9 29.3  1.9 182.6 50.1 2.9  3.2  3.9
1923 49.2 18.4 34.1  5.2 184.5 57.2 2.9  2.8  4.7
1924 50.6 19.4 33.9  3.0 189.7 57.1 3.1  3.5  3.8
1925 52.6 20.1 35.4  5.1 192.7 61.0 3.2  3.3  5.5
1926 55.1 19.6 37.4  5.6 197.8 64.0 3.3  3.3  7.0
1927 56.2 19.8 37.9  4.2 203.4 64.4 3.6  4.0  6.7
1928 57.3 21.1 39.2  3.0 207.6 64.5 3.7  4.2  4.2
1929 57.8 21.7 41.3  5.1 210.6 67.0 4.0  4.1  4.0
1930 55.0 15.6 37.9  1.0 215.7 61.2 4.2  5.2  7.7
1931 50.9 11.4 34.5 -3.4 216.7 53.4 4.8  5.9  7.5
1932 45.6  7.0 29.0 -6.2 213.3 44.3 5.3  4.9  8.3
1933 46.5 11.2 28.5 -5.1 207.1 45.1 5.6  3.7  5.4
1934 48.7 12.3 30.6 -3.0 202.0 49.7 6.0  4.0  6.8
1935 51.3 14.0 33.2 -1.3 199.0 54.4 6.1  4.4  7.2
1936 57.7 17.6 36.8  2.1 197.7 62.7 7.4  2.9  8.3
1937 58.7 17.3 41.0  2.0 199.8 65.0 6.7  4.3  6.7
1938 57.5 15.3 38.2 -1.9 201.8 60.9 7.7  5.3  7.4
1939 61.6 19.0 41.6  1.3 199.9 69.5 7.8  6.6  8.9
1940 65.0 21.1 45.0  3.3 201.2 75.7 8.0  7.4  9.6
1941 69.7 23.5 53.3  4.9 204.5 88.4 8.5 13.8 11.6
```

Figure E21.2 Data for Klein Model I

The variables are:

c	=	consumption
p	=	private profits
wp	=	private wage bill
i	=	investment
klag	=	lagged value of capital stock
x	=	total demand
wg	=	government wage bill
g	=	government spending
t	=	indirect business taxes plus net exports
a	=	year - 1931

Klein's model is estimated using

```
READ        ; Nvar = 10  ; Nobs  = 22
            ; Names = year,c,p,wp,i,klag,x,wg,g,t $
CREATE      ; w = wp + wg ; a  = year - 1931 $
CREATE      ; plag = p[-1] ; xlag = x[-1] $
SAMPLE      ; 2-22 $
NAMELIST    ; cons = one,p,plag,w
            ; invs = one,p,plag,klag
            ; wage = one,x,xlag,a
            ; exog = one,g,t,wg,a,plag,klag,xlag $
2SLS        ; Lhs = c ; Rhs = cons ; Inst = exog $
2SLS        ; Lhs = i  ; Rhs = invs ; Inst = exog $
2SLS        ; Lhs = wp ; Rhs = wage ; Inst = exog $
```

(Some of the results that are repeated or are superfluous are omitted.)

```
-----------------------------------------------------------------------------
Two stage    least squares regression ............
LHS=C        Mean                    =     53.99524
             Standard deviation      =      6.86087
             Number of observs.      =           21
Model size   Parameters              =            4
             Degrees of freedom      =           17
Residuals    Sum of squares          =      17.7490
             Standard error of e     =      1.02179
Fit          R-squared               =       .97671
Not using OLS or no constant. Rsqrd & F may be < 0
Instrumental Variables:
ONE      G        T         WG       A         PLAG
KLAG     XLAG
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       C|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    16.5548***    1.32079    12.53  .0000     13.9661   19.1435
       P|      .01730       .11805      .15  .8835     -.21407    .24867
    PLAG|      .21623**     .10727     2.02  .0438      .00599    .42648
       W|      .81018***    .04025    20.13  .0000      .73129    .88907
--------+--------------------------------------------------------------------
Two stage    least squares regression ............
LHS=I        Mean                    =      1.26667
             Standard deviation      =      3.55195
Model size   Parameters              =            4
             Degrees of freedom      =           17
Residuals    Sum of squares          =      23.5141
             Standard error of e     =      1.17609
Fit          R-squared               =       .88488
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       I|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    20.2782***    7.54271     2.69  .0072      5.4948   35.0616
       P|      .15022       .17323      .87  .3858     -.18930    .48974
    PLAG|      .61594***    .16279     3.78  .0002      .29689    .93500
    KLAG|    -.15779***     .03613    -4.37  .0000     -.22859   -.08698
--------+--------------------------------------------------------------------
Two stage    least squares regression ............
LHS=WP       Mean                    =     36.36190
             Standard deviation      =      6.30440
Model size   Parameters              =            4
             Degrees of freedom      =           17
Residuals    Sum of squares          =      8.09926
             Standard error of e     =       .69024
Fit          R-squared               =       .98741
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      WP|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.50030       1.14778     1.31  .1912     -.74931   3.74990
       X|      .43886***    .03563    12.32  .0000      .36902    .50870
    XLAG|      .14667***    .03884     3.78  .0002      .07056    .22279
       A|      .13040***    .02914     4.47  .0000      .07328    .18751
--------+--------------------------------------------------------------------
```

E21.2.5 Specification Tests: Hausman and Wu

Two specification tests for exogeneity have been developed for the linear model. The Hausman test is based on a comparison of 2SLS to OLS. The model is specified as

$$y = \boldsymbol{\beta}_1'\mathbf{x}_1 + \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon$$

where $\mathbf{x}_2$ is K_2 variables. The question is whether the covariance of x_2 and ε is nonzero (that is, whether $\mathbf{x}_2$ is endogenous). Two competing estimators are **b**(ols) and **b**(2sls). The test is based on the difference, **d** = [**b**(2sls) - **b**(ols)]. Under the null hypothesis of exogeneity, plim **d** = **0**; under the alternative it is not. The Hausman (1978) test uses a Wald statistic to test the joint hypothesis that **d** equals zero. For the covariance matrix, the theorem in Hausman prescribes the difference in the two covariance matrices, **V**(2sls) - **V**(ols). A refinement needed to insure nonnegative definiteness of the matrix is to use instead of $\hat{\sigma}^2_{2sls}(\mathbf{X}'\mathbf{X})^{-1}$ instead of $s^2_{ols}(\mathbf{X}'\mathbf{X})^{-1}$ – this is a robust estimator of σ^2. The difference matrix then becomes $\hat{\sigma}^2\ [(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{P_Z}\mathbf{X})^{-1}]$. A second complication is that this matrix is singular, so a generalized inverse matrix that uses only K_2 of the dimensions is employed.

The Hausman test can be constructed as follows:

```
? These three lines are specific to the application

NAMELIST   ; x = the Rhs variables in the model $
NAMELIST   ; z = the full set of instruments, including exogenous elements of X $
CREATE     ; y = the dependent variable $

? These remaining lines carry out the test.  They are generic and need not be changed.

2SLS       ; Quietly ; Lhs = y ; Rhs = x ; Inst = z $
MATRIX     ; vh = varb - ssqrd*<x'x> ; dh = b - <x'x>*x'y $
MATRIX     ; list ; hausman = dh'*ginv(vh)*dh $
```

The number of degrees of freedom for the Hausman test is the number of variables in $\mathbf{x}_2$.

Note the following about the Hausman test.

1. The specification is a joint test against the exogeneity of the variables that appear in the *x* list that are not in the *z* list.

2. Applying this test to individual variables in the lists is not a valid test of any hypothesis. It is not possible for one element of the OLS estimator (that associated with a particular variable) to be inconsistent while the others are consistent. If any of the variables in *x* are endogenous, the entire OLS estimator is inconsistent, not just specific elements. By this construction, one could carry out the Hausman test for $K_2 > 1$ variables by just using one of the elements of *dh*. But, this would not test a different hypothesis; it would just waste the information contained in the other elements of *dh*.

3. This test is not useable for nonlinear models. It is specifically proposed for the linear model with possibly endogenous right hand side variables.

We carried out the Hausman test for the consumption function in Klein's Model I. For the application, *x* is *cons*, *z* is *exog* and *y* is *c*. The result of the test is

```
HAUSMAN |             1
--------+--------------
       1|       7.45221
```

The critical chi squared for two degrees of freedom and 95% significance is 5.99. We conclude on the basis of the test that at least one of *p* and *w* are endogenous in the consumption function.

The second test is the Wu test (also attributed to Durbin and Hausman and reiterated in Davidson and MacKinnon (1993)). The Wu test is a simple variable addition test based on least squares. The test can be carried out by the following steps:

Step 1. For each possibly endogenous variable in the equation, compute the residuals from a regression of that variable on the full set of exogenous variables.

Step 2. Add the residuals to the least squares regression. The test is carried out by testing the joint hypothesis that the coefficients on the added residuals are zero.

Mechanically, there is a much simpler way to carry out this test. The relevant F statistic is computed as

$$\text{F}[K_2, n-K_1-K_2-K_2] = \frac{(ss_{ols} - ss_{augmented})/K_2}{ss_{augmented}/(n-K_1-K_2-K_2)},$$

where ss_{ols} is the sum of squares in the original least squares regression and $ss_{augmented}$ is the sum of squares in the least squares regression to which the K_z elements of **z** that are not contained in $\mathbf{x}_1$ are added to the equation. Note that the denominator degrees of freedom includes the additional K_2 not K_z – the test is whether the K_2 coefficients on the added residuals are zero. Since this is based on OLS regressions, the F statistic is exactly $1/K_2$ times the chi squared statistic that would result if a Wald statistic were used instead.

This test is automated. The command is

REGRESS ; Lhs = y variable ; Rhs = x variables
; Inst = z variables ; Wu test $

Include the full set of instruments in **z** – that includes $\mathbf{x}_1$ plus the additional instrumental variables that are not contained in $\mathbf{x}_1$. These K_Z variables are denoted $\mathbf{z}_2$. *LIMDEP* will sort out internally what variables are contained in $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{z}_2$ (the part of **z** that is not in $\mathbf{x}_1$).

The following shows in detail how this would be applied to the consumption function.

REGRESS ; Lhs = c
; Rhs = one, p, plag, w
; Inst = one,g,t,wg,a,plag,klag,xlag
; Wu test $

In this setup, $\mathbf{x}_1$ = (*one,plag*), $\mathbf{x}_2$ = (*p,w*), $\mathbf{z}_2$ = (*g,t,wg,a,klag,xlag*). Applying this test to the consumption function set up earlier, we would use

REGRESS ; Lhs = c ; Rhs = cons ; Inst = Exog ; Wu test $

The results are as follows:

```
-----------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=C         Mean                    =      53.99524
              Standard deviation      =       6.86087
              No. of observations     =            21  Degrees of freedom
Regression    Sum of Squares          =       923.550           3
Residual      Sum of Squares          =       17.8794          17
Total         Sum of Squares          =       941.430          20
              Standard error of e     =       1.02554
Fit           R-squared               =        .98101  R-bar squared =   .97766
--------+--------------------------------------------------------------------
        |                  Standard             Prob.      95% Confidence
       C|  Coefficient       Error        t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|    16.2366***     1.30270    12.46  .0000     13.6834   18.7898
       P|     .19293**       .09121     2.12  .0495      .01417    .37170
    PLAG|     .08988         .09065      .99  .3353     -.08778    .26755
       W|     .79622***      .03994    19.93  .0000      .71793    .87451
--------+--------------------------------------------------------------------
+-----------------------------------------------------------------------------+
| Wu test for exogeneity of variables in RHS that are not listed in   |
| INST, beginning with P        . F[ 2,     15] =      5.603. P value  |
| for this F statistic is  .0000. (If < .05, reject exogeneity.)      |
+-----------------------------------------------------------------------------+
```

Note that Wu and Hausman are not the same test. Wu is an F test while Hausman is a Wald test, and they are based on different sums of squared residuals. Thus, twice the Wu statistic in the preceding does not produce the Hausman statistic.

E21.3 Autocorrelation with a Lagged Dependent Variable

If you are using **2SLS** to estimate an equation with a lagged endogenous variable and autocorrelation, such as:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma y_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t.$$

you can use Hatanaka's (1974) efficient estimator, which is asymptotically equivalent to maximum likelihood for normally distributed disturbances, u_t. The procedure is as follows:

Step 1. Use instrumental variables to estimate $[\boldsymbol{\beta},\gamma]$. Any consistent estimator will do. A suitable instrumental variable for the lagged value of y_t might be the lagged value of the prediction of y_t from a regression on $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$.

Step 2. Using the consistent estimator in Step 1, estimate ρ consistently by the autocorrelation of the residuals,

$$e_t = y_t - \mathbf{b}_{IV}'\mathbf{x}_t - c_{IV}y_{t-1}.$$

That is, compute the residuals using actual values, not predictions.

Step 3. Now, use the Cochrane-Orcutt transformation to do GLS based on the original data, but add an additional regressor to the model, e_{t-1}. (The transformation is not applied to the lagged residual.)

Step 4. The efficient estimator of ρ is the original estimator plus the slope on the lagged residual in the regression at Step 3. The asymptotic covariance for this estimate is that provided for the slope in Step 3. I.e., the GLS regression in Step 3 provides the full set of covariances.

This procedure uses the **2SLS** command, not **REGRESS**. The command is

```
2SLS      ; Lhs = y ; Rhs = x ; Inst = full set of instruments
          ; AR1 ; Hatanaka $
```

To use the **2SLS** command builder, click the Autocorrelation button on the Options page to open a dialog box that offers Hatanaka's estimator as an option. (See Figure E21.3.)

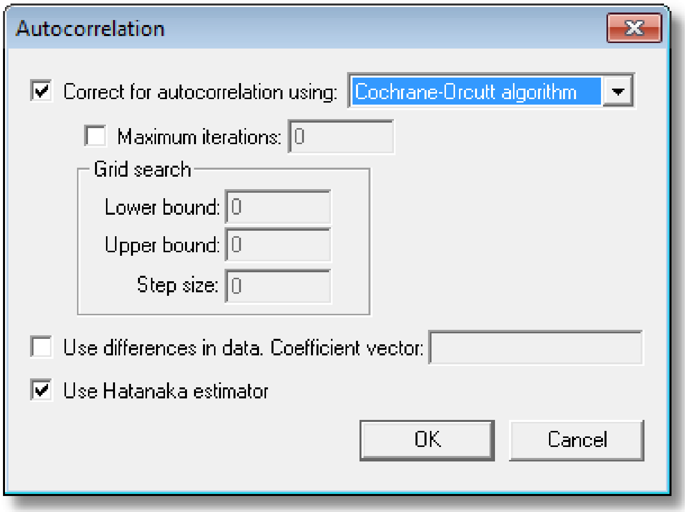


Figure E21.3 Command Builder for Hatanaka Estimator

Note that the set of instruments includes:

- all exogenous variables in *x* on the Rhs,
- *one* if it is included in the Rhs,
- additional instrumental variables.

For estimating a simultaneous equations model with first order autoregressive disturbances, the list of instruments should include the lagged values of all endogenous and exogenous variables in the reduced form. (See Pindyck and Rubinfeld (1991) and Greene (2011).)

For example, for a linear model that contains a constant, two regressors and a lagged dependent variable,

```
CREATE    ; ylag = y[-1]
          ; x1lag = x1[-1]
          ; x2lag = x2[-1] $
SAMPLE    ; 3 - ... end of sample $
2SLS      ; Lhs = y
          ; Rhs = one,x1,x2,ylag
          ; Inst = one,x1,x2,x1lag,x2lag
          ; AR1 ; Hatanaka $
```

Note that we have also begun the sample period for this estimator at observation 3. Since the lagged value of y_{t-1} is needed for the Cochrane-Orcutt transformation, two observations at the beginning of the sample will be incomplete.

E21.4 Alternatives to 2SLS

Two estimators (other than OLS) are proposed as alternatives to 2SLS. The limited information maximum likelihood (LIML) estimator based on assuming the disturbances are normally distributed may have better small sample properties than 2SLS and, unlike 2SLS, is invariant to normalization. The same result is obtained regardless of which variable in a structural equation is labeled the 'dependent' variable. Ackerberg and Devereux's (2009) JIVE estimator is a jackknife estimator that is intended to remedy some of the small sample bias of 2SLS. The 'improved,' IJIVE estimator, as suggested, is an extension.

E21.4.1 LIML

The LIML estimator is derived at length in numerous sources such as Davidson and MacKinnon (2004) and Greene (2011). We note only the mechanics of the computation here. In the single equation in a system, which we write in terms of the full n observations as

$$\mathbf{Y} = \mathbf{Y}\boldsymbol{\gamma} + \mathbf{X}^0\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where the set of endogenous variables is $\mathbf{Y}$ and the 'included' exogenous variables are $\mathbf{X}^0$. (In the consumption function in our earlier example, $\mathbf{y}$ is $\mathbf{c}$, $\mathbf{Y}$ is $(\mathbf{p},\mathbf{w})$ and $\mathbf{X}^0$ is $(\mathbf{1},\mathbf{plag})$.). Consider the residuals in the linear regressions of $\mathbf{Y}^0 = (\mathbf{y},\mathbf{Y})$ on $\mathbf{X}^0$,

$$\mathbf{E}^0 = \mathbf{M}^0\mathbf{Y}^0 = (\mathbf{I} - \mathbf{X}^0(\mathbf{X}^{0\prime}\mathbf{X}^0)^{-1}\mathbf{X}^{0\prime})\mathbf{Y}^0.$$

(The residuals are computed column by column then arranged next to each other in $\mathbf{E}^0$.) Then, the covariance matrix estimator (mean squares and cross products) is

$$\mathbf{W}^0 = (1/n)\mathbf{E}^{0\prime}\mathbf{E}^0.$$

Now, repeat the computation using not $\mathbf{X}^0$, but $\mathbf{X}^1$, which is all of the exogenous variables in the system. (In the consumption function example, $\mathbf{X}^1$ would be (**one,g,t,wg,a,plag,klag,xlag**). Note that $\mathbf{X}^1$ contains $\mathbf{X}^0$ plus at least M additional variables, where M is the number of variables in $\mathbf{Y}$. (This number would be $M = 2$ in the consumption function example.) Then, based on this regression,

$$\mathbf{W}^1 = (1/n)\mathbf{E}^{1\prime}\mathbf{E}^1.$$

Define λ = the smallest characteristic root of $(\mathbf{W}^1)^{-1}\mathbf{W}^0$. (The root is real even though the product matrix is asymmetric and greater than one if $\mathbf{X}^1$ contains $\mathbf{X}^0$ plus more than M additional variables.) Now, $\mathbf{W}^0$ and $\mathbf{W}^1$ are partitioned into $w_{yy}{}^0$, $\mathbf{w}_{\mathbf{Y}y}{}^0$ and $\mathbf{W}_{\mathbf{YY}}{}^0$ based on $\mathbf{y}$ and $\mathbf{Y}$, and $\mathbf{W}^1$ likewise. The components of the LIML estimator are computed as

$$\hat{\gamma}_{LIML} = [\mathbf{W}^0_{\mathbf{YY}} - \lambda\mathbf{W}^1_{\mathbf{YY}}]^{-1}\left(\mathbf{w}^0_{\mathbf{Y}y} - \lambda\mathbf{w}^1_{\mathbf{Y}y}\right)$$

$$\hat{\beta}_{LIML} = \left(\mathbf{X}^{0\prime}\mathbf{X}^0\right)\mathbf{X}^{0\prime}\left(\mathbf{y} - \mathbf{Y}^0\hat{\gamma}_{LIML}\right)$$

The asymptotic covariance matrix is estimated using the same matrix as 2SLS save for the computation of the residuals based on the LIML estimator, and replacing 1.0 with λ in the upper left block of the matrix. (See equation (10-56) in Greene (2011, p. 329).)

By construction, $\lambda > 1$ if the number of instruments is greater than the number needed. (In the consumption function example, there are two endogenous variables and eight instrumental variables.) Thus, the system is overidentified by four variables. A test of the overidentifying restrictions is based on

$$c = n(\lambda - 1).$$

The limiting distribution of c is chi squared with degrees of freedom equal to the number of overidentifying restrictions – four in our example.

The LIML estimator is requested with

LIML ; Lhs = y, variables in $\mathbf{Y}^0$
; Rhs = variables in $\mathbf{X}^0$
; Inst = variables in X, including $\mathbf{X}^0$ and at least M more $

The command is otherwise the same as 2SLS in the options, such as residuals, fitted values and hypothesis tests. There is no 'robust' covariance matrix available, since the estimator is based on a specific assumption.

For the consumption function in our example, the estimator would be

LIML ; Lhs = c,p,w
; Rhs = one,plag
; Inst = one,g,t,wg,a,plag,klag,xlag $

The results with a comparison to 2SLS are as follows:

```
-----------------------------------------------------------------------------
LmtdInfoMLE  for linear sim eqn model ............
LHS=C        Mean                    =     53.99524
             Standard deviation      =      6.86087
             Number of observs.      =           21
Model size   Parameters              =            4
             Degrees of freedom      =           17
Residuals    Sum of squares          =      33.0967
             Standard error of e     =      1.39530
Spec.Test    Smallest root <W1>W0 =      1.49875
             Chi sqd. test overID =     10.47366
             No. of over ID insts.=            4
             P value for chi-sqd. =       .03316
Instrumental Variables:
ONE       G         T         WG        A         PLAG
KLAG      XLAG
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       C|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
       P|    -.22251        .20175    -1.10  .2701     -.61793    .17291
       W|     .82256***     .05538    14.85  .0000      .71402    .93110
Constant|    17.1477***    1.84030     9.32  .0000     13.5407   20.7546
    PLAG|     .39603**      .17360     2.28  .0225      .05578    .73627
--------+--------------------------------------------------------------------
```

```
--------+-------------------------------------------------------------
Two stage    least squares regression ...........
LHS=C        Mean                    =      53.99524
             Standard deviation      =       6.86087
             Number of observs.      =            21
Model size   Parameters              =             4
             Degrees of freedom      =            17
Residuals    Sum of squares          =       17.7490
             Standard error of e     =       1.02179
Fit          R-squared               =        .97671
             Adjusted R-squared      =        .97260
Model test   F[  3,    17] (prob) =   237.6(.0000)
Diagnostic   Log likelihood          =     -28.03169
             Restricted(b=0)         =     -69.72792
             Chi-sq [  3]  (prob) =  83.4(  .0000)
Info criter. Akaike Info. Criter. =        .21276
Not using OLS or no constant. Rsqrd & F may be < 0
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       C|  Coefficient       Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------
Constant|    16.5548***     1.32079    12.53  .0000     13.9661   19.1435
       P|     .01730         .11805      .15  .8835     -.21407    .24867
    PLAG|     .21623**       .10727     2.02  .0438      .00599    .42648
       W|     .81018***      .04025    20.13  .0000      .73129    .88907
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E21.4.2 JIVE Estimator

The jackknife instrumental variable estimator (JIVE), developed in Ackerberg and Devereux (2009) (based on Phillips and Hale (1977), Staiger and Stock (1997), Angrist, Imbens and Krueger (1999) and others, is a straightforward modification of 2SLS. Write the 2SLS estimator as a true IV estimator,

$$\mathbf{b}_{2SLS} = \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

where $\mathbf{X} = (\mathbf{X}_1,\mathbf{X}_2)$ and $\mathbf{X}_2$ is the set of K_2 endogenous variables and $\hat{\mathbf{X}}$ is the set of predictions in the regressions of the columns of $\mathbf{X}$ on all of the instrumental variables, $\mathbf{Z}$, which includes $\mathbf{X}_1$ and some additional variables. In these regressions, the predictions are computed using the least squares coefficients in the linear regressions of each column of $\mathbf{X}$ on $\mathbf{Z}$. The matrix whose columns are the regression coefficients used for the predictions is $\mathbf{P} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_2$. (The predictions of $\mathbf{X}_1$ are themselves, $\mathbf{X}_1$.) The JIVE estimator is constructed by using an observations specific set of least squares coefficients to compute each set of predictions. That is, for row i of $\hat{\mathbf{X}}$, to compute the predictions, instead of using the same matrix of coefficients, $\mathbf{P}$, we compute $\mathbf{P}_{(i)}$ by omitting observation *i* from the first step regression – hence the jackknife aspect of the computation. (Ackerberg and Devereux provide a 'one pass' computation that can obviate computing *n* regressions. However, with the sample sizes typically in use for this sort of computation, one would expect the time savings to be trivial.) They also propose an 'improved' estimator, IJIVE, that is obtained by first partialling out $\mathbf{X}_2$ from $\mathbf{X}_1$, $\mathbf{y}$, and the other instruments. (Davidson and MacKinnon (2006 and others) are convinced that this estimator has no moments and stridently argue against its use.)

The estimator is obtained as a modification of 2SLS;

2SLS ; … ; JIVE or IJIVE $

The other features and options of 2SLS remain as earlier. This computation only changes the computation of the coefficient vector and the covariance matrix.

The earlier example is extended here with

LIML ; Lhs = c, p,w ; Rhs = one,plag
; Inst = one,g,t,wg,a,plag,klag,xlag
; JIVE $

```
-----------------------------------------------------------------------------
Two stage     least squares regression ............
LHS=C         Mean                   =        53.99524
              Standard deviation     =         6.86087
              Number of observs.     =              21
Model size    Parameters             =               4
              Degrees of freedom     =              17
Residuals     Sum of squares         =         17.7490
              Standard error of e    =         1.02179
Fit           R-squared              =          .97671
              Adjusted R-squared     =          .97260
Model test    F[  3,    17] (prob) =   237.6(.0000)
Diagnostic    Log likelihood         =       -28.03169
              Restricted(b=0)        =       -69.72792
              Chi-sq [  3]  (prob) =  83.4(  .0000)
Info criter. Akaike Info. Criter. =          .21276
Not using OLS or no constant. Rsqrd & F may be < 0
Instrumental Variables:
ONE       G         T         WG        A         PLAG
KLAG      XLAG
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       C|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    16.5548***     1.32079    12.53  .0000     13.9661   19.1435
       P|     .01730         .11805      .15  .8835     -.21407    .24867
    PLAG|     .21623**       .10727     2.02  .0438      .00599    .42648
       W|     .81018***      .04025    20.13  .0000      .73129    .88907
--------+--------------------------------------------------------------------
Two stage     least squares regression ............
Residuals     Sum of squares         =         49.6921
              Standard error of e    =         1.70970
Fit           R-squared              =          .93480
              Adjusted R-squared     =          .92329
Model test    F[  3,    17] (prob) =    81.2(.0000)
Diagnostic    Log likelihood         =       -38.84161
              Restricted(b=0)        =       -69.72792
              Chi-sq [  3]  (prob) =  61.8(  .0000)
Info criter. Akaike Info. Criter. =         1.24228
Instrumental Variables using JIVE (jackknife):
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       C|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    17.6126***     3.86726     4.55  .0000     10.0329   25.1923
       P|    -.37527         .87666     -.43  .6686    -2.09348   1.34295
    PLAG|     .51456         .67472      .76  .4457     -.80788   1.83699
       W|     .82676***      .05334    15.50  .0000      .72221    .93130
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E21.5 Nonlinear IV Estimation

Estimation of the parameters of the nonlinear model $y_i = f(\mathbf{x}_i,\boldsymbol{\beta}) + \varepsilon_i$ can be done by nonlinear least squares. (See Amemiya (1987).) If either $\mathbf{x}_i$ is correlated with ε_i or the model is identified by the orthogonality conditions $E[\mathbf{z}_i\varepsilon_i] = \mathbf{0}$, where $\mathbf{z}$ is a set of instrumental variables, then an appropriate estimator is nonlinear instrumental variables.

The nonlinear IV estimator is requested with

NLSQ **; Lhs = the dependent variable**
; Fcn = function specification
; Labels = labels for parameters
; Start = starting values
; Inst = list of instrumental variables $

The command and other options are exactly as described in Chapter E14. The only new feature added here is the set of instrumental variables. This is also a form of GMM estimator. The GMM estimator is described in Section E21.6.

The nonlinear IV procedure involves a set of instrumental variables, $\mathbf{z}_1, \mathbf{z}_2,..., \mathbf{z}_K$. Suppose these are combined in an $n\times K$ matrix $\mathbf{Z}$. Let the vector $\boldsymbol{\varepsilon}$ denote the $n\times 1$ column of residuals $\varepsilon_i = y_i - f(\mathbf{x}_i,\boldsymbol{\beta})$. Then, the nonlinear least squares estimator described in Chapter E14 is found by solving the optimization problem

$$\text{Minimize wrt } \boldsymbol{\beta} \;\; \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\varepsilon}(\boldsymbol{\beta}).$$

For the nonlinear instrumental variables estimation problem (NLIV), the estimation criterion is

$$\text{Minimize wrt } \boldsymbol{\beta} \;\; \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta}).$$

This replicates two stage least squares for linear functions. You may also combine nonlinear two stage least squares with weighted least squares. In this case, we define a diagonal matrix $\mathbf{W}$ whose diagonal elements are the weights, w_i. The weighted nonlinear IV procedure (NLWIV) is

$$\text{Minimize wrt } \boldsymbol{\beta} \;\; \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{W}\mathbf{Z}(\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}\boldsymbol{\varepsilon}(\boldsymbol{\beta}).$$

To request the nonlinear instrumental variables estimation method, you will use

; Inst = list of variables in Z

in the **NLSQ** command. Add

; Wts = weighting variable

for the nonlinear weighted instrumental variables estimator.

The number of instruments you provide must be at least as large as the number of parameters that you estimate. For purposes of this calculation, *LIMDEP* ignores the possibility of fixed parameters. Thus, if your model has six parameters, and you fix two of them, you must still provide at least six instrumental variables. Although this may seem like a restriction, it is trivial to work around it, simply by fixing the values in the function definition, and not defining them as parameters in the model formulation.

The asymptotic covariance matrix estimated for the nonlinear instrumental variables estimator is

$$\text{Est.Asy.Var.}[\mathbf{b}] = \hat{\sigma}^2 [\mathbf{G}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}]^{-1}$$

where the rows of **G** are the derivatives of $f(\mathbf{x}_i,\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and

$$\hat{\sigma}^2 = (1/n)\Sigma_i[y_i - f(\boldsymbol{\beta},\mathbf{x}_i)]^2.$$

($1/n$ is replaced with $1/(n$ - #parameters) if you select **; Dfc**.)

To apply the method, we revisit the health care data used in several previous examples. The model is

$$\text{Income} = \exp(\beta_0 + \beta_1 educ + \beta_2 age + \beta_3 healthsatisfaction) + \varepsilon.$$

The model is estimated first by nonlinear least squares. It is believed that health satisfaction (*hsat*) is endogenous. Instruments to be used are marital status, public insurance, addon insurance and children in the household. (One might question the endogeneity of the insurance purchases.) The initial ordinary (nonlinear) least squares estimates are obtained first with

```
NLSQ        ; Lhs = hhninc
            ; Fcn = Exp(b0+b1*educ+b2*age+b3*hsat)
            ; Labels = b0,b1,b2,b3
            ; Start = -1,0,0,0  $
```

The nonlinear IV estimator is invoked with

```
NLSQ        ; Lhs = hhninc
            ; Fcn = Exp(b0+b1*educ+b2*age+b3*hsat)
            ; Labels = b0,b1,b2,b3
            ; Start = -1,0,0,0
            ; Inst = one,educ,age,married,public,addon,hhkids $
```

We are interested in the partial effects of education on income.

```
PARTIALS    ; Function = Exp(b0+b1*educ+b2*age+b3*hsat)
            ; Labels = b0,b1,b2,b3
            ; Effects: educ & age = 25(5)60 ; Plot $
```

Results are as follows:

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=HHNINC    Mean                   =         .44476
              Standard deviation     =         .21659
              Number of observs.     =           3377
Model size    Parameters             =              4
              Degrees of freedom     =           3373
Residuals     Sum of squares         =        144.618
              Standard error of e    =         .20694
Fit           R-squared              =         .08682
              Adjusted R-squared     =         .08709
Model test    F[  3,  3373] (prob) =   106.9(.0000)
Diagnostic    Log likelihood         =      528.11382
              Restricted(b=0)        =      374.76923
              Chi-sq [  3]  (prob) = 306.7(  .0000)
Info criter. Akaike Info. Criter. =       -3.14828
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|   -1.71782***      .05743   -29.91  .0000    -1.83039  -1.60525
      B1|     .05006***      .00284    17.63  .0000      .04450    .05563
      B2|     .00402***      .00071     5.65  .0000      .00263    .00542
      B3|     .02268***      .00385     5.90  .0000      .01514    .03021
--------+--------------------------------------------------------------------
Instrumental Variables (NLIV).....................
Nonlinear     least squares regression ............
LHS=
              Standard deviation     =         .21659
              Number of observs.     =           3377
Model size    Parameters             =              4
              Degrees of freedom     =           3373
Residuals     Sum of squares         =        435.938
              Standard error of e    =         .35929
Fit           R-squared              =       -1.75271
              Adjusted R-squared     =       -1.75190
Diagnostic    Log likelihood         =    -1334.98421
              Restricted(b=0)        =      374.76923
Info criter. Akaike Info. Criter. =       -2.04487
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|   -4.70470***      .94574    -4.97  .0000    -6.55833  -2.85108
      B1|     .02900***      .00654     4.43  .0000      .01617    .04182
      B2|     .01589***      .00316     5.02  .0000      .00969    .02209
      B3|     .38942***      .10804     3.60  .0003      .17767    .60117
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------
Partial Effects  Analysis for User Specified Function
--------------------------------------------------------------------
Effects on function with respect to EDUC
Results are computed by average over sample observations
Partial effects for continuous EDUC     computed by differentiation
Effect is computed as derivative     = df(.)/dx
--------------------------------------------------------------------
df/dEDUC            Partial    Standard
(Delta method)      Effect      Error     |t|  95% Confidence Interval
--------------------------------------------------------------------
APE. Function        .01291     .00291    4.44      .00722      .01861
AGE     = 25.00      .00990     .00254    3.90      .00492      .01487
AGE     = 30.00      .01072     .00263    4.07      .00555      .01588
AGE     = 35.00      .01160     .00274    4.24      .00624      .01696
AGE     = 40.00      .01256     .00284    4.42      .00699      .01814
AGE     = 45.00      .01360     .00296    4.59      .00779      .01941
AGE     = 50.00      .01472     .00309    4.76      .00866      .02079
AGE     = 55.00      .01594     .00324    4.92      .00959      .02229
AGE     = 60.00      .01726     .00341    5.07      .01058      .02394
```

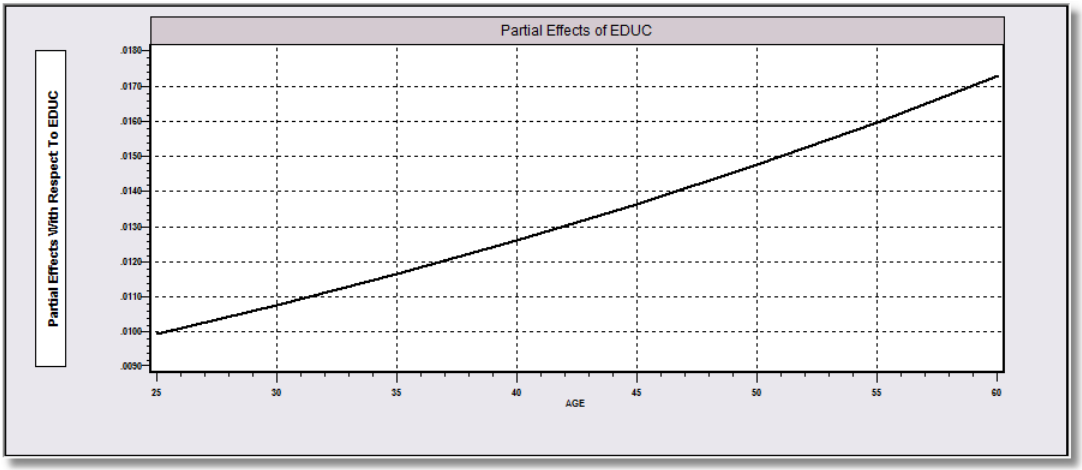


Figure E21.4 Partial Effects in Nonlinear Regression

E21.6 NLSQ/GMM Estimation

LIMDEP can be used for formal GMM estimation of econometric models. Although the methodology is common to all of them, we provide several approaches. The nonlinear least squares estimator presented in the preceding section is based on the least squares criterion

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$$

which minimizes the simple sum of squares of a set of residuals. As noted earlier, different weighting schemes and the use of instrumental variables extends this to more general GMM interpretations. A somewhat more general estimator results from using instrumental variables, with

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta}).$$

The yet more general estimation criterion,

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$$

allows for instrumental variables and a weighting matrix, $\boldsymbol{\Omega}$. Depending on the choice of the weighting matrix, this will produce GMM estimators of various sorts. Finally, consider the less structured GMM criterion:

$$q = \bar{\mathbf{m}}'\Sigma\, \bar{\mathbf{m}}$$

where $$\bar{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)$$

based on a set of L ‘orthogonality conditions,’

$$E[\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)] = \mathbf{0}.$$

E21.6.1 GMM Estimation of Single Equation Nonlinear Models

LIMDEP’s **NLSQ** command can be used for obtaining GMM estimates of the parameters in an equation. (Reference is made to Hansen (1982) or Pagan and Vella (1989) for details on the method.) The following will briefly present the relevant background, then give the command structure.

Let $\boldsymbol{\beta}$ be the vector of parameters being estimated. The estimation criterion is

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\beta})$ is the column of residuals in the form

$$\varepsilon_i(\boldsymbol{\beta}) = y_i - f(\mathbf{x}_i,\boldsymbol{\beta}),$$

and $\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}$ is the expected value of

$$\mathbf{W} = (1/n)\Sigma_i\mathbf{z}_i\varepsilon_i^2\mathbf{z}_i'.$$

This matrix must be estimated using the starting values provided for the estimator. Each column of the matrix $\mathbf{Z}$ contains the observations derived from the orthogonality conditions

$$E[\varepsilon_i(\boldsymbol{\beta})z_{ik}] = 0$$

or, in a more compact vector notation,

$$E[\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{z}_k] = 0 \text{ for the column, and}$$

$$E[\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}] = \mathbf{0} \text{ for the entire set of variables.}$$

The matrix $\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}$ is the ‘optimal weighting matrix,’ such that

$$\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z} = \mathrm{E}[(1/n)\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}].$$

After estimation, the estimated asymptotic covariance matrix for this estimator is given by

$$\text{Est.Asy.Var.}[\mathbf{b}] = [\mathbf{G}'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}]^{-1},$$

where each row of $\mathbf{G}$ is $\partial\varepsilon_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$.

Note in $(1/n)\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}$ that if the disturbances in the model are uncorrelated and homoscedastic, the appropriate matrix to use would simply be $\mathbf{Z}'\mathbf{Z}$. The scalar, σ^2, would be irrelevant. In this instance, GMM estimation reduces simply to nonlinear instrumental variables. Thus, the criterion reduces precisely to that given earlier. But, if the disturbances are heteroscedastic and/or autocorrelated, then the nonscalar matrix $\boldsymbol{\Omega}$ presents a new difficulty. Consider, first, the case of heteroscedasticity. In this case, estimation of $\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}$ is exactly analogous to computation of the White estimator for heteroscedastic disturbances in the classical regression case. (See Chapter E10.) Given a set of consistent estimates for the elements of $\mathbf{W}$, this is at least straightforward. But, if the disturbances are autocorrelated as well, then there are nonzero off diagonal elements in $\boldsymbol{\Omega}$. If a finite lag length can be specified (i.e., a truncated or finite moving average representation), then the Newey-West estimator (see Chapter E11) can be used, instead. *LIMDEP* uses this approach. The consistent estimator needed to compute the elements of $\mathbf{W}$ must be provided as the starting values for the estimator. One approach would be to estimate the model ignoring the heteroscedasticity and/or autocorrelation just to get the consistent (albeit, inefficient) estimates to use as starting values.

The foregoing is applied with the **NLSQ** command. The basic format would be:

```
NLSQ        ; Labels = list
            ; Start = set of values
            ; Fcn = expression for the residual y(i) - f(.)
            ; Inst = list of instruments, z(k) $
and         ; Pds = 0
```

to request the White estimator

or **; Pds = L** (e.g., **Pds = 5**)

to use the Newey-West estimator. Note that to request the heteroscedasticity estimator, **; Pds = 0** must be provided. This is not the default. The **; Pds** specification requests the computation of a nonscalar **W**; the number of periods (zero or positive) dictates how the computation is to be done. *Also, note that for GMM estimation, you are not providing the name of a Lhs variable.*

In the **NLSQ** command, for GMM estimation, you specify the residual, not just the function, and you do not name the Lhs variable. This is an important difference in the command. For example, to fit the function

$$f(x_i,\boldsymbol{\beta}) = \text{Exp}(\beta_1 + \beta_2 x)$$

use the following commands:

For nonlinear least squares:

; Lhs = y
; Fcn = Exp(b1 + b2*x)

For nonlinear instrumental variables:

; Lhs = y
; Fcn = Exp(b1 + b2 *x)
; Inst = the list of IVs

For GMM:

; Fcn = y - Exp(b1 + b2*x)
; Inst = the list of IVs
; Pds = 0 or the number

E21.6.2 Technical Note on Optimization

The **NLSQ** command maintains all the accounting information to ensure that the nonlinear optimization problem is analyzed as a regression. You can also compute nonlinear least squares coefficients using the **MINIMIZE** command. This will produce the same estimates, but it will not produce the same estimated asymptotic covariance matrix for the coefficients. **MINMIZE** simply accumulates the estimated Hessian of the criterion function, but does not scale it to account for the disturbance variance. Consider the example of a linear regression, using the Grunfeld data used earlier (Greene (2011, Table F10.4)):

```
REGRESS     ; Lhs = i ; Rhs = one,f,c $
NLSQ        ; Lhs = i ; Labels = b1,b2,b3 ; Start = 0,0,0 ;
            ; Fcn = b1+b2*f+b3*c
            ; Dfc $
MINIMIZE    ; Fcn = (i - b1 - b2*f - b3*c)^2
            ; Labels = b1,b2,b3 ; Start = 0,0,0 $
```

All three produce the same coefficients, and the first two produce the same asymptotic covariance matrix. In the first two cases, that covariance matrix is the conventional Est.Var[$\mathbf{b}_{LS}$] = $s^2(\mathbf{X}'\mathbf{X})^{-1}$. But, for the third estimation problem, the estimated asymptotic covariance matrix is computed as

$$\text{Est.Asy.Var}[\mathbf{b}_{MINIMIZE}] = \left[\sum_{i=1}^{n}(-2e_i\mathbf{x}_i)(-2e_i\mathbf{x}_i)'\right]^{-1}$$

This matrix is the BHHH estimator for the function specified. There is no definite relationship between the two matrices; it depends on the data. For the way it is specified above, in large samples, the covariance matrix for the **MINIMIZE** command would resemble $(1/4s^4)$ times that from **NLSQ**, but that should not be used as any kind of rule of thumb.

The essential point in this result is that the **MINIMIZE** command does not directly specify that the problem is a least squares problem or a regression model. From the point of view of the program, there is nothing to distinguish this from any other optimization problem. The **NLSQ** command requests not only a particular kind of optimization, but also a particular arrangement of and interpretation of the results.

```
--------------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=I        Mean                    =      145.95825
             Standard deviation      =      216.87530
             No. of observations     =            200  Degrees of freedom
Regression   Sum of Squares          =    .760409E+07            2
Residual     Sum of Squares          =    .175585E+07          197
Total        Sum of Squares          =    .935994E+07          199
             Standard error of e     =       94.40840
Fit          R-squared               =         .81241  R-bar squared =   .81050
Model test   F[  2,   197]           =      426.57573  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =    -1191.80236  Akaike I.C.   =  9.11015
             Restricted (b=0)        =    -1359.15096  Bayes  I.C.   =  9.15962
             Chi squared [  2]       =      334.69719  Prob C2 > C2* =   .00000
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       t     |t|>T*        Interval
--------+-----------------------------------------------------------------------
Constant|   -42.7144***     9.51168    -4.49  .0000    -61.3569  -24.0718
       F|      .11556***     .00584    19.80  .0000      .10412    .12700
       C|      .23068***     .02548     9.05  .0000      .18075    .28061
--------+-----------------------------------------------------------------------


--------------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
(Results identical to OLS are omitted)
--------+-----------------------------------------------------------------------
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    1755850.
--------------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function   1755850.48409
Restricted log likelihood        .00000
Chi squared [   3 d.f.]   3511700.96818
Significance level               .00000
Estimation based on N =    200, K =   0
Inf.Cr.AIC  =********* AIC/N = ********
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*        Interval
--------+-----------------------------------------------------------------------
      B1|   -42.7144***      .00076 ********  .0000    -42.7159  -42.7129
      B2|      .11556***   .3572D-06 ********  .0000      .11556    .11556
      B3|      .23068***   .8903D-06 ********  .0000      .23068    .23068
--------+-----------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E21.7 General Specifications of the GMM Estimator

A somewhat more general form of the GMM estimation procedure departs from a set of 'orthogonality conditions,'

$$E[m_{il}(\boldsymbol{\beta},\mathbf{x}_i)] = 0,\ l = 1,\ldots,L$$

where $\boldsymbol{\beta}$ is the vector of parameters to be estimated, $\mathbf{x}_i$ is a set of variables that is assumed to be in the set of information that defines the 'moment condition,' and $m_{il}(.)$ is one of L expectations that the model specifies to equal zero. The GMM estimator is obtained by finding the estimator, $\mathbf{b}$, that makes the empirical moment,

$$\overline{m}_l = \frac{1}{n}\sum_{i=1}^{n} m_{il}(\mathbf{b},\mathbf{x}_i)$$

mimic the population expectation as closely as possible.

Note the following about the GMM estimator:

- If there are L functionally independent conditions specified and $K = L$ parameters to be estimated, it will generally be possible to find a $\mathbf{b}$ that makes the empirical moments match the population expectations.

- If $L > K$, then it will generally not be possible to make the moments all equal zero, and instead, we will have to minimize some criterion which makes the moments 'close' to zero. This is the GMM estimation problem.

- If $L < K$, then there are more parameters to be estimated than there are moment conditions specified, and, since they are functionally independent, the L moment conditions will not be sufficient to identify the parameters, and estimation will be impossible.

E21.7.1 GMM Estimation

Collect the L moment specifications in the column vector

$$\overline{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)$$

The GMM estimator is the minimum distance estimator which minimizes the quadratic form

$$q = \overline{\mathbf{m}}'\ \boldsymbol{\Sigma}\ \overline{\mathbf{m}}$$

for some choice of positive definite matrix $\boldsymbol{\Sigma}$. Different choices of $\boldsymbol{\Sigma}$ will produce different estimators. At this point, we turn to formulating the command for the GMM estimator. A brief application will be shown next, then the remaining details of using the estimator will be given. Some technical details will follow.

The essential command structure for the GMM estimator is

GMME **; Fn1** = definition of the first moment condition
; Fn2 = definition of the second moment condition
; ... up to 50 orthogonality conditions
; Labels = the symbols used for the parameters,
; Start = starting values for the optimization **$**

This basic command – note that **Σ** is not specified, requests minimization of the simple sum of squares. The default specification, therefore, is **Σ** = **I**. The number of parameters may not exceed the number of functions. The function definitions can make use of all the tools discussed earlier for specifying nonlinear regressions. They may also specify instrumental variables, as shown in the examples below.

Example 1:

Suppose $y_1...,y_n$ are a sample of n independent observations from the gamma distribution,

$$f(y) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda y} y^{P-1}, \; y \geq 0, \; \lambda, P > 0.$$

Then, the following expectations hold

$$E[y] = P/\lambda,$$

$$E[y^2] = P(P+1)/\lambda^2,$$

$$E[1/y] = \lambda/(P\text{-}1), \; P > 1,$$

$$E[\log y] = \Psi(P) - \log\lambda,$$

where $\Psi(P)$ is the Psi function, $d\log\Gamma(P)/dP$. Any two moments could be used for estimation of the parameters. To use the two which, it turns out, define the maximum likelihood estimator, consider the first and the fourth. The command would be

GMME ; Fn1 = y - p/lambda
; Fn2 = Log(y) - Psi(p) + Log(lambda)
; Start = ... the starting values
; Labels = p,lambda $

Example 2: (From Ruud (2000)

Hansen and Singleton’s classic (1982) paper on consumption and asset pricing suggests the moment equations

$$E\left[z_{tj} \left\{ \left(\frac{1+r_t}{1+\delta} \right) \left(\frac{C_t}{C_{t-1}} \right)^{\gamma-1} - 1 \right\} \middle| t-1 \right] = 0$$

for a set of instrumental variables z_{tj} where t indexes periods, C_t is consumption, r_t is return, and δ and γ are the parameters to be estimated. Ruud suggests using the instrumental variables obtained by differentiating the function in brackets with respect to 1/(1+δ) and γ, which produces,

$$z_{t1} = (1+r_{t-1})\left(\frac{C_{t-1}}{C_{t-2}}\right)^{\gamma-1} \text{ and } z_{t2} = z_{t1} \times \log\left(\frac{C_{t-1}}{C_{t-2}}\right)$$

We could set this up for estimation as follows:

```
SAMPLE    ; 1 - whatever is appropriate $
CREATE    ; ct1 = c / c[-1]
          ; lagct1 = cc1[-1]
          ; If(_obsno > 2) loglag = Log(lagct1)
          ; r1 = 1+r
          ; lagr1 = r1[-1] $
SAMPLE    ; 3 - whatever is appropriate
GMME      ; Labels = delta,gamma
          ; Start = 0,0
          ; Fn1 = (r/(1+delta) * ct1^(gamma-1) - 1) * lagr1 * ctl^(gamma-1)
          ; Fn2 = (r/(1+delta) * ct1^(gamma-1) - 1) * lagr1 * ctl^(gamma-1) * loglag $
```

We note, this can be made simpler to specify and to estimate by slightly reparameterizing the function. Let θ = 1/(1+δ) and τ = γ - 1. Making the substitutions, we would obtain the same results with

```
GMME      ; Labels = delta,gamma
          ; Start = 0,0
          ; Fn1 = (r1 * theta * ct1^tau - 1) * lagr1 * ctl^tau
          ; Fn2 = (r1 * theta * ct1^tau - 1) * lagr1 * ctl^tau * loglag $
WALD      ; Fn1 = 1/theta - 1
          ; Fn2 = tau + 1  $  (We do this to see our original parameters.)
```

E21.7.2 The Weighting Matrix

The GMM estimator defined above is consistent regardless of what positive definite matrix **Σ** is used in the minimization. (Indeed, if the problem is 'exactly identified,' that is, if there are the same number of equations as parameters), then, as has been widely documented elsewhere, the identical solution will be obtained for all matrices **Σ**. However, in terms of the efficiency of the estimator, not all choices are the same – in this discussion, we now consider only 'overidentified' problems, in which there are more equations than parameters. You may specify any matrix you like to be used in the optimization by adding

; Sigma = the name of the matrix

to the command. The name given must be that of a positive definite matrix with number of rows and columns equal to the number of moment equations.

E21.7.3 The Optimal Weighting Matrix

As noted, you may specify any matrix you wish for the weighting matrix in the criterion function. For GMM estimation, the 'optimal' weighting matrix is

$$\mathbf{\Sigma^*} = \{\mathrm{Var}[\bar{\mathbf{m}}]\}^{-1}$$

This matrix can be estimated if one has in hand any consistent estimator of the model parameters. Thus, let **b** be that estimator. Then, the estimator would be

$$\mathbf{S}^* = (1/n)(1/n)\sum_{i=1}^{n}\mathbf{m}_i(\mathbf{b},\mathbf{x}_i)\mathbf{m}_i(\mathbf{b},\mathbf{x}_i)'$$

A natural way to proceed, then, would be to use two steps:

Step 1. Use the default **Σ** = **I** to obtain the initial consistent estimates of the parameters.

Step 2. After computing **S***, redo the estimation while specifying **Σ** to be the inverse of this estimate.

When you use the **GMME** command, *LIMDEP* automatically saves **S*** for you as a matrix named *sigma*. So, to do the two steps, you would proceed as follows: The first step in GMM obtains consistent estimates. The weighting matrix is **I**.

```
GMME     ; Fn1 = definition of the first moment condition
         ; Fn2 = definition of the second moment condition
         ; ... up to 50 orthogonality conditions
         ; Labels = the symbols used for the parameters,
         ; Start = starting values for the optimization  $
```

Then, compute optimal weighting matrix as inverse of covariance matrix of moments

```
MATRIX   ; optimalw = <sigma> $
```

The second step has the same estimation problem, now with the optimal weighting matrix.

```
GMME     ; Fn1 = definition of the first moment condition
         ; Fn2 = definition of the second moment condition
         ; ... up to 50 orthogonality conditions
         ; Labels = the symbols used for the parameters,
         ; Start = starting values for the optimization
         ; Sigma = optimalw $
```

E21.7.4 Other Options

GMME is an optimization command that is largely the same as **NLSQ** and **MINIMIZE**. All other options that are available for the nonlinear optimization procedures, including output display and convergence are useable here as well. Moreover, the full range of specification options are available for defining the moment equations; that is, all functions, using quadrature, linear, bilinear, and quadratic forms, use of namelists, and so on, may all be used as they are in other optimization problems.

E21.7.5 Application

The following is Example 1 suggested earlier. It is based on 20 observations on a random variable '*y*' to which we fit a gamma distribution with parameters λ and P. (See Example 13.5 in Greene (2011).) The data are

$$\mathbf{y}' = 20.5, 31.5, 47.7, 26.2, 44.0, 8.28, 30.8, 17.2, 19.9, 9.96,\\ 55.8, 25.2, 29.0, 85.5, 15.1, 28.5, 21.4, 17.7, 6.42, 84.9$$

We first obtain the maximum likelihood estimates by maximizing the log likelihood function directly:

```
MAXIMIZE    ; Fcn = p*Log(l) - Lgm(p) - l*y + (p-1)*Log(y)
            ; Labels = l,p
            ; Start = .1,2 $
```

The GMM estimator based on the first and fourth moments will replicate the maximum likelihood estimator.

```
GMME        ; Labels =  l,p
            ; Start = .1,2
            ; Fn1 =  p/l - y  ? (changed sign of this for convenience.)
            ; Fn2 =  Log(y) - Psi(p) + Log(l) $
```

Note, however, that the asymptotic covariance matrix will differ – a finite sample difference – because of the different formulas used to do the computations. It seems useful to pursue that difference here, as we can derive the results in full detail for this simple problem. We use the BHHH estimator for the asymptotic covariance matrix for the MLE. For the gamma model above,

$$\partial \log L/\partial \lambda = \Sigma_i (P/\lambda - y)$$

$$\partial \log L/\partial P = \Sigma_i (\log\lambda - \Psi(P) + \log y)$$

Note that the first order conditions for the MLE are $n\bar{\mathbf{m}} = \mathbf{0}$. Let $\mathbf{M}$ be the 20×2 matrix whose ith row is the derivative shown above for the ith observation. Then, the estimator of the asymptotic covariance matrix for the MLE is

$$\text{Est.Asy.Var[MLE]} = (\mathbf{M}'\mathbf{M})^{-1}.$$

For the GMM estimator, $\mathbf{\Sigma} = \mathbf{I}$ while $\mathbf{G}$ turns out to be a sum of constants, so the n disappears;

$$\mathbf{G} = \begin{bmatrix} -P/\lambda^2 & 1/\lambda \\ 1/\lambda & -\Psi'(P) \end{bmatrix}$$

Inserting these in the formula for the asymptotic covariance matrix of the GMM estimator, we obtain after canceling

$$\text{Est.Asy.Var}[GMM] = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{M}'\mathbf{M}\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}$$

As can be seen, this differs from the formula for the MLE. Since $\mathbf{G}'\mathbf{G}$ and $(1/n)\mathbf{M}'\mathbf{M}$ converge to the same matrix, we see that the difference is due to finite sample variation.

Finally, we obtain the full GMM estimator, using all four moment equations, and two steps to obtain the efficient estimator at the second step.

This is the maximum likelihood estimator.

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable                 Function
Log likelihood function            85.37567
Restricted log likelihood            .00000
Chi squared [    2 d.f.]          170.75134
Significance level                   .00000
Estimation based on N =      20, K =   0
Inf.Cr.AIC  = -170.751 AIC/N =    -8.538
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .07707***      .02708     2.85  .0044      .02400     .13014
       P|    2.41060***      .87683     2.75  .0060      .69206    4.12915
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the GMM estimator based on the same two moments as used by the maximum likelihood estimator. Not surprisingly, the parameter estimates are the same.

```
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    .1203629E-14
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function            .00000
Restricted log likelihood          .00000
Chi squared [   2 d.f.]            .00000
Significance level                1.00000
GMM Criterion function             .00000
Degrees of freedom = #eqn-#parms =    0
Significance level                1.00000
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .07707***      .02555     3.02  .0026      .02698    .12716
       P|    2.41060***      .60848     3.96  .0001     1.21800   3.60321
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the GMM estimators based on all four moments.

```
GMME    ; Labels = l,p
        ; Start = .1,2
        ; Fn1 = y-p/l
        ; Fn2 = 1 / y - l/(p-1)
        ; Fn3 = y^2 - p*(p+1)/l^2
        ; Fn4 = Log(y) - Psi(p) + Log(l) $
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function            .00180
Restricted log likelihood          .00000
Chi squared [   2 d.f.]            .00361
Significance level                 .99820
GMM Criterion function             .00180
Degrees of freedom = #eqn-#parms =    2
Significance level                 .99910
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .06580***      .01890     3.48  .0005      .02876    .10284
       P|    2.05830***      .50345     4.09  .0000     1.07156   3.04504
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the second step, efficient GMM estimators based on the optimal weighting matrix.

MATRIX ; optimalw = <sigma>$
GMME ; Labels = l,p
; Start = .1,2
; Fn1 = y-p/l
; Fn2 = 1/y - l/(p-1)
; Fn3 = y^2 - p*(p+1)/l^2
; Fn4 = Log(y) - Psi(p) + Log(l)
; Sigma = optimalw $

```
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function         1.97522
Restricted log likelihood        .00000
Chi squared [   2 d.f.]         3.95043
Significance level               .13873
GMM Criterion function          1.97522
Degrees of freedom = #eqn-#parms =    2
Significance level               .37247
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .12449***     .03403     3.66  .0003      .05780    .19118
       P|    3.35894***     .64628     5.20  .0000     2.09225   4.62563
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E21.7.6 Technical Details for the GMM Estimator

The underlying theory for the GMM estimator is well documented in the current literature, including the current textbooks such as Greene (2011), Ruud (2000), and Hayashi (2000), so it will be omitted here, and only final results will be shown.

The estimation criterion used is

$$q = \frac{1}{2}\,\bar{\mathbf{m}}'\boldsymbol{\Sigma}\bar{\mathbf{m}}$$

(The 1/2 is purely for convenience - it allows the 2 to disappear from the derivatives.)

NOTE: The output displayed by the program reports 2q, not q. That is, your final results will report the value of the quadratic form, not one half times it.

The first order conditions for minimizing q are

$$\frac{\partial q}{\partial \boldsymbol{\beta}} = \mathbf{G}' \boldsymbol{\Sigma} \, \bar{\mathbf{m}} = \mathbf{0},$$

where $$\mathbf{G} = \frac{\partial \bar{\mathbf{m}}}{\partial \boldsymbol{\beta}'}$$

There are L equations and K parameters and $L \geq K$. Thus, $\mathbf{G}$ is an $L \times K$ matrix of partial derivatives. (Note, as well, that $\mathbf{G}$ is a sample mean.) If there are K moment equations used to identify the K parameters, then assuming that $\boldsymbol{\Sigma}$ is positive definite and that the moment equations are functionally independent so that $\mathbf{G}$ has an inverse, then we can premultiply the first order condition by $(\mathbf{G}' \boldsymbol{\Sigma})^{-1}$ and obtain the simpler necessary condition, $\bar{\mathbf{m}} = \mathbf{0}$. The solution to this is independent of $\boldsymbol{\Sigma}$, which establishes the earlier claim that $\boldsymbol{\Sigma}$ is irrelevant to the solution to an exactly identified problem.

The asymptotic covariance matrix is computed using the estimated parameters, and

$$\text{Est.Var}[\mathbf{b}] = [\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G}]^{-1} \mathbf{G}' \boldsymbol{\Sigma} \mathbf{S}^* \boldsymbol{\Sigma} \mathbf{G} [\mathbf{G}' \boldsymbol{\Sigma} \mathbf{G}]^{-1}$$

where $\mathbf{S}^*$ was defined earlier. If you have specified the optimal weighting matrix, $\boldsymbol{\Sigma} = (\mathbf{S}^*)^{-1}$, then the estimated variance reduces to the familiar result,

$$\text{Est.Var}[\mathbf{b}] = [\mathbf{G}' (\mathbf{S}^*)^{-1} \mathbf{G}]^{-1}$$

If the model is exactly identified, then q is minimized at zero. (See the example above.) If not, then q will be positive. The theoretical result that $2q$ will have a limiting chi squared distribution with degrees of freedom equal to the number of overidentifying restrictions (equations minus parameters) can be used to test restrictions in this framework. (The multiplier, 2, appears because in our formulation of the problem, we initially divided by 2.) For two nested models, with $q0$ being the unrestricted one and $q1$ embodying the restrictions, $2(q1 - q0)$ can be used to test the restrictions – refer this statistic to the chi squared table with degrees of freedom equal to the number of restrictions.

E22: 2SLS for Panel Data

E22.1 Introduction

This chapter will present estimation of linear models for panel data by two stage least squares and instrumental variables. It combines the results in Chapters E16-18 (panel data models) and E21 (instrumental variables estimation).

E22.2 Application

The data used in the application below were analyzed in Cornwell and Rupert (1988). (See Baltagi (2005), page 122 for further analysis.) Unfortunately, the data are not available at the archive site for the journal. They were downloaded from the website for Baltagi's text, http://www.wiley.com/legacy/wileychi/baltagi/supp/WAGES.xls. These data are a microeconomic panel data set of observations on 595 individuals for seven years. Variables in the data set are:

exp	=	work experience
wks	=	weeks worked
occ	=	occupation, 1 if blue collar
ind	=	1 if manufacturing industry
south	=	1 if resides in south
smsa	=	1 if resides in a city (SMSA)
ms	=	1 if married
fem	=	1 if female
union	=	1 if wage set by union contract
ed	=	years of education
blk	=	1 if individual is black
lwage	=	log of wage

The model estimated below is

$$\log wks_{it} = \beta_1 + \beta_2 lwage_{it} + \beta_3 occ_{it} + \beta_4 fem_{it} + \beta_5 ed_{it} + \varepsilon_{it} + u_i$$

The instrumental variables used for the *lwage* variable are *union* membership and *south*, both dummy variables.

E22.3 2SLS Estimation with Fixed Effects

A basic estimator for a fixed effects model with one endogenous variable on the Rhs is obtained as follows:

```
SETPANEL    ; Group = id variable ; Pds = count variable to create $
REGRESS     ; Lhs = endogenous variable
            ; Rhs = all instruments
            ; Panel
            ; Keep = fittedy
            ; Output = 5 $
REGRESS     ; Lhs = dependent variable
            ; Rhs = endogenous variable, other variables
            ; Panel
            ; Inst = fittedy $
```

The list of instrumental variables includes only the predicted value for the one variable for which instruments are needed. The command can be extended to models with more than one endogenous variable on the right hand side, as well by producing a fitted value for each one, including the more than one endogenous variables at the beginning of the Rhs list, and including the corresponding list of fitted values in the Inst list. Note, the Rhs list should not include *one*.

The following computes the 2SLS estimates for the *logwks* equation in the application

```
CREATE      ; i = Trn(7,0) $
SETPANEL    ; Group = i ; Pds = ti $
REGRESS     ; Lhs = lwage
            ; Rhs = one,occ,fem,ed,union,south
            ; Panel
            ; Keep = fittedy
            ; Output = 5 $
REGRESS     ; Lhs = logwks
            ; Rhs = lwage, occ,fem,ed
            ; Panel
            ; Inst = fittedy $
```

```
+----------------------------------------------------------------+
| Variable = ____________ Variable Groups    Max    Min   Average |
| TI         Group sizes  I            595      7      7       7.0 |
+----------------------------------------------------------------+
Command requests fitted values only. Output is suppressed

-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGWKS   Mean                 =       3.83748
             Standard deviation   =        .14796
             No. of observations  =          4165  Degrees of freedom
Regression   Sum of Squares       =       .727604          4
Residual     Sum of Squares       =       90.4309       4160
Total        Sum of Squares       =       91.1585       4164
             Standard error of e  =        .14744
Fit          R-squared            =        .00798  R-bar squared =   .00703
Model test   F[  4,  4160]        =       8.36780  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    2065.85755  Akaike I.C.   = -3.82748
             Restricted (b=0)     =    2049.16888  Bayes  I.C.   = -3.81988
             Chi squared [  4]    =      33.37733  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]    =     589.18552  Prob C2 > C2* =   .00000
[High values of  LM favor FEM/REM over base model]
Baltagi-Li form of LM Statistic   =     589.18552  [= BP if balanced panel]
Moulton/Randolph form:SLM N[0,1]  =     24.578092
--------------------------------------------------
Panel Data Analysis of LOGWKS              [ONE way]
               Unconditional ANOVA (No regressors)
Source         Variation   Deg. Free.   Mean Square
Between         30.49694        594.         .05134
Residual        60.66157       3570.         .01699
Total           91.15850       4164.         .02189
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   LWAGE|    .02314***      .00736     3.14  .0017       .00871    .03756
     OCC|   -.00487         .00596     -.82  .4137     -.01656    .00681
     FEM|   -.02203***      .00813    -2.71  .0067     -.03796   -.00609
      ED|   -.00161         .00110    -1.45  .1459     -.00377    .00056
Constant|   3.70860***      .04822    76.91  .0000     3.61409   3.80311
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=LOGWKS   Mean                 =       3.83748
             Standard deviation   =        .14796
             No. of observations  =          4165  Degrees of freedom
Regression   Sum of Squares       =       .708599           598
Residual     Sum of Squares       =       90.4499          3566
Total        Sum of Squares       =       91.1585          4164
             Standard error of e  =        .15926
Fit          R-squared            =        .00777  R-bar squared =  -.15862
Model test   F[598,  3566]        =        .04672  Prob F > F*   =  1.00000
Diagnostic   Log likelihood       =    2065.41994  Akaike I.C.   = -3.54204
             Restricted (b=0)     =    2049.16888  Bayes  I.C.   = -2.63103
             Chi squared [598]    =      32.50211  Prob C2 > C2* =  1.00000
```

```
--------------------------------------------------
Panel:Groups Empty      0,     Valid data      595
             Smallest   7,     Largest           7
             Average group size in panel      7.00
Variances    Effects a(i)         Residuals e(i,t)
              8.756725                     .025365
These  2 variables have no within group variation.
FEM      ED
F.E. estimates are based on a generalized inverse.
--------+-----------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z     |z|>Z*         Interval
--------+-----------------------------------------------------------------
   LWAGE|     .36781          .23562     1.56  .1185      -.09399    .82961
     OCC|     .05496**        .02179     2.52  .0117       .01225    .09767
     FEM|        0.0    .....(Fixed Parameter).....
      ED|        0.0    .....(Fixed Parameter).....
--------+-----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
--------------------------------------------------------------------------
```

```
+--------------------------------------------------------------------+
|             Test Statistics for the Regression Model               |
+--------------------------------------------------------------------+
|       Model              Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only      2049.16894         91.15850     .00000 |
|(2)  Group effects only      2897.34924         60.66157     .33455 |
|(3)  X - variables only      2065.85761         90.43090     .00798 |
|(4)  X and group effects     2065.42000         90.44991     .00777 |
+--------------------------------------------------------------------+
```

E22.4 IV Estimators for Panel Data

Two stage least squares for panel data estimators is extended to include random effects. In the preceding section, it is shown how you may include predicted values of the regressors in the right hand side of the equation. The estimator then adjusts the computation of variance estimators for the presence of the fitted value. The extension described here adds a full two stage least squares treatment for other panel data models. The essential model is

$$y_{it} = \alpha_i + \boldsymbol{\beta_1}'\mathbf{x}_{1,it} + \boldsymbol{\beta_2}'\mathbf{x}_{2,it} + \varepsilon_{it}.$$

In the different specifications, α_i may be a fixed effect, a random effect, or unspecified. (The chosen estimator is robust to either.) Variables in $\mathbf{x}_{2,it}$ are assumed to be correlated with ε_{it}. A set of instrumental variables, $\mathbf{z}_{it}$ is provided. Five estimators are supported: conventional two stage least squares, fixed effects, first differences, random effects and group means. The estimator is the linear estimator for these panel data settings, using instrumental variables rather than OLS, GLS, or FGLS. Let $\mathbf{X}_i$ denote the K_1+K_2 columns in the structural variables, and let $\mathbf{Z}_i$ denote the K_1+K_z instrumental variables – note, presumably, some of the variables in $\mathbf{Z}_i$ are those in $\mathbf{x}_{1,it}$ and there must be at least K_2 additional instrumental variables in $\mathbf{Z}_i$ so that the model is identified. The matrices have T_i rows or observations. Then, the various estimators of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1{}',\boldsymbol{\beta}_2{}')'$ are

Two Stage Least Squares (2SLS)

$$\hat{\boldsymbol{\beta}}_{2S} = \left[\sum_{i=1}^{N} \hat{\mathbf{X}}_i{}'\mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \hat{\mathbf{X}}_i{}'\mathbf{y}_i\right]$$

where $\hat{\mathbf{X}}_i = \mathbf{Z}_i(\mathbf{Z}_i{}'\mathbf{Z}_i)^{-1}\mathbf{Z}_i{}'\mathbf{X}$

Fixed Effects (FE)

$$\hat{\boldsymbol{\beta}}_{FE} = \left[\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_i^0 \mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_i^0 \mathbf{y}_i\right]$$

where $\mathbf{M}_i^0$ is the $T_i \times T_i$ matrix that creates deviations from means.

First Differences (FD)

$$\hat{\boldsymbol{\beta}}_{FD} = \left[\sum_{i=1}^{N} (\mathbf{Z}_i' \mathbf{D}_i')(\mathbf{D}_i\mathbf{X}_i)\right]^{-1} \left[\sum_{i=1}^{N} (\mathbf{Z}_i' \mathbf{D}_i')(\mathbf{D}_i\mathbf{y}_i)\right]$$

where $\mathbf{D}_i$ is the $(T_i\text{-}1) \times T_i$ matrix that creates first differences. The first observation is lost.

Random Effects (RE)

$$\hat{\boldsymbol{\beta}}_{RE} = \left[\sum_{i=1}^{N} \mathbf{Z}_i' \hat{\boldsymbol{\Omega}}_i^{-1} \mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{Z}_i' \hat{\boldsymbol{\Omega}}_i^{-1} \mathbf{y}_i\right]$$

where $\boldsymbol{\Omega}_i^{-1}$ is the $T_i \times T_i$ matrix that creates partial deviations from means for the two step FGLS estimator. The variance components are computed using the simple two stage least squares estimator and the fixed effects estimator.

Group Means (MEANS)

$$\hat{\boldsymbol{\beta}}_{MEANS} = \left[\sum_{i=1}^{N} (\mathbf{Z}_i' \mathbf{a}_i')(\mathbf{a}_i\mathbf{X}_i)\right]^{-1} \left[\sum_{i=1}^{N} (\mathbf{Z}_i' \mathbf{a}_i')(\mathbf{a}_i\mathbf{y}_i)\right]$$

where $\mathbf{a}_i$ is a $1 \times T_i$ vector with all elements equal to $1/T_i$; it creates a group mean vector.

Some of these estimators (2SLS, RE, MEANS) are inconsistent under the fixed effects assumption. All are consistent under the random effects assumption, but some (2SLS, FE) are inefficient. The fixed effects and first difference estimators are always consistent. Also, the treatment of the constant term differs from one to the next. The estimator will sort this out, as can be seen in the example below.

Commands

The command for this estimator is

2SLS ; Lhs = dependent variable
; Rhs = set of independent variables
; Inst = full list of instruments

plus, for any of the panel data estimators,

; Pds = the usual panel data setup, balanced or not
; Panel

and exactly one of the following

; Fixed Effects
; Random Effects
; Differences to use first differences
; Means to use group means $

Options **; Keep**, **; Res** and **; Covariance Matrix** operate as usual. With **; Fixed Effects**, **; Par** saves the *alphafe* matrix containing the estimated constant terms. If the model contains time invariant variables, as in the example below, the fixed effects estimator uses generalized inverses. This will be noted in the results. But don't expect good results.

We fit the model by conventional 2SLS and then using the four estimators detailed above.

```
2SLS    ; Lhs = logwks
        ; Rhs = one,lwage,occ,fem,ed
        ; Inst = one,occ,fem,south,union,ed $
2SLS    ; Lhs = logwks
        ; Rhs = one,lwage,occ,fem,ed
        ; Inst = one,occ,fem,south,union,ed
        ; Panel ; Fixed effects $
2SLS    ; Lhs = logwks
        ; Rhs = one,lwage,occ,fem,ed
        ; Inst = one,occ,fem,south,union,ed
        ; Panel ; Random Effects $
2SLS    ; Lhs = logwks
        ; Rhs = one,lwage,occ,fem,ed
        ; Inst = one,occ,fem,south,union,ed
        ; Panel ; Differences $
2SLS    ; Lhs = logwks
        ; Rhs = one,lwage,occ,fem,ed
        ; Inst = one,occ,fem,south,union,ed
        ; Panel ; Means $
```

(Some superfluous results and results that are repeated in the outputs are omitted.)

```
----------------------------------------------------------------------------
Two stage    least squares regression ...........
LHS=LOGWKS   Mean                 =        3.83748
             Standard deviation   =         .14796
             Number of observs.   =           4165
Model size   Parameters           =              5
             Degrees of freedom   =           4160
Residuals    Sum of squares       =        126.324
             Standard error of e  =         .17426
Fit          R-squared            =        -.38742
             Adjusted R-squared   =        -.38876
Not using OLS or no constant. Rsqrd & F may be < 0
Model was estimated on Jun 09, 2011 at 08:42:56 AM
Instrumental Variables:
ONE       OCC       FEM       SOUTH     UNION     ED
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|    5.19877***      .23141    22.47  .0000     4.74522   5.65233
   LWAGE|    -.21748***      .03724    -5.84  .0000     -.29047   -.14448
     OCC|    -.04186***      .00898    -4.66  .0000     -.05946   -.02425
     FEM|    -.14102***      .02032    -6.94  .0000     -.18085   -.10118
      ED|     .00996***      .00218     4.58  .0000      .00569    .01422
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
Panel Data Instrumental Variables Estimator. LHS Variable =   LOGWKS
Fixed effects (within) estimator y(i,t) = a(i) + x(i,t)b + e(i,t)
Consistent for both FE and RE models, (in)efficient for FE (RE) model
Mean of   LOGWKS =     3.83748   Std. Dev. of   LOGWKS =      .14794
Estimated residual standard deviation     =      .12326
Sum of squared deviations (y - fitted)^2 =    63.28094
Correlation actual and fitted values      =      .30738
(Note, this is not a proportion of variation explained.)
Panel group sizes: Minimum =    7, Maximum =    7, Mean =      7.000
Total sample size is     4165 observations in     595 groups.
Model has time invariant variables FEM     , ED      ,         , ...
FE model is unidentified. Using G2 inverse for RE and 1st differences
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|       0.0    .....(Fixed Parameter).....
   LWAGE|    -.01978         .01230    -1.61  .1077     -.04389    .00432
     OCC|    -.11347         .13234     -.86  .3912     -.37285    .14590
     FEM|       0.0    .....(Fixed Parameter).....
      ED|       0.0    .....(Fixed Parameter).....
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Panel Data Instrumental Variables Estimator. LHS Variable =   LOGWKS
Random effects (fgls)  estimator y(i,t) = a + x(i,t)b + e(i,t) + u(i)
Consistent and efficient for RE model, inconsistent for FE model
Estimated residual standard deviation     =      .25308
Sum of squared deviations (y - fitted)^2 =   266.77186
Correlation actual and fitted values      =      .00286
(Note, this is not a proportion of variation explained.)
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|    3.28914       2.06493     1.59  .1112     -.75805   7.33633
   LWAGE|    -.02477        .03759     -.66  .5100     -.09845    .04892
     OCC|     .06559        .37324      .18  .8605     -.66594    .79712
     FEM|    -.52323        .43400    -1.21  .2280    -1.37386    .32740
      ED|     .05578        .13074      .43  .6696     -.20046    .31202
--------+-----------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Panel Data Instrumental Variables Estimator. LHS Variable =   LOGWKS
First difference estimator for  Dy(i,t) = Dx(i,t)b + De(i,t)
Consistent but inefficient for both FE and RE models
Estimated residual standard deviation     =     2.21439
Sum of squared deviations (y - fitted)^2 = 20423.21660
Correlation actual and fitted values      =      .00519
(Note, this is not a proportion of variation explained.)
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|        0.0    .....(Fixed Parameter).....
   LWAGE|     .24600        .99573      .25  .8049    -1.70560   2.19760
     OCC|    -.02305        .56598     -.04  .9675    -1.13235   1.08626
     FEM|        0.0    .....(Fixed Parameter).....
      ED|        0.0    .....(Fixed Parameter).....
--------+-----------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Panel Data Instrumental Variables Estimator. LHS Variable =   LOGWKS
Grp means (between) estimator ybar(i) = a + xbar(i)b + ebar(i) + u(i)
Consistent only for RE model, not FE; inefficient for RE model
Estimated residual standard deviation     =      .17856
Sum of squared deviations (y - fitted)^2 =   132.79147
Correlation actual and fitted values      =      .00176
(Note, this is not a proportion of variation explained.)
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  LOGWKS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|    5.34743***     .23565    22.69  .0000     4.88557   5.80929
   LWAGE|    -.23703***     .03766    -6.29  .0000     -.31085   -.16321
     OCC|    -.05897***     .01028    -5.74  .0000     -.07912   -.03881
     FEM|    -.15259***     .02070    -7.37  .0000     -.19317   -.11202
      ED|     .00933***     .00218     4.29  .0000      .00507    .01360
--------+-----------------------------------------------------------------------
```

E23: Hausman-Taylor and Arellano-Bond Estimators

E23.1 Introduction

This chapter will detail estimation of several linear models for panel data. The essential structure for most of them is an 'effects' model,

$$y_{it} = \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

in which variation across groups (individuals) or time is captured in simple shifts of the regression function – i.e., changes in the intercepts. These models are the *fixed effects* (FE) and *random effects* (RE) models. Several variations on this structure can be analyzed with this estimator, including both one and two factor models, models of autocorrelation, and simultaneous equations models. This chapter also presents some major extensions including multifactor random effects models, the Hausman and Taylor estimator for random effects and the Arellano, Bond and Bover estimator for dynamic panel data models.

E23.2 The Hausman and Taylor Estimator for Random Effects

Hausman and Taylor's (1981) estimator for the random effects model is provided to overcome one of the major shortcomings of the REM, the possible correlation between the independent variables and the random effects. The following will sketch the technical aspects; the user is referred to their paper for full details.

The random effects model is formulated with the possibility that there may be time invariant independent variables. We thus write it in this form:

$$y_{it} = \boldsymbol{\beta}_1'\mathbf{x}_{1it} + \boldsymbol{\beta}_2'\mathbf{x}_{2it} + \boldsymbol{\gamma}_1'\mathbf{f}_{1i} + \boldsymbol{\gamma}_2'\mathbf{f}_{2i} + \varepsilon_{it} + u_i, \text{ where } \boldsymbol{\beta} = (\boldsymbol{\beta}_1',\boldsymbol{\beta}_2')' \text{ and } \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1',\boldsymbol{\gamma}_2')'$$

where

$$E[u_i] = 0,\ \mathrm{Var}[u_i] = \sigma_u^2\ \ \mathrm{Cov}[\varepsilon_{it},u_i] = 0,$$

$$\mathrm{Var}[\varepsilon_{it} + u_i] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,$$

$$\mathrm{Corr}[\varepsilon_{it} + u_i,\varepsilon_{is} + u_i] = \rho = \sigma_u^2 / \sigma^2.$$

There are four sets of variables in the model,

	Uncorrelated with u_i	Correlated with u_i
Time varying	$\mathbf{x}_{1it}$ is KX_1 variables	$\mathbf{x}_{2it}$ is KX_2 variables
Time invariant	$\mathbf{f}_{1i}$ is KF_1 variables	$\mathbf{f}_{2i}$ is KF_2 variables

Note that as stated, the model already embodies an important assumption. The formulation assumes that you can distinguish a set of variables $\mathbf{x}_1$ that is uncorrelated with u_i. In *LIMDEP*'s formulation of the model, any of the remaining three sets of variables are optional; your model may include any or all of these remaining three sets, but it must include set $\mathbf{x}_1$. *For identification purposes, KX_1 must be at least as large as KF_2 (KF_2 may be zero).* At the outset, we note that if your model contains neither $\mathbf{x}_2$ nor $\mathbf{f}_2$, then you should not use this estimator, as you can use the ordinary random effects estimator.

By construction, any OLS or GLS estimators of this model are inconsistent when KX_2 or KF_2 are positive (that is, when the model contains variables that are correlated with the random effects). Hausman and Taylor have proposed an instrumental variables estimator that uses only the information within the model (i.e., as already stated). The strategy for estimation is based on the following logic. First, by taking deviations from group means, we find that

$$y_{it} = \boldsymbol{\beta}_1'(\mathbf{x}_{1it} - \overline{\mathbf{x}}_{1i}) + \boldsymbol{\beta}_2'(\mathbf{x}_{2it} - \overline{\mathbf{x}}_{2i}) + \varepsilon_{it} - \overline{\varepsilon}_i$$

which implies that $\boldsymbol{\beta}$ can be consistently estimated by least squares, in spite of the correlation between $\mathbf{x}_2$ and u. This is the familiar, fixed effects, least squares dummy variable estimator. Now, in the original model, Hausman and Taylor show that the group means can be used as (KX_1+KX_2) instrumental variables for estimation of $(\boldsymbol{\beta},\boldsymbol{\gamma})$. Since $\mathbf{f}_1$ is uncorrelated with the disturbances, it can likewise serve as a set of KZ_1 instrumental variables. That leaves a necessity for KF_2 instrumental variables. The authors show that the group means for $\mathbf{x}_1$ can serve as these remaining instruments, and the model will be identified so long as KX_1 is greater than or equal to KF_2. As before, feasible GLS is better than OLS, and available. Likewise, FGLS is an improvement over simple IV estimation of the model, which is consistent but inefficient.

The authors propose the following set of steps:

Step 1. Use consistent but inefficient estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to estimate the variance components.

Step 2. Use weighted FGLS with instrumental variables to take full advantage of the known information about the variances at a second step.

The specific procedure is as follows:

Step 1. Obtain the LSDV (fixed effects) estimator of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\boldsymbol{\beta}_2')'$ based on $\mathbf{x}_1$ and $\mathbf{x}_2$. The residual variance estimator from this step is a consistent estimator of σ_ε^2.

Step 2. Form the within groups residuals from this regression at Step 1, e_{it}. Stack the group means of these residuals in a full sample length data vector. Thus, $e_{it}^* = \overline{e}_i$, $i = 1,...,T_i$, i = 1,...,N. These group means are used as the dependent variable in a two stage least squares regression on $\mathbf{f}_1$ and $\mathbf{f}_2$ with instrumental variables $\mathbf{f}_1$ and $\mathbf{x}_1$. (Note the identification requirement that KX_1, the number of columns in $\mathbf{x}_1$ be at least as large as KF_2, the number of columns in $\mathbf{f}_2$.) This provides a consistent estimator of $\boldsymbol{\gamma}$. The residual variance in this regression is a consistent estimator of $\sigma^{*2} = \sigma_u^2 + Q\sigma_\varepsilon^2$ where $Q = \text{plim}(1/N)\Sigma_i(1/T_i)$. (This is just $1/T$ in a balanced sample, but we do not require this.) From this estimator and the estimator of σ_ε^2 in Step 1, we deduce an estimator of σ_u^2.

Step 3. The final step is a weighted instrumental variable estimator. The transformation of y_{it} and $(\mathbf{x}_{1it},\mathbf{x}_{2it},\mathbf{f}_{1i},\mathbf{f}_{2i})$ is

$$v_{it}^* = v_{it} - (1 - \theta_i)\,\overline{v}_i \text{ where } \theta_i = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i\sigma_u^2}}$$

where v_{it} denotes any of the aforementioned variables and $\overline{v}_i$ denotes a group mean. Note in the case of the time invariant variables, the group mean is the original variable, and the transformation just multiplies the variable by θ_i. The instrumental variables are $\mathbf{x}_{1it} - \overline{\mathbf{x}}_{1i}$, $\mathbf{x}_{2it} - \overline{\mathbf{x}}_{2i}$, $\mathbf{z}_{1i}$ and $\overline{\mathbf{x}}_{1i}$. Note for the fourth set of instruments, the group mean is repeated for each member of the group.

In order to implement this estimator with *LIMDEP*, several steps are required, but the final, most complicated one has been completely automated. The program below shows the set of steps. Most of the steps for this estimator use familiar parts of the program. *The final* **REGRESS** *command contains the important settings which specifically request the Hausman and Taylor estimator.*

This is a program template for application of the Hausman and Taylor estimator for the random effects model.

1. Define four sets of variables. Note, any of *x*2, *f*1, or *f*2 may contain no variables. In this case, in the **CALC** command, set the associated count to zero. Be sure to define the compound namelists appropriately as well. If the model contains a constant term, include one in *f*1.

```
CREATE or otherwise set up group ID.  We call this i $
SETPANEL    ; Group = i ; Pds = ti $
NAMELIST    ; x1 = ... the set of variables  $
NAMELIST    ; x2 = ... the set of variables  $
NAMELIST    ; f1 = ... one, the set of variables $
NAMELIST    ; f2 = ... the set of variables  $
NAMELIST    ; x = x1,x2 ; f = f1,f2 ; exog = x1,f1 ; all = x1,x2,f1,f2 $
CREATE      ; y = the dependent variable $
CALC        ; kx1 = Col(x1) ; kx2 = Col(x2) ; kf1 = Col(f1) ; kf2 = Col(f2) $
```

2. Compute the LSDV estimator based only on [**x1**,**x2**] to produce β and σ_ε^2. The panel specification may be by **; Pds = number** or **; Str = variable.**

```
REGRESS     ; Lhs = y ; Rhs = x ; Fixed ; Panel  $
MATRIX      ; bw = b $
CALC        ; s2e = ssqrd$
```

3. Form within groups residuals, then get the group means, expanded to the full sample length. This is a special command created for this purpose. Panel specification should now be the one created automatically by the regression above. This command also creates the estimate of *Q* and calls it *avg_ti*. See Step 4 below. (It can be done with **CREATE**, but we need *Q* as well so we use the regression.)

```
CREATE      ; dwit = y - x'bw $
CREATE      ; dwi = Group Mean(dwit, Pds = ti) $
```

4. Regress these group means on [**f1**,**f2**] with instruments [**f1**,**x1**] to estimate γ and σ_*^2. (This is just 2SLS.) Then get the variance components estimator.

```
2SLS        ; Lhs = dwi ; Rhs = f ; Inst = exog $
CALC        ; s2s = ssqrd ; s2u = s2s - avg_ti*s2e $
```

5. This is the Hausman and Taylor procedure which has been automated. The model is set up as usual for the random effects model. The **; Start** specification requests the estimator. This must provide six values, exactly as shown below.

```
REGRESS     ; Lhs = y ; Rhs = all ; Panel ; Random
            ; Start = kx1,kx2,kf1,kf2,s2e,s2u $
```

To illustrate the Hausman and Taylor estimator, we will fit a log wage equation using the Cornwell and Rupert data examined in the previous section. The model is the one specified by Cornwell and Rupert,

$$\begin{aligned}\log wage_{it} = \; & \beta_{1,1}wks_{it} + \beta_{1,2}south_{it} + \beta_{1,3}smsa_{it} + \beta_{1,4}ms_{it} \\ & + \beta_{2,1}exp_{it} + \beta_{2,2}exp_{it}^2 + \beta_{2,3}occ_{it} + \beta_{2,4}ind_{it} + \beta_{2,5}union_{it} \\ & + \gamma_{1,1} + \gamma_{1,2}fem_i + \gamma_{1,4}blk_i \\ & + \gamma_{2,1}ed_i + \varepsilon_{it} + u_i\end{aligned}$$

We take weeks worked and union membership to be endogenous in the model. The following commands adapt the Hausman and Taylor routine to this specification. Results follow. We note, this is precisely the model specified by Cornwell and Rupert, and these are their data. Our results resemble theirs, but are not close enough to 'match' within rounding error. One possible explanation is that the estimator depends crucially on the two variance estimators, and there are numerous ways to estimate them. Cornwell and Rupert do not document how they did this computation.

? [Specific for this application]

```
CREATE      ; i = Trn(7,0) $
SETPANEL    ; Group = i ; Pds = ti $
NAMELIST    ; x1= wks,south,smsa,ms $
NAMELIST    ; x2= exp,exp*exp,occ,ind,union $
NAMELIST    ; f1 = one,fem,blk $
NAMELIST    ; f2 = ed $
NAMELIST    ; x = x1,x2
            ; f  = f1,f2
            ; exog = x1,f1
            ; all = x1,x2,f1,f2 $
```

? [Generic for estimation of Hausman and Taylor]

```
CREATE    ; y = lwage $
CALC      ; kx1 = Col(x1) ; kx2 = Col(x2)
          ; kf1 = Col(f1) ; kf2 = Col(f2) $
REGRESS   ; Lhs = y
          ; Rhs = x
          ; Fixed ; Panel $
MATRIX    ; bw = b $
CALC      ; s2e = ssqrd $
CREATE    ; dwit =  y - x'bw $
CREATE    ; dwi = Group Mean(dwit, Pds = ti) $
2SLS      ; Lhs = dwi ; Rhs = f ; Inst = exog $
CALC      ; s2s = ssqrd ; s2u = s2s - avg_ti*s2e $
REGRESS   ; Lhs = y
          ; Rhs = all
          ; Panel ;  Random
          ; Start = kx1,kx2,kf1,kf2,s2e,s2u $
```

The results of all the procedures are shown below.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y        Mean                 =       6.67635
             Standard deviation   =        .46151
             No. of observations  =          4165  Degrees of freedom
Regression   Sum of Squares       =       279.778           9
Residual     Sum of Squares       =       607.127        4155
Total        Sum of Squares       =       886.905        4164
             Standard error of e  =        .38226
Fit          R-squared            =        .31545  R-bar squared =   .31397
Model test   F[  9,  4155]        =     212.74706  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =   -1899.53722  Akaike I.C.   = -1.92093
             Restricted (b=0)     =   -2688.80603  Bayes  I.C.   = -1.90572
             Chi squared [  9]    =    1578.53762  Prob C2 > C2* =   .00000
B-P test     Chi squared [  1]    =    3881.34495  Prob C2 > C2* =   .00000
[High values of  LM favor FEM/REM over base model]
Baltagi-Li form of LM Statistic   =    3881.34495  [= BP if balanced panel]
Moulton/Randolph form:SLM N[0,1]  =     63.316479
--------------------------------------------------
Panel Data Analysis of Y                   [ONE way]
               Unconditional ANOVA (No regressors)
Source         Variation   Deg. Free.    Mean Square
Between        646.25374        594.        1.08797
Residual       240.65119       3570.         .06741
Total          886.90494       4164.         .21299
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     WKS|     .00446***      .00118     3.78  .0002      .00215    .00677
   SOUTH|    -.11368***      .01345    -8.45  .0000     -.14004   -.08732
    SMSA|     .15858***      .01303    12.17  .0000      .13305    .18411
      MS|     .32033***      .01585    20.21  .0000      .28927    .35139
     EXP|     .03611***      .00236    15.32  .0000      .03149    .04073
 EXP*EXP|    -.00066***   .5186D-04   -12.63  .0000     -.00076   -.00055
     OCC|    -.31762***      .01349   -23.54  .0000     -.34407   -.29117
     IND|     .03213**       .01277     2.52  .0119      .00711    .05716
   UNION|     .06975***      .01392     5.01  .0000      .04246    .09704
Constant|    5.88024***      .06035    97.43  .0000     5.76194   5.99853
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=Y        Mean                 =       6.67635
             Standard deviation   =        .46151
             No. of observations  =          4165  Degrees of freedom
Regression   Sum of Squares       =       804.638        603
Residual     Sum of Squares       =       82.2673       3561
Total        Sum of Squares       =       886.905       4164
             Standard error of e  =        .15199
Fit          R-squared            =        .90724  R-bar squared =   .89154
Model test   F[603,  3561]        =      57.76006  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    2262.88725  Akaike I.C.   = -3.63446
             Restricted (b=0)     =   -2688.80603  Bayes  I.C.   = -2.71585
             Chi squared [603]    =    9903.38656  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)   =       .146506
--------------------------------------------------
Panel:Groups Empty      0,     Valid data      595
             Smallest   7,     Largest           7
             Average group size in panel      7.00
Variances    Effects a(i)         Residuals e(i,t)
              1.068764                     .023102
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     WKS|     .00084         .00060     1.39  .1633     -.00034    .00201
   SOUTH|    -.00186         .03430     -.05  .9567     -.06909    .06536
    SMSA|    -.04247**       .01943    -2.19  .0288     -.08055   -.00439
      MS|    -.02973         .01898    -1.57  .1174     -.06693    .00748
     EXP|     .11321***      .00247    45.81  .0000      .10837    .11805
 EXP*EXP|    -.00042***   .5459D-04    -7.66  .0000     -.00053   -.00031
     OCC|    -.02148         .01378    -1.56  .1192     -.04849    .00554
     IND|     .01921         .01545     1.24  .2136     -.01106    .04948
   UNION|     .03278**       .01492     2.20  .0280      .00354    .06203
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
|              Test Statistics for the Regression Model                |
+----------------------------------------------------------------------+
|       Model              Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only    -2688.80597        886.90494      .00000 |
|(2)  Group effects only       27.58464        240.65119      .72866 |
|(3)  X - variables only    -1899.53716        607.12650      .31545 |
|(4)  X and group effects    2262.88731         82.26732      .90724 |
+----------------------------------------------------------------------+
|                       Hypothesis Tests                               |
|         Likelihood Ratio Test           F Tests                      |
|         Chi-squared   d.f.    Prob           F     num    denom  P value |
|(2) vs (1)   5432.78    594  .0000      16.14   594    3570   .00000 |
|(3) vs (1)   1578.54      9  .0000     212.75     9    4155   .00000 |
|(4) vs (1)   9903.39    603  .0000      57.76   603    3561   .00000 |
|(4) vs (2)   4470.61      9  .0000     761.75     9    3561   .00000 |
|(4) vs (3)   8324.85    594  .0000      38.25   594    3561   .00000 |
+----------------------------------------------------------------------+

-----------------------------------------------------------------------------
Two stage    least squares regression ............
LHS=DWI      Mean                    =        4.64877
             Standard deviation      =        1.03307
             Number of observs.      =           4165
Model size   Parameters              =              4
             Degrees of freedom      =           4161
Residuals    Sum of squares          =        3703.87
             Standard error of e     =         .94347
Fit          R-squared               =         .16573
             Adjusted R-squared      =         .16513
Not using OLS or no constant. Rsqrd & F may be < 0
Instrumental Variables:
WKS        SOUTH      SMSA       MS         ONE        FEM
BLK
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     DWI|  Coefficient       Error        z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2.86187***      .31650      9.04  .0000      2.24155   3.48219
     FEM|    -.12947***      .04759     -2.72  .0065      -.22274   -.03619
     BLK|    -.27853***      .06643     -4.19  .0000      -.40872   -.14833
      ED|     .14181***      .02447      5.80  .0000       .09385    .18977
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .023102
            Var[u]              =       .886838
            Corr[v(i,t),v(i,s)] =       .974611
            Sum of Squares              .000000
            R-squared                   .000000
Estimated using Hausman and Taylor IV estimator
Variance components provided by ;START = values

--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |X1 = time varying variables uncorrelated with u(i)
     WKS|     .00091         .00060     1.52  .1290     -.00026    .00208
   SOUTH|     .00716         .03255      .22  .8259     -.05663    .07095
    SMSA|    -.04176**       .01940    -2.15  .0314     -.07979   -.00373
      MS|    -.03636*        .01886    -1.93  .0539     -.07332    .00060
        |X2 = time varying variables assumed correlated with u(i)
     EXP|     .11297***      .00247    45.74  .0000      .10813    .11781
 EXP*EXP|    -.00042***   .5459D-04    -7.68  .0000     -.00053   -.00031
     OCC|    -.02139         .01378    -1.55  .1206     -.04841    .00562
     IND|     .01884         .01544     1.22  .2224     -.01143    .04911
   UNION|     .03035**       .01490     2.04  .0416      .00115    .05955
        |F1 = time invariant variables uncorrelated with u(i)
Constant|    2.88435***      .85189     3.39  .0007     1.21467   4.55402
     FEM|    -.13687         .12714    -1.08  .2817     -.38605    .11232
     BLK|    -.28182         .17643    -1.60  .1102     -.62762    .06397
        |F2 = time invariant variables assumed correlated with u(i)
      ED|     .14053**       .06580     2.14  .0327      .01156    .26950
--------+--------------------------------------------------------------------
```

E23.3 Arellano, Bond, and Bover's Estimator for Dynamic Panel Data Models

This estimator is for the dynamic random effects model

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \boldsymbol{\beta}_1'\mathbf{x}_{1it} + \boldsymbol{\beta}_2'\mathbf{x}_{2it} + \boldsymbol{\gamma}_1'\mathbf{f}_{1i} + \boldsymbol{\gamma}_2'\mathbf{f}_{2i} + \varepsilon_{it} + u_i \\ &= \alpha y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{f}_i + \varepsilon_{it} + u_i \\ &= \alpha y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{f}_i + v_{it} \\ &= \boldsymbol{\delta}'\mathbf{w}_{it} + v_{it}. \end{aligned}$$

where the terms in the equation are the same as in the Hausman and Taylor (HT) model. Subscripts '1' denote variables that are uncorrelated with u_i while subscripts '2' indicate variables that are correlated with u_i. This model differs by its inclusion of the lagged dependent variable. In principle, one could include additional lags, but our formulation includes only one. Note, also, that variables $\mathbf{f}_i$ are time invariant. If there is a constant term in the model, it is part of $\mathbf{f}_{1i}$.

E23.3.1 Technical Background

Instrumental variables estimation of the model without the lagged dependent variable is discussed in the previous section on the HT estimator. The Arellano et al. (ABB) estimator uses GMM instead. The HT approach is consistent in this setting, but ABB show that efficiency gains are available by using a larger set of moment conditions. In order to present the command structure for this estimator, it is necessary to lay out first some of the mathematical formulation of the ABB estimator. Let

$$\mathbf{w}_{it} = [y_{i,t\text{-}1},\ \mathbf{x}_{1it}',\ \mathbf{x}_{2it}',\ \mathbf{f}_{1i}',\ \mathbf{f}_{2i}']'$$

$$= (1 \ + KX_1 + KX_2 + KF_1 + KF_2) \times 1 = K{\times}1 \text{ vector}$$

and $$\mathbf{W}_i = \begin{bmatrix} \mathbf{w}_{i1}' \\ \mathbf{w}_{i2}' \\ \vdots \\ \mathbf{w}_{iT_i}' \end{bmatrix} = \text{ the full set of Rhs data for group } i \text{, and } \mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT_i} \end{bmatrix}$$

Note that $\mathbf{W}_i$ is, in principle, a $T_i{\times}K$ matrix. However, owing to the lagged dependent variable, only T_i-1 observations are available. To avoid a cumbersome, cluttered notation, we leave this distinction embedded in the notation for the moment. Later, when necessary, we will make it explicit. It will reappear in the formulation of the instrumental variables. A total of T_i-1 observations will be available for constructing the IV estimators. (Users with very short panels, i.e., two or three observations, are warned at this point that this estimator requires at least three periods of data to be useable, but four or more will be preferable.)

We now form the matrix of instrumental variables. Readers are referred to Hausman and Taylor (1981), Arellano et al. (1991, 1995, 1998), Ahn and Schmidt (1995) and Amemiya and MaCurdy (1995) for discussion of the various possibilities. We will form a matrix $\mathbf{Z}_i$ consisting of T_i-1 rows constructed the same way for T_i-2 observations and a final row that will be different, as discussed below. The matrix will be of the form

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i1}' & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{z}_{i2}' & \cdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{m}_i' \end{bmatrix}.$$

The instrumental variables contained in $\mathbf{z}_{it}'$ can include the following from within the model:

(Z Type 0) $\mathbf{x}_{it}$	(i.e., current values of the time varying variables)
(Z Type 1) $\mathbf{x}_{it}$ and $\mathbf{x}_{i,t\text{-}1}$	(i.e., current and one lag of the time varying variables)
(Z Type 2) $\mathbf{x}_{i1},...,\mathbf{x}_{iTi}$	(i.e., all current, past and future values of the time varying variables)
(Z Type 3) $\mathbf{x}_{i1},...\mathbf{x}_{i,t}$	(i.e., all current and past values of the time varying variables)

The time invariant variables that are uncorrelated with u_i, that is $\mathbf{f}_{1i}$, are always appended at the end of the nonzero part of each row. We should note, it may seem that including $\mathbf{x}_2$ in the instruments would be invalid. However, we will be using deviations from group means. While the original variables are correlated with u_i, by construction, the group mean deviations are not. The issue of correlation between the transformed lagged y_{it} and the deviations of ε_{it} is discussed in the papers cited.

The final row of $\mathbf{Z}_i$ is important to the construction. Two possibilities are provided:

(M Type 1) $\mathbf{f}_{1i}'$ and $\overline{\mathbf{x}}_{1i}$ (produces the Hausman and Taylor estimator)

(M Type 2) $\mathbf{f}_{1i}'$ and $\mathbf{x}_{1i,1}', \mathbf{x}_{1i,2}',..., \mathrm{x}_{1i,Ti}$ (Amemiya and MaCurdy).

Note that the **m** variables are exogenous time invariant variables, $\mathbf{f}_{1i}$, and the exogenous time varying variables, either condensed into the single group mean or in the raw form, with the full set of T_i observations.

As Ahn and Schmidt show, there are potentially huge numbers of additional orthogonality conditions in this model owing to the relationship between first differences and second moments. We do not consider those. As it stands, the number of instrumental variables contained in $\mathbf{Z}_i$ is potentially enormous even in a moderately sized model. The matrix $\mathbf{Z}_i$ could be huge. Consider the Z Type 3 and M Type 2 forms in a model with 10 time varying right hand side variables and suppose T_i is 15. Then, there are 15 rows and roughly 15×(10×15) or 2,250 columns. (This makes one wonder about the practicality of the Ahn and Schmidt estimator which involves potentially thousands of instruments in a model containing only a handful of parameters. The order of the computation grows with the square of T_i.)

To construct the estimator, we will require a transformation matrix, $\mathbf{H}_i$ constructed as follows. Let **i** denote a $T_i \times 1$ column of ones. Then,

$$\mathbf{M}_i = \mathbf{I} - (1/T_i)\mathbf{i}\,\mathbf{i}'$$

This is the matrix that creates deviations from means. Let $\mathbf{M}_{i1}$ denote the first T_i-1 rows of this matrix. Then, finally,

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{M}_{i1} \\ \dfrac{1}{T_i}\mathbf{i}' \end{bmatrix}$$

Thus, $\mathbf{H}_i$ replaces the last row of $\mathbf{M}_i$ with a row with all elements equal to $1/T_i$. The effect is as follows: if $\mathbf{q}_i$ is T_i observations on a variable, then $\mathbf{H}_i\mathbf{q}_i$ produces $\mathbf{q}_i$* in which the first T_i-1 observations are converted to deviations from group means and the last observation is the group mean.

Finally, for purposes of GMM estimation, consider three candidates for the covariance matrix of $\mathbf{v}_i = (v_{i1},v_{i2},...,v_{iTi})$

($\mathbf{\Omega}$ Type 1) $\mathbf{\Omega} = \sigma_\varepsilon^2\mathbf{I}$ (uncorrelated, misspecified by construction)

($\mathbf{\Omega}$ Type 2) $\mathbf{\Omega} = \sigma_\varepsilon^2\,\mathbf{I} + \sigma_u^2\mathbf{i}\,\mathbf{i}'$ (variance components, random effects)

($\mathbf{\Omega}$ Type 3) $\mathbf{\Omega} = E[\mathbf{v}_i\mathbf{v}_i']$ (robust, unrestricted, $E[\boldsymbol{\varepsilon\varepsilon}'] = \mathbf{\Sigma} \neq \sigma^2\,\mathbf{I}$)

We leave aside for the moment the problem of computing an estimator of $\mathbf{\Omega}$. The ABB estimator of δ is a two step GMM estimator in which the two steps are defined by which form of $\mathbf{\Omega}$ is used. In the first step, the consistent, but inefficient estimator based on $\mathbf{\Omega}$ Type 1 is used to obtain an estimator of δ that enables estimation of the appropriate $\mathbf{\Omega}$. At the second step, the more efficient estimator based on $\mathbf{\Omega}$ Type 2 or 3 is used. (Note, again, this is not going to be 'fully' efficient because there remain moment conditions based on first differences and higher moments that are not being used - see Ahn and Schmidt (1995).) The ABB estimator is

$$\hat{\delta} = \left[\left(\sum_{i=1}^{N} (\mathbf{W}_i'\mathbf{H}_i')\mathbf{Z}_i\right)\left(\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{H}_i\mathbf{\Omega}_i\mathbf{H}_i')\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{H}_i\mathbf{W}_i)\right)\right]^{-1} \times$$
$$\left(\sum_{i=1}^{N} (\mathbf{W}_i'\mathbf{H}_i')\mathbf{Z}_i\right)\left(\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{H}_i\mathbf{\Omega}_i\mathbf{H}_i')\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{H}_i\mathbf{y}_i)\right)$$

The estimator of the asymptotic covariance matrix for this estimator is the inverse matrix in square brackets. The two step estimator is computed as follows:

Step 1. Computation of $\hat{\delta}$ based on $\mathbf{\Omega} = \sigma_\varepsilon^2\mathbf{I}$. (Note, σ_ε^2 is not necessary. It falls out of the matrix product. After Step 1, a set of residuals $\hat{\mathbf{v}}_i$ is computed. Two estimators of $\mathbf{\Omega}$ are now available. You may provide specific values for σ_ε^2 and σ_u^2. If so, then Type 2 is computed. (In this case, the first step is actually superfluous, but results will be reported nonetheless.) Otherwise,

$$\hat{\mathbf{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{H}_i\,\hat{\mathbf{v}}_i\,\hat{\mathbf{v}}_i\,\mathbf{H}_i'$$

Step 2. Recompute the estimator using the new estimator of $\mathbf{\Omega}$.

E23.3.2 Command

The command for this procedure is as follows: Note that the entire estimator is self contained, unlike the HT estimator of the previous section. Use

SETPANEL	**; … $**
REGRESS	**; Lhs = dependent variable**
	; Rhs = variables in x_1 then
	variables in x_2 then
	variables in f_1 (including one if needed) then
	variables in f_2
	; Start = kx_1, kx_2, kf_1, kf_2
	; Panel
	; DPD (for dynamic panel data) $

If your model contains a constant term, include it in the list for $\mathbf{f}_1$. This procedure allows unbalanced panels, but it is quite intricate to do so. The panel data specification,

; Pds = count variable

assumes each group starts with the first period. If this is not the case, use

; Date = date variable, first - last

for example,

; Date = year, 1980 - 1987

to specify starting and ending points in the sample. All dates must fall in that interval. This allows you to change starting dates for specific observation groups with balanced or unbalanced panels. Your date variable, *year* in the example above, will provide the dates for the specific observation. We would note, in spite of the additional flexibility this allows, you *should not* use this estimator with panels that have gaps in them. The lagged values of the variables cannot be placed properly if the data set contains gaps. Results will be unpredictable, and almost certainly unusable.

Some of the four sets of variables may not be present in your model. You will indicate this by placing zeros in the **; Start** list. Thus, for example, if you have no exogenous time invariant variables, you will set KF_1 to zero. (We note, like many of these, the setup is much more complicated than the command it requires.)

NOTE: The model is assumed to contain a lagged dependent variable. Do not include this on your Rhs. It will be added to your model. If your model does not contain a lagged dependent variable, you should be using the HT estimator described in the previous section.

With the command as given, the default settings for the different arrangements are

Instruments: *Z* Type 1, present and a single lagged value
M Type 1, Hausman and Taylor, group means, not the original data

Covariances: $\mathbf{\Omega}$ Type 3, computed after the first step

You can change the first of these by using

; Pattern = Z Type, M Type if desired

where the specifications are in the following table:

		M Type = last row of $\mathbf{Z}_i$	
Z Type = instruments		1 = means	2 = all data
0	$\mathbf{x}_t$, $\mathbf{f}_1$	**0**	**0, A**
1	$\mathbf{x}_t$, $\mathbf{x}_{t-1}$, $\mathbf{f}_1$	**1**	**1, A**
2	$\mathbf{x}_t$, $\mathbf{x}_{t-1}$,...,$\mathbf{x}_1$, $\mathbf{f}_1$	**P**	**P, A**
3	$\mathbf{x}_T$, $\mathbf{x}_{T-1}$,...,$\mathbf{x}_1$, $\mathbf{f}_1$	**A**	**A, A**

For example, to use all previous values and the Amemiya/MaCurdy form of the last row of Z_i, you would use

; Pattern = P, A

Note the following for this estimator:

- *Z* Type 3 may only be used with a balanced panel with a fixed starting date.

- This estimator has no handler for missing data. You must set up the sample without missing values before invoking the estimator.

- You may use **; Model = kx1,kx2,kf1,kf2** instead of **; Start = kx1,kx2,kf1,kf2**. This change in format is just for convenience. Both forms are retained to maintain compatibility with earlier versions.

- Time dummy variables can be included in the model if desired – they must be created separately. However, they are likely to be problematic when constructing the instruments. They can proliferate.

This estimator generates huge numbers of moment conditions, particularly if you use the '*A*' pattern – that is, ALL moment conditions in the sample. For example, suppose that $T_i = 10$, $KX_1 = 3$, $KX_2 = 1$, $KF_1 = 2$ (including *one*) and $KF_2 = 1$, which does not seem like a large model. Then, the number of moment conditions (columns in $\mathbf{Z}_i$) is

KZ = 8 periods of lagged data after the one dropped initially ×
(8 groups of available exogenous variables × 4 variables in $\mathbf{x}_1$ and $\mathbf{x}_2$ + 2 variables in $\mathbf{f}_1$) +
the number of variables in the last row, which might be $KX_1+KF_1 = 5$ for *M* Type 1
= 277 columns

This does not seem like a large model, but it is. The problem is that the number of columns proliferates with the square of the number of periods. The Ahn and Schmidt (1995) approach would add roughly 10(10(9)/2 + 8) = 530 additional columns to $\mathbf{Z}_i$ for a total approaching 900, for the purpose of estimating eight parameters! A matrix with 900 columns might, in itself, present no obvious obstacle. However, note that the center matrix in the computation of $\hat{\delta}$ is square with this number of rows, so in order to compute the Ahn and Schmidt estimator, one would have to manipulate a 900×900 inverse matrix. Ahn and Schmidt do not mention this practical burden.

Arellano and Bond (1998) mention briefly the possibility that $\mathbf{Z}_i$ may have many columns, and one might have to drop some of them (some 'less informative instruments'). We begin this paring by dropping the Ahn and Schmidt estimator. Then, *Z* Type 1 is used to produce a fairly small set of instruments. Given that the model is already vastly overidentified, it seems that Type 1, a single lagged value of the exogenous variables, ought to be sufficient. For the model size considered here, this would produce an order of 9(5+5) + 5 or 95 columns in $\mathbf{Z}_i$. This is still very large, but manageable. The end result of this discussion is that this estimator is practical only when T_i is relatively small, and best with small T_i and quite large *N*. We recommend using only the default pattern, '1,' that is, a single lagged value.

Finally, if you wish to specify the **Ω** Type 2, GLS estimator, then you must provide values for σ_u^2 and σ_ε^2, in that order, after KF_2 in the **; Start** specification (which will now contain six values instead of four.) You might use a prior GLS estimator to estimate the variance components in your model. The resulting command for the DPD estimator would be

REGRESS ; Lhs = dependent variable
; Rhs = all independent variables
; Pds = the value
; Panel
; DPD
; Start = kx1,kx2,kf1,kf2,s2u,s2e $

There are no other options for this model other than those already mentioned. After estimation, you may use

; List to display fitted values
; Keep = name to retain fitted values
; Res = name to retain residuals

all as usual for regression estimators.

E23.3.3 A Test Statistic for the Specification

Bhargava and Sargan (1983) propose a test statistic for the specification of the model. This is computed as part of the results for the estimator. See the results below. The computation is

$$\chi^2[KZ - Kmodel] = \left(\sum_{i=1}^{N} (\hat{\mathbf{v}}_i'\mathbf{H}_i')\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{H}_i\mathbf{\Omega}_i\mathbf{H}')_i\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{H}_i\hat{\mathbf{v}}_i)\right)$$

The statistic has a limiting chi squared distribution with degrees of freedom equal to the number of overidentifying restrictions (moment conditions). As noted earlier, in these models, there are typically few parameters and very many moments, so that the degrees of freedom is likely to be quite large.

E23.3.4 Technical Notes

The earlier result lays out the computation of the estimator and the asymptotic covariance matrix. A detail to be added concerns the precise form of the matrices of instrumental variables. Your data for this estimator are assumed to consist of groups of T_i observations,

$$Data_{it} = y_{it},\ \mathbf{x}_{1it},\ \mathbf{x}_{2it}, \mathbf{f}_{1i}, \mathbf{f}_{2i}, t = 1,...,T_i.$$

That is, each data group contains T_i rows. We assume that the first row contains known initial values for all variables, but since there is no y_{i0} in the data set, the model applies to observations two through T_i. The general form of the model is

$$y_{it} = \alpha y_{i,t-1} + \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}_{1it} + \boldsymbol{\beta}_2'\mathbf{x}_{2it} + \boldsymbol{\gamma}_1'\mathbf{f}_{1i} + \boldsymbol{\gamma}_2'\mathbf{f}_{2i} + \varepsilon_{it} + u_i, t = 2,...,T_i.$$

The matrices of instruments are assembled as follows: (Note that any of $\mathbf{x}_{2it}$, $\mathbf{f}_{1i}$ and $\mathbf{f}_{2i}$ may not actually be present in your model. First, define the vector placed in the last row.

$$\mathbf{m}_i' = [\,\bar{\mathbf{x}}_{i1}{}'\ \ \mathbf{f}_{1i}'\,] \qquad \text{for } M \text{ Type 1} \qquad (KM = KF_1 + KX_1)$$

or
$$\mathbf{m}_i' = [\mathbf{x}_{1i2}, \ldots, \mathbf{x}_{1iT(i)}\ \mathbf{f}_{1i}'] \qquad \text{for } M \text{ Type 2} \qquad (KM = KF_1 + (T_i - 1)KX_1)$$

Type 1 uses the group mean, including only the observations used in the computation of the coefficients (i.e., dropping the first observation in each group) – this is $(1 \times KX_1{+}KF_1)$. Type 2 uses the original data that would be used to compute the group mean, so this vector has $(T_i\text{-}1) \times KX_1{+}KF_1$ elements.

The matrices of instrumental variables contain T_i - 1 rows, once again, assuming that the first row of data contains only initial values. Rows 1 to T_i - 2 are built up from the data set, while the last row contains only $\mathbf{m}_i'$ and zeros. The particular forms of these matrices are as follows, where we illustrate with a model in which T_i = 4. This will be sufficiently general to extend to other cases. With T_i = 4, there are three observations to use for estimation in each group, as one is lost for the lagged dependent variable. In the following, let $\mathbf{x}_{it}'$ denote $[\mathbf{x}_{1it}{}', \mathbf{x}_{2it}{}']$, the full set of time varying right hand side variables. Then,

Z Type 0, current data only:

$$\begin{bmatrix} \mathbf{x}_{i1}' & \mathbf{f}_{1i}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{x}_{i2}' & \mathbf{f}_{1i}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_i' \end{bmatrix}$$

This matrix has number of columns equal to $(T_i\text{-}2)\times[(KX_1{+}KX_2) + KF_1] + KM$.

Z Type 1, current and one period lagged data:

$$\begin{bmatrix} \mathbf{x}_{i1}' & \mathbf{x}_{i2}' & \mathbf{f}_{1i}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{x}_{i2}' & \mathbf{x}_{i3}' & \mathbf{f}_{1i}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_i' \end{bmatrix}$$

This matrix has number of columns equal to $(T_i\text{-}2)\times[2(KX_1{+}KX_2) + KF_1] + KM$.

Z Type 2, current past and future data:

$$\begin{bmatrix} \mathbf{x}_{i1}' & \mathbf{x}_{i2}' & \mathbf{x}_{i3}' & \mathbf{x}_{i4}' & \mathbf{f}_{1i}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{x}_{i1}' & \mathbf{x}_{i2}' & \mathbf{x}_{i3}' & \mathbf{x}_{i4}' & \mathbf{f}_{1i}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_i' \end{bmatrix}$$

This matrix has number of columns equal to $(T_i\text{-}2)\times[T_i(KX_1{+}KX_2) + KF_1] + KM$.

Z Type 3, current and lagged data:

$$\begin{bmatrix} \mathbf{x}'_{i1} & \mathbf{x}'_{i2} & \mathbf{f}'_{1i} & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{x}'_{i1} & \mathbf{x}'_{i2} & \mathbf{x}'_{i3} & \mathbf{f}'_{1i} & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}'_i \end{bmatrix}$$

This matrix has number of columns equal to $[T_i \times (T_i\text{-}1)/2\ -\ 1](KX_1+KX_2) + (T_i\text{-}2)KF_1 + KM$.

The matrix $\mathbf{W}_i$ consists of T_i - 1 rows of the right hand side data for the observations in group i.

E23.3.5 An Application

To illustrate the Arellano, Bond and Bover estimator, we extend the Cornwell and Rupert log wage equation examined in the previous section. The model is now

$$\begin{aligned} \log wage_{it} = \; & \alpha \log wage_{i,t-1} \\ & + \beta_{1,1} wks_{it} + \beta_{1,2} south_{it} + \beta_{1,3} smsa_{it} + \beta_{1,4} ms_{it} \\ & + \beta_{2,1} exp_{it} + \beta_{2,2} exp_{it}^2 + \beta_{2,3} occ_{it} + \beta_{2,4} ind_{it} + \beta_{2,5} union_{it} \\ & + \gamma_{1,1} + \gamma_{1,2} fem_i + \gamma_{1,4} blk_i \\ & + \gamma_{2,1} ed_i + \varepsilon_{it} + u_i \end{aligned}$$

The commands used to specify the model are

```
CREATE      ; i = Trn(7,0) $
SETPANEL    ; Group = i ; Pds = ti $
NAMELIST    ; x1= wks,south,smsa,ms $
NAMELIST    ; x2= exp,exp*exp,occ,ind,union $
NAMELIST    ; f1 = one,fem,blk $
NAMELIST    ; f2 = ed $
NAMELIST    ; x = x1,x2 $
NAMELIST    ; f  = f1,f2 $
NAMELIST    ; exog = x1,f1 $
NAMELIST    ; all = x1,x2,f1,f2 $
CALC        ; kx1 = Col(x1) ; kx2 = Col(x2)
            ; kf1 = Col(f1) ; kf2 = Col(f2) $
REGRESS     ; Lhs = lwage
            ; Rhs = all
            ; Start = kx1,kx2,kf1,kf2
            ; Panel
            ; DPD
            ; List $
```

Results are given for the first step IV and the second step GMM estimators. The **; List** request generates an extremely long list. We show only a few lines of this. One would normally not use this option with a large sample such as this one. Finally, the initial, one step instrumental variable estimates are shown last. They would appear at the beginning of the displayed results when the program is used.

```
-----------------------------------------------------------------------------
Arellano/Bond/Bover IV Estimator for Dynamic
Panel Data Models
2nd step GMM/IV with robust VC
Pattern requested for instrumental variables is:
[f(1)] + current and one lag of [x(1),x(2)]
Hausman and Taylor form for last row of Z matrix
Bhargava/Sargan Spec. Test:  3515.00305
Degrees of freedom =  98, Prob = .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Exogenous variables uncorrelated with u(i)
     WKS|      .00908***      .00040    22.97  .0000      .00831    .00986
   SOUTH|      .11813***      .00902    13.09  .0000      .10045    .13582
    SMSA|     -.13325***      .01438    -9.27  .0000     -.16144   -.10506
      MS|     -.19757***      .01085   -18.22  .0000     -.21882   -.17631
        |Exogenous variables may be correlated with u(i)
     EXP|      .10674***      .00190    56.11  .0000      .10301    .11047
 EXP*EXP|         0.0    .....(Fixed Parameter).....
     OCC|      .05345***      .01072     4.99  .0000      .03245    .07445
     IND|     -.17999***      .01253   -14.36  .0000     -.20455   -.15543
   UNION|      .13734***      .01067    12.87  .0000      .11643    .15825
        |Time invariant variables uncorrelated with u(i)
Constant|     8.44355***      .16780    50.32  .0000     8.11466   8.77244
     FEM|    -1.19378    .....(Fixed Parameter).....
     BLK|     -.09318***      .02595    -3.59  .0003     -.14403   -.04233
        |Time invariant variables correlated with u(i)
      ED|      .68292***      .00869    78.57  .0000      .66589    .69996
        |Lagged value of the dependent variable
LWAGElag|    -1.88041    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Predicted values and residuals for a fitted dynamic panel model
===============================================================
Individual  Period         Actual     Prediction        Residual
         1       1       5.560680 [No data, T=0] [No data, T=0]
         1       2       5.720310       4.634063        1.086247
         1       3       5.996450       4.392103        1.604347
         1       4       5.996450       4.310541        1.685909
         1       5       6.061460       4.056660        2.004800
         1       6       6.173790       3.816224        2.357566
         1       7       6.244170       3.751004        2.493166
         2       1       6.163310 [No data, T=0] [No data, T=0]
         2       2       6.214610       8.671972       -2.457362
         2       3       6.263400       8.469102       -2.205702
         2       4       6.543910       8.232034       -1.688124
         2       5       6.697030       8.130897       -1.433867
         2       6       6.791220       8.017610       -1.226390
         2       7       6.815640       7.916062       -1.100422
```

```
        3       1       5.652490 [No data, T=0] [No data, T=0]
        3       2       6.436150       6.968390       -.532240
        3       3       6.548220       6.744594       -.196374
        3       4       6.602590       6.690278       -.087688
        3       5       6.695800       6.589141        .106659
        3       6       6.778780       6.609846        .168934
        3       7       6.860660       6.517380        .343280
```

These are the initial, one step instrumental variable estimates of the dynamic model.

```
-----------------------------------------------------------------------------
Arellano/Bond/Bover IV Estimator for Dynamic
Panel Data Models
One step GMM/IV estimator
Pattern requested for instrumental variables is:
[f(1)] + current and one lag of [x(1),x(2)]
Hausman and Taylor form for last row of Z matrix
Bhargava/Sargan Spec. Test:     5.08942
Degrees of freedom =  98, Prob =1.00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
   LWAGE|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Exogenous variables uncorrelated with u(i)
     WKS|     .00122         .00101     1.21  .2268     -.00076    .00320
   SOUTH|    -.00390         .01981     -.20  .8442     -.04273    .03494
    SMSA|     .01883         .04095      .46  .6457     -.06143    .09908
      MS|    -.04553*        .02653    -1.72  .0861     -.09753    .00646
        |Exogenous variables may be correlated with u(i)
     EXP|     .00849*        .00456     1.86  .0630     -.00046    .01743
 EXP*EXP|        0.0    .....(Fixed Parameter).....
     OCC|    -.04102         .02684    -1.53  .1264     -.09362    .01158
     IND|     .04169         .03145     1.33  .1850     -.01995    .10333
   UNION|    -.00958         .02745     -.35  .7272     -.06338    .04423
        |Time invariant variables uncorrelated with u(i)
Constant|     .91945*        .52122     1.76  .0777     -.10213   1.94102
     FEM|    -.08323***      .02983    -2.79  .0053     -.14169   -.02477
     BLK|    -.06278         .06108    -1.03  .3040     -.18249    .05693
        |Time invariant variables correlated with u(i)
      ED|     .00472         .03680      .13  .8979     -.06740    .07684
        |Lagged value of the dependent variable
LWAGElag|     .83916***      .04756    17.65  .0000      .74595    .93237
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

E24: Linear Systems of Regression Equations – SURE and 3SLS

E24.1 Introduction

This chapter and Chapter E25 present methods of estimating the parameters of the regression system

$$y_1 = f_1(\mathbf{x}_1,\boldsymbol{\beta}) + \varepsilon_1$$

$$y_2 = f_2(\mathbf{x}_2,\boldsymbol{\beta}) + \varepsilon_2$$

$$\ldots$$

$$y_M = f_M(\mathbf{x}_M,\boldsymbol{\beta}) + \varepsilon_M$$

or

$$\mathbf{y} = \mathbf{f}(\mathbf{X},\boldsymbol{\beta}) + \boldsymbol{\varepsilon},.$$

We assume $E[\boldsymbol{\varepsilon}|\text{all } \mathbf{x}] = \mathbf{0}$

and $E[\boldsymbol{\varepsilon\varepsilon}'|\text{all } \mathbf{x}] = \boldsymbol{\Sigma}.$

As stated, the model is a possibly nonlinear system of seemingly unrelated regressions. However, for some settings (e.g., the linear model of Section E24.4), the **x** vectors on the right hand sides of the equations may include endogenous variables, y_j, from other equations. That is, we also accommodate systems of simultaneous equations. The linear models also allow autocorrelation of the disturbances. Systems of nonlinear equations are shown in Chapter E25.

E24.2 Linear SURE Models Estimated by GLS

The seemingly unrelated linear regression equations (**SURE**) model is:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,$$

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2,$$

$$\ldots$$

$$\mathbf{y}_M = \mathbf{X}_M\boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M.$$

$$E[\boldsymbol{\varepsilon}_i|\mathbf{X}_1,\ldots] = \mathbf{0},$$

$$E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'|\mathbf{X}_1,\ldots] = \sigma_{ij}\mathbf{I}.$$

There are M, up to 20, equations. There are n observations in total. There must be the same number of observations for all equations. The disturbances across equations are allowed to be freely correlated. The parameter vector obtained by stacking $\boldsymbol{\beta}_m$ may have up to 150 parameters.

Collect the M disturbances for a particular observation in a column vector

$$\boldsymbol{\varepsilon}_t = [\varepsilon_{t1},\varepsilon_{t2}, \ldots \varepsilon_{tM}]'.$$

The model specifies $E[\boldsymbol{\varepsilon}_t|\mathbf{X}_1,\ldots] = \mathbf{0}$, $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' |\mathbf{X}_1,\ldots] = \boldsymbol{\Sigma}.$

The estimator also allows for first order autocorrelation,

$$\varepsilon_{i,t} = \rho_i \varepsilon_{i,t\text{-}1} + u_{i,t}.$$

There are two estimators for this model in *LIMDEP*. The two step (or iterative) feasible GLS procedure (FGLS) uses Zellner's technique. The second is maximum likelihood, which is suitable for constrained, singular systems, such as translog demand systems. This is presented in Section E24.3.

E24.2.1 Command for SURE Estimation

The command for the GLS estimator for a system of regression equations is:

SURE **; Lhs = y1, y2, ..., ym** (your list of Lhs variables)
; Eq1 = list of Rhs variables for first equation
; Eq2 = list of Rhs variables for second equation
; ...
; EqM = list of Rhs variables for Mth equation $

HINT: This application is a convenient one for namelists. Namelists will be particularly useful if all of the equations share a set of variables. See Chapter R6 for details.

HINT: If all equations have the same set of Rhs variables *and if there are no linear constraints imposed*, then SURE is the same as equation by equation ordinary least squares. If linear constraints are imposed, this is no longer true.

E24.2.2 Options for the Generalized Least Squares Procedure

If no further specifications are given on the command, the procedure is allowed to iterate to convergence. This is a globally concave problem for which convergence is guaranteed unless the data are very badly conditioned. You can restrict the number of iterations with

; Maxit = maximum iterations

To use Zellner's efficient two step estimator for the system, that is, using the OLS residuals to estimate $\mathbf{\Sigma}$, use

; Maxit = 0

To obtain equation by equation OLS estimates, use

; Maxit = 99

Linear constraints may be imposed on the coefficients in the same way as described for the single equation, linear regression model in Chapter E8. The parameters of the equations are stacked as $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', ..., \boldsymbol{\beta}_M']'$, then the constraints are imposed as if this were a single equation model. For example, the following is a part of the model estimated in our example below:

SURE **; Lhs = igm,ic**
; Eq1 = one,fgm,cgm
; Eq2 = one,fc,cc
; CLS: b(4) - b(1) = 0, b(5) - b(2) = 0, b(6) - b(3) = 0 $

The linear restriction imposes cross equation equality of the corresponding pairs of coefficients. (With this restriction imposed, the model is the TSCS model presented in Chapter E15.) (The new specification of restrictions and hypothesis tests shown in Chapter E8, using the variable names, is not useable here because the program has no way to know which equation to use if a variable appears in more than one equation.)

With restrictions imposed, you will see two full sets of output. The initial set, without the restrictions imposed, is presented in full. Then, the restricted least squares estimates are presented.

TECHNICAL NOTE: The restricted GLS estimator is *not* the maximum likelihood estimator, even if it is allowed to iterate to convergence. The RGLS estimator is computed using the restricted least squares formula, *after* the unrestricted estimates are obtained. Therefore, the RGLS estimator is a function of the unrestricted estimator, not an iterative estimator in its own right.

Autocorrelation may be of two forms:

Model 1: $\varepsilon_{i,t} = \rho_i \varepsilon_{i,t\text{-}1} + u_{i,t}$ (equation specific autocorrelation)

Model 2: $\varepsilon_{i,t} = \rho\varepsilon_{i,t\text{-}1} + u_{i,t}$ (common autocorrelation).

Model 2 differs in that the same correlation coefficient is used for all equations. For these estimators, simply add

; Model = 1

or **; Model = 2**

to the command. The autocorrelation coefficients are estimated by using $r_i = 1 - DW_i/2$, where DW_i is the Durbin-Watson statistic computed using the single equation, equation by equation ordinary least squares residuals. If you specify **; Model = 2**, the common estimate is the simple (unweighted) average of the M individual estimates.

The estimated **Σ** is a 'weighting' matrix which greatly influences the final results. You can specify your own, rather than allowing *LIMDEP* to use the second step OLS residuals to form one. To specify a particular **Σ** matrix, use

; Sigma = name of matrix
; Maxit = 1

The setting of the maximum iterations at 1 is needed to prevent *LIMDEP* from recomputing **Σ** at another iteration. If necessary, you may also weight observations with

; Wts = weighting variable

The estimated disturbance covariance matrix, **S** is not displayed in the final output unless you request it with

; List

Note that this is not the same as a request for a listing of the fitted values.

Because this is a multiple equation estimator, it does not produce a set of fitted values or residuals. But, these are simple to obtain just by extracting the coefficients from the saved results and using **CREATE** with each parameter vector to create the linear function of the variables.

E24.2.3 Model Output for the GLS Estimator

Initial results include a trace of the log likelihood function for the iterations. Then, for each equation, the usual sorts of regression results, including fit measures, coefficient estimates, etc. are given. If the model is fit with a correction for autocorrelation, the diagnostic statistics include the autocorrelation coefficient estimated using the OLS residuals and the Durbin-Watson statistic and estimated autocorrelation for the corrected GLS residuals. That is, the values displayed are for

$$\hat{u}_{i,t} = e_{i,t} - r_i e_{i,t-1}.$$

The other saved results for this estimator are:

Matrices:	*b*	=	stacked coefficient vector,
	varb	=	estimated asymptotic covariance matrix for **B**,
	sigma	=	**S**, the final sample estimate of $\boldsymbol{\Sigma}$.
Scalars:	*logl*	=	log likelihood
		=	$-(MT/2)[1 + \log 2\pi + \log \det(\mathbf{S})]$,
	traceofs	=	trace(**S**).

In addition, this model creates a coefficient matrix named *b_sure* that is built up from the separate coefficient vectors, into a matrix whose each column corresponds to an equation.

Consider a small example using the Grunfeld data:

$$I_{it} = \alpha_1 + \alpha_2 C_{it} + \varepsilon_{1t},$$

$$F_{it} = \beta_1 + \beta_2 Year_t + \varepsilon_{2t}.$$

```
SAMPLE      ; 1-200 $
SURE        ; Lhs = i,f
            ; Eq1 = one,c
            ; Eq2 = one,year $
```

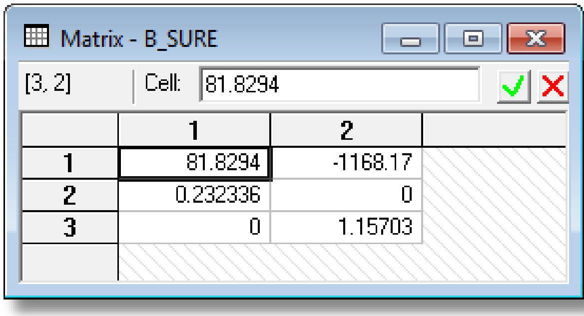
The estimates for the two equations using the 200 observations will be

```
Criterion function for GLS is log-likelihood.
Iteration    0, GLS           =   -2946.243
Iteration    1, GLS           =   -2913.801
Iteration    2, GLS           =   -2911.505
Iteration    3, GLS           =   -2911.333
Iteration    4, GLS           =   -2911.320
Iteration    5, GLS           =   -2911.319
Iteration    6, GLS           =   -2911.319
Iteration    7, GLS           =   -2911.319
Iteration    8, GLS           =   -2911.319
GLS             has converged.
```

```
-----------------------------------------------------------------------------
Estimates for equation: I.........................
Generalized  least squares regression ............
LHS=I         Mean                    =     145.95825
              Standard deviation      =     216.87530
              Number of observs.      =           200
Model size    Parameters              =             2
              Degrees of freedom      =           198
Residuals     Sum of squares          =    .626965E+07
              Standard error of e     =     177.94630
Fit           R-squared               =        .32340
              Adjusted R-squared      =        .31998
Model test    F[  1,    198] (prob) =    94.6(.0000)
Diagnostic    Log likelihood          =   -1319.07904
              Restricted(b=0)         =   -1359.15096
              Chi-sq [  1]  (prob) =  80.1(  .0000)
Info criter. Akaike Info. Criter. =      10.37291
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      23.4374 Log-Likelihood  =   -2911.3194
Durbin-Watson  .202 Autocorrelation =        .8989
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    81.8294***    14.28533     5.73  .0000     53.8307  109.8282
       C|     .23234***      .02451     9.48  .0000      .18431    .28037
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Estimates for equation: F.........................
Generalized  least squares regression ............
LHS=F         Mean                    =    1081.68110
              Standard deviation      =    1314.46969
              Number of observs.      =           200
Model size    Parameters              =             2
              Degrees of freedom      =           198
Residuals     Sum of squares          =    .340175E+09
              Standard error of e     =    1310.74576
Fit           R-squared               =        .00066
              Adjusted R-squared      =       -.00439
Model test    F[  1,    198] (prob) =      .1(.7177)
Diagnostic    Log likelihood          =   -1718.45298
              Restricted(b=0)         =   -1719.52417
              Chi-sq [  1]  (prob) =   2.1(  .1433)
Info criter. Akaike Info. Criter. =      14.36665
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      23.4374 Log-Likelihood  =   -2911.3194
Durbin-Watson  .155 Autocorrelation =        .9227
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       F|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -1168.17       18282.87     -.06  .9491   -37001.94  34665.60
    YEAR|    1.15703        9.40223      .12  .9021   -17.27100  19.58507
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The additional matrix created, *b_sure*, will be as follows:



	1	2
1	81.8294	-1168.17
2	0.232336	0
3	0	1.15703

Figure E24.1 Coefficient Matrix

Note that although it is clear from the specification of the model, the matrix, itself, contains no internal information to identify which variable corresponds to each row. The set of variables is constructed by moving through the equation lists, in order, and assembling the union of all of the sets of names in the order in which they occur in the stacked list.

The labels for the *Last Model* estimated are constructed from the equation specification. Each name is constructed as the first three characters from the name of the Lhs variable, then an underscore, then the first four characters of the Rhs variable. In the example,

SURE ; Lhs = igm,ic ; Eq1 = one,fgm,cgm ; Eq2 = one,fc,cc $

the set of labels would be [**igm_one, igm_fgm, igm_cgm, ic_one, ic_fc, ic_cc**]. (See Section R9.7.3 for use of the *Last Model*.)

There are no residuals or fitted values produced internally. But, you can retrieve these from the other results. In most cases, this will involve some setup that is specific to the model at hand.

E24.2.4 The Translog System

The translog function and some related models are estimated in various forms in the setting of multivariate regressions. Normally, the restrictions in the model are cross equation equality restrictions not usually viewed as testable, but as part of the model. Christensen and Greene's (1976) homothetic translog cost function provides an example. The model is:

$$
\begin{aligned}
\log(c/p_f) &= \alpha + \beta\log y + \gamma(\log y)^2/2 + \delta_k\log(p_k/p_f) + \delta_l\log(p_l/p_f)\\
&\quad + \theta_{kk}\log(p_k/p_f)^2/2 + \theta_{ll}\log(p_l/p_f)^2/2 + \theta_{kl}\log(p_k/p_f)\log(p_l/p_f) + \varepsilon_c,\\
s_k &= \delta_k + \theta_{kk}\log(p_k/p_f) + \theta_{kl}\log(p_l/p_f) + \varepsilon_k,\\
s_l &= \delta_l + \theta_{kl}\log(p_k/p_f) + \theta_{ll}\log(p_l/p_f) + \varepsilon_l,
\end{aligned}
$$

where c = cost,
y = output,
s_k = cost share of capital,
s_l = cost share of labor,
p_k, p_l, and p_f are the unit prices for capital, labor, and fuel, respectively.

The restriction of linear homogeneity in the input prices is imposed by normalizing cost and the other prices by the price of fuel. Using the obvious mnemonics for the variables in the model, and assuming that all variables are in log form, the commands for estimating a model such as this would be:

```
SURE     ; Lhs = cost,sk,sl
         ; Eq1 = one,y,y2,lpkf,lplf,lpkf2,lpkl2,lpkfplf
         ; Eq2 = one,lpkf,lplf
         ; Eq3 = one,lpkf,lplf
         ; CLS: b(9)  - b(4) = 0,   ? δk  in 1st and 2nd equation
                b(10) - b(6) = 0,   ? θkk in 1st and 2nd equation
                b(11) - b(8) = 0,   ? θkl in 1st and 2nd equation
                b(12) - b(5) = 0,   ? δl  in 1st and 3rd equation
                b(13) - b(8) = 0,   ? θkl in 1st and 3rd equation
                b(14) - b(7) = 0    $ θll in 1st and 3rd equation
```

This imposes all of the necessary constraints across the second and third equations. It is generally observed that very large gains in efficiency often follow when the cross equation restrictions are imposed. This underscores the substantial collinearity in the unrestricted equation and raises the question of whether it can be estimated at all. In practical terms, if the data in the unrestricted equations are so collinear that the model cannot be estimated, then the restricted estimates will not be computable either. Ultimately, they are functions of the unrestricted estimates. But, for systems such as the translog model, this problem is circumvented by the estimator described in the next section.

As specified above, one is not guaranteed to obtain the same parameter estimates if a different variable is chosen as the numeraire. This is normally handled by obtaining maximum likelihood estimates, rather than two step GLS estimates. As noted earlier, using **SURE** in the fashion specified above does not produce maximum likelihood estimators, even with iteration. The problem is easily solved using the direct maximum likelihood estimator described in Section E24.3.

E24.2.5 Generalized Least Squares

In Chapter E15, a Time Series/Cross Section model was fit using 20 years of data for five firms. The following continues that example by relaxing the constraint of equal parameter vectors across equations. The model commands are as follows: (We begin by transforming the first 100 observations in the raw data set to the 20 observations used here.) First, move the data up to the first 20 rows of the data set.

```
CREATE     ; igm = i       ; fgm = f       ; cgm = c
           ; ich  = i[+20] ; fch  = f[+20] ; cch  = c[+20]
           ; ige  = i[+40] ; fge  = f[+40] ; cge  = c[+40]
           ; iwe  = i[+60] ; fwe  = f[+40] ; cwe  = c[+40]
           ; ius  = i[+80] ; fus  = f[+40] ; cus  = c[+40] $
SAMPLE     ; 1-20 $
NAMELIST   ; xgm = one,fgm,cgm ; xch = one,fch,cch
           ; xge  = one,fge,cge ; xwe = one,fwe,cwe
           ; xus  = one,fus,cus ; y = igm,ich,ige,iwe,ius $
```

Fit the basic model by iterated FGLS.

SURE ; Lhs = y
; Eq1 = xgm ; Eq2 = xch ; Eq3 = xge ; Eq4 = xwe ; Eq5 = xus $

Estimate the model with autocorrelation, with separate coefficients for each equation.

SURE ; Lhs = y
; Eq1 = xgm ; Eq2 = xch ; Eq3 = xge ; Eq4 = xwe ; Eq5 = xus
; Model = 1 $

Constrained FGLS imposes cross equation equality of coefficient vectors

SURE ; Lhs = y
; Eq1 = xgm ; Eq2 = xch ; Eq3 = xge ; Eq4 = xwe ; Eq5 = xus
; CLS: b(4)-b(1) = 0, b(7)-b(1) = 0, b(10)-b(1) = 0, b(13)-B(1) = 0,
b(5)-b(2) = 0, b(8)-b(2) = 0, b(11)-b(2) = 0, b(14)-B(2) = 0,
b(6)-b(3) = 0, b(9)-b(3) = 0, b(12)-b(3) = 0, b(15)-B(3) = 0 $

The listing below shows parts of the output from these commands. In the first, the full set of results is shown for the set of equations. For the autocorrelation model, the results which have changed are listed. Finally, in the constrained model, only one coefficient vector is estimated, so only the diagnostic statistics are shown. Some superfluous lines of results are omitted in each case.

The first is the base case, iterated FGLS estimates.

```
Criterion function for GLS is log-likelihood.
Iteration    0, GLS            =   -473.1602
Iteration    1, GLS            =   -469.5564
Iteration    2, GLS            =   -469.4238
Iteration    3, GLS            =   -469.4187
Iteration    4, GLS            =   -469.4182
Iteration    5, GLS            =   -469.4182
Iteration    6, GLS            =   -469.4182
Iteration    7, GLS            =   -469.4182
GLS              has converged.

-----------------------------------------------------------------------------
Estimates for equation: IGM.......................
Generalized  least squares regression ...........
LHS=IGM       Mean                   =      608.02000
              Standard deviation     =      309.57463
              Number of observs.     =             20
Model size    Parameters             =              3
              Degrees of freedom     =             17
Residuals     Sum of squares         =        122985.
              Standard error of e    =       85.05539
Fit           R-squared              =         .92054
              Adjusted R-squared     =         .91119
Model test    F[  2,    17] (prob) =     98.5(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|       32.7524 Log-Likelihood  =      -469.4182
Durbin-Watson   .986 Autocorrelation =          .5072
```

```
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
     IGM|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -179.671**     86.11059    -2.09  .0369    -348.445   -10.898
     FGM|     .12491***      .02072     6.03  .0000      .08429    .16552
     CGM|     .37993***      .03249    11.70  .0000      .31625    .44360
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Estimates for equation: ICH......................
Generalized  least squares regression ...........
LHS=ICH      Mean                 =      410.47500
             Standard deviation   =      125.39943
             Number of observs.   =             20
Model size   Parameters           =              3
             Degrees of freedom   =             17
Residuals    Sum of squares       =        139449.
             Standard error of e  =       90.56989
Fit          R-squared            =         .45090
             Adjusted R-squared   =         .38630
Model test   F[  2,    17] (prob) =     7.0(.0061)
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      32.7524 Log-Likelihood  =    -469.4182
Durbin-Watson 1.100 Autocorrelation =        .4501
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
     ICH|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    60.4012       98.22032      .61  .5386   -132.1071  252.9095
     FCH|     .11561**       .04749     2.43  .0149      .02253    .20869
     CCH|     .41414***      .10911     3.80  .0001      .20029    .62800
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
Estimates for equation: IGE......................
Generalized  least squares regression ...........
LHS=IGE      Mean                 =      102.29000
             Standard deviation   =       48.58450
             Number of observs.   =             20
Model size   Parameters           =              3
             Degrees of freedom   =             17
Residuals    Sum of squares       =        12073.6
             Standard error of e  =       26.64977
Fit          R-squared            =         .68329
             Adjusted R-squared   =         .64602
Model test   F[  2,    17] (prob) =    18.3(.0001)
Not using OLS or no constant. Rsqrd & F may be < 0
Model was estimated on Jun 09, 2011 at 10:47:05 AM
Log|W|      32.7524 Log-Likelihood  =    -469.4182
Durbin-Watson  .947 Autocorrelation =        .5265
```

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     IGE|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
Constant|   -23.1643       25.57100     -.91  .3650    -73.2826   26.9539
     FGE|     .03825***      .01207     3.17  .0015      .01460    .06190
     CGE|     .12797***      .02208     5.79  .0000      .08469    .17125
--------+-------------------------------------------------------------

----------------------------------------------------------------------
Estimates for equation: IWE......................
Generalized  least squares regression ...........
LHS=IWE      Mean                 =       86.12350
             Standard deviation   =       42.72556
             Number of observs.   =             20
Model size   Parameters           =              3
             Degrees of freedom   =             17
Residuals    Sum of squares       =        4918.21
             Standard error of e  =       17.00902
Fit          R-squared            =         .83318
             Adjusted R-squared   =         .81355
Model test   F[  2,    17] (prob) =    42.5(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Model was estimated on Jun 09, 2011 at 10:47:05 AM
Log|W|      32.7524 Log-Likelihood  =    -469.4182
Durbin-Watson 1.479 Autocorrelation =        .2603
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     IWE|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
Constant|   -34.5394*      18.24896    -1.89  .0584    -70.3067    1.2279
     FWE|     .03521***      .00898     3.92  .0001      .01762    .05281
     CWE|     .13070***      .01509     8.66  .0000      .10112    .16028
--------+-------------------------------------------------------------

----------------------------------------------------------------------
Estimates for equation: IUS......................
Generalized  least squares regression ...........
LHS=IUS      Mean                 =       61.80250
             Standard deviation   =       15.16693
             Number of observs.   =             20
Model size   Parameters           =              3
             Degrees of freedom   =             17
Residuals    Sum of squares       =        1153.98
             Standard error of e  =        8.23900
Fit          R-squared            =         .68938
             Adjusted R-squared   =         .65284
Model test   F[  2,    17] (prob) =    18.9(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      32.7524 Log-Likelihood  =    -469.4182
Durbin-Watson 2.158 Autocorrelation =       -.0791
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     IUS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    18.1185**      8.00495     2.26  .0236      2.4290   33.8079
     FUS|     .01513***      .00385     3.92  .0001      .00757    .02268
     CUS|     .03579***      .00680     5.27  .0000      .02247    .04911
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Iteration    1, GLS           =   -461.9742
Iteration    2, GLS           =   -461.9118
Iteration    3, GLS           =   -461.9110
Iteration    4, GLS           =   -461.9110
Iteration    5, GLS           =   -461.9110
GLS               has converged.

-----------------------------------------------------------------------------
Estimates for equation: IGM.......................
Generalized  least squares regression ............
LHS=IGM      Mean                    =      608.02000
             Standard deviation      =      309.57463
             Number of observs.      =             20
Model size   Parameters              =              3
             Degrees of freedom      =             17
Residuals    Sum of squares          =        81395.8
             Standard error of e     =       69.19530
Fit          R-squared               =         .94741
             Adjusted R-squared      =         .94122
Model test   F[  2,    17] (prob) =   153.1(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      32.0017 Log-Likelihood  =     -461.9110
Durbin-Watson 1.423 Autocorrelation =         .2886
RHO used for AR(1) corrected FGLS  =       .531273
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     IGM|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -46.5149       79.64759     -.58  .5592   -202.6213  109.5915
     FGM|     .09281***      .01701     5.46  .0000      .05947    .12616
     CGM|     .40669***      .04276     9.51  .0000      .32288    .49049
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Criterion function for GLS is log-likelihood.
Iteration    0, GLS           =   -473.1602
Iteration    1, GLS           =   -469.5564
Iteration    2, GLS           =   -469.4238
Iteration    3, GLS           =   -469.4187
Iteration    4, GLS           =   -469.4182
Iteration    5, GLS           =   -469.4182
Iteration    6, GLS           =   -469.4182
Iteration    7, GLS           =   -469.4182
GLS               has converged.
```

```
-----------------------------------------------------------------------------
Estimates for equation: IGM......................
Generalized  least squares regression ...........
LHS=IGM      Mean                 =      608.02000
             Standard deviation   =      309.57463
             Number of observs.   =             20
Model size   Parameters           =              3
             Degrees of freedom   =             17
Residuals    Sum of squares       =    .327447E+07
             Standard error of e  =      438.88021
Fit          R-squared            =       -1.11562
             Adjusted R-squared   =       -1.36452
Not using OLS or no constant. Rsqrd & F may be < 0
Log|W|      40.8290 Log-Likelihood  =     -550.1839
Durbin-Watson  .050 Autocorrelation =         .9752
Wald test:Chi-squared[12]=2038.6522, Prob =  .0000

--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     IGM|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -120.093***     5.89234   -20.38  .0000    -131.642   -108.544
     FGM|     .07445***      .00288    25.82  .0000      .06880     .08010
     CGM|     .05133***      .00541     9.48  .0000      .04072     .06195
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The figure below plots the residuals equation by equation for the five equations. This is a counterpart to the first 100 points in the figure at the end of Chapter E15. As can be seen by comparing the two figures, the restriction of identical coefficients in the two equations brings a considerable change in the fit of some of the equations.

```
SAMPLE     ; 1-100 $
CREATE     ; If (_obsno <= 20)
             e = igm - b(1) - b(2)*fgm - b(3)*cgm $
CREATE     ; If (_obsno > 20 & _obsno <= 40)
             e = ich[-20] - b(4) - b(5)*fch[-20] - b(6)*cch[-20] $
CREATE     ; If (_obsno > 40 & _obsno <= 60)
             e = ige[-40] - b(7)  - b(8)*fge[-40] - b(9)*cge[-40] $
CREATE     ; If (_obsno > 60 & _obsno <= 80)
             e = iwe[-60] - b(10) - b(11)*fwe[-60] - b(12)*cwe[-60] $
CREATE     ; If (_obsno > 80)
             e = ius[-80] - b(13) - b(14)*fus[-80] - b(15)*cus[-80] $
CREATE     ; obs = Trn(1,1) $
PLOT       ; Lhs = obs ; Rhs = e
           ; Bars = 0 ; Spikes = 20.5,40.5,60.5,80.5 ; Fill ; Endpoints = 0,100
           ; Title = Residuals from separate equations $
```

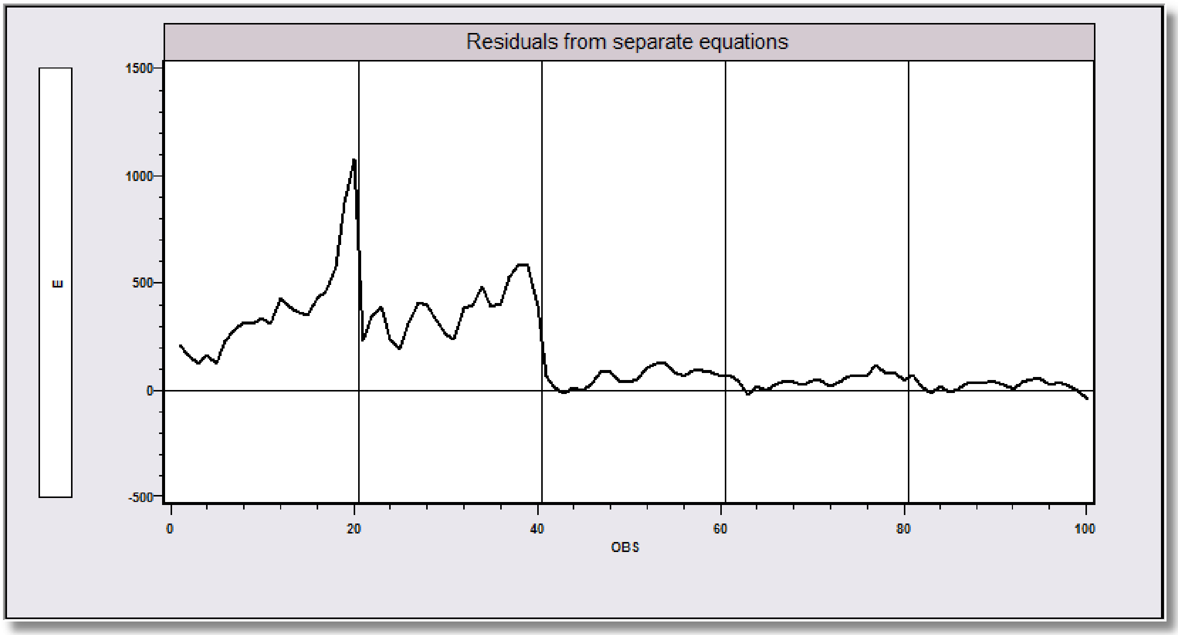


Figure E24.2 Plot of Residuals from Separate Equations

E24.2.6 Technical Details for Generalized Least Squares

The generalized least squares (GLS) approach to estimation is based on the 'stacked' system,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}$$

or $\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$

where $\quad \mathrm{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$

and $\quad \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}\otimes\mathbf{I}.$

The GLS estimator is

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}[\mathbf{X}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{y}].$$

The feasible GLS (FGLS) estimator is obtained in two steps. At the first, single equation ordinary least squares is used one equation at a time to compute $\mathbf{b}_i$. Then, $\mathbf{b}_i$ is used to obtain residuals $\mathbf{e}_i$, which are used to compute

$$s_{ij} = \mathbf{e}_i'\mathbf{e}_j / T.$$

FGLS is then computed using this estimator of **Σ**. The estimated asymptotic covariance matrix is the estimate of the inverse matrix in brackets above. If desired, the estimator can then be allowed to iterate to convergence. Convergence is checked at the *i*th iteration using $\text{Max}(|b_k(i)/b_k(i\text{-}1) - 1|) <$ 1.d-9. That is, the largest absolute percentage change in any parameter from one iteration to the next must be less than 1.d-9.

The GLS procedure is based exactly on the textbook formulas. In computing the estimate of **Σ**, we do not make any corrections for degrees of freedom. But, results given with the initial output from the regressions for each equation provide the values needed if you wish to make the correction later. The model results do not include the correlation matrix for the residuals. Since the covariance matrix is kept as *sigma* this can also be computed later. The commands would be:

```
SURE        ; ... $
MATRIX      ; se = Diag (sigma) ; se = Isqr(se) * sigma * Isqr(se) $
```

NOTE: With no constraints imposed, this iterative SURE estimator converges to the maximum likelihood estimator. This is not the case if there are constraints imposed. We will return to this subject below.

In the autocorrelation model, the parameters are estimated twice. In the first pass, the model is fit with no autocorrelation. The autocorrelation coefficients are then estimated using $r_j = 1 - ½DW_j$. In the next pass, the models are fit using the Prais-Winsten transformation – the first observation is transformed by $\sqrt{1-r_j^2}$, not dropped, as it would be for the Cochrane-Orcutt estimator.

E24.3 Maximum Likelihood Estimation of Constrained Linear Systems

The **SURE** command will produce feasible GLS estimates for the multivariate regression model. It can also be allowed to iterate in order to produce maximum likelihood estimates. This does the intended job if there are no restrictions on the parameters. But, if you use the iterated GLS estimator, then impose linear restrictions, (as we did in the previous example) the restricted estimator will be a hybrid of the GLS and MLE, as *LIMDEP* will take the unconstrained MLE and apply the constrained GLS formula to it to get the constrained estimator. Although the unconstrained estimator is MLE, this is not the way to get the constrained MLE. If you have a restricted model, you can use one of the following procedures:

Use **SURE**'s FGLS procedure, with the restrictions, and limit the number of iterations to one. This will obtain the constrained two step GLS estimator (see, e.g., Greene (2011) or Johnston and DiNardo (1997)). But, if you are estimating a translog model or other singular system of demand equations in which you have dropped an equation to achieve nonsingularity, these estimates will not be invariant to which equation you drop. Use the MLE procedure described below.

The following is for models in which the constraints are equalities of the parameters across (or within) equations. (See Greene (2011).) (For other types, use constrained FGLS.) Consider, for example, the following model:

	Variable			
Equation	*one*	x_1	x_2	x_3
y_1	a_1	a_2	a_3	a_4
y_2	a_3		b_1	
y_3	a_4			c_1

This model has eight parameters with two equalities. We view the set of parameters as arranged in a 'parameter matrix,' such as the one in the box above. It has number of rows equal to the number of equations and number of columns equal to the total number of independent variables (including *one*) in the model. The fact that not all variables appear in all equations shows up as empty cells (or zeros) in the matrix. Note the arrangement which implies that each column of the parameter matrix applies to one of the independent variables, and each row corresponds to an equation, or dependent variable.

E24.3.1 Command for ML Estimation of Constrained SURE Systems

To set up such a model, you must inform *LIMDEP* of the dimensions of the problem, what the nonzero values in the parameter matrix are and where they are. Your command does that as follows:

```
SURE        ; Lhs     = list of dependent variables
            ; Rhs     = full list of independent variables
            ; Labels  = the labels to use for the parameters
            ; Pattern = the parameter matrix $
```

Note the following aspects of this command:

- Dimensions of the problem: The number of rows in the parameter matrix equals the number of variables in your Lhs list. The number of columns in the matrix equals the number of variables in your Rhs list.

- The Labels are the names you want to use for the parameters in the model. These may be any symbols with up to five characters. (Anything over five is truncated.) The one exception is that you may not use the '*' character in a label.

- The Pattern is simply a listing of the rows of the parameter matrix, with labels and zeros, moving rowwise through the matrix and separating values with commas. As will be obvious from the examples, the best way to set this up is, literally, to lay out the matrix in the command.

- You may use n*0, e.g., 10*0, where 'n' is from 2 to 50 to provide a string of zeros.

The full command for the example above would be:

```
SURE      ; Lhs     = y1,y2,y3
          ; Rhs     = one,x1,x2,x3
          ; Pattern = a1, a2, a3, a4,
                      a3, 0,  b1, 0,
                      a4, 0,  0,  c1
          ; Labels  = a1, a2, a3, a4, b1, c1 $
```

Note that the pattern matrix automatically (and visually) imposes any equality constraints on the parameters within or across equations. For example, the fact that *a*3 is used in two places in the matrix ensures that this constraint will be imposed.

For laying out the parameter matrix, it will often help to arrange the Rhs list and Pattern list exactly above one another in correspondence. For example, the command for the translog model given earlier, using the same variable names, would be

```
SURE      ; Lhs     = cost,sk,sl
          ; Labels  = a,    cy,    dk,    dl,     tkk,    tll,     tkl,  cyy
          ; Rhs     = one,  lpkf,  lplf,  lpkf2,  lplf2,  lpkfplf, y,    y2
          ; Pattern = a,    dk,    dl,    tkk,    tll,    tkl,     cy,   cyy
                      dk,   tkk,   tkl,           5*0
                      dl,   tkl,   tll,           5*0  $
```

There are no other options for this model.

E24.3.2 Model Output for the Maximum Likelihood Estimator

Model output for the constrained MLE consists of an initial trace of

1. log likelihood function,
2. log determinant of **S**,
3. $\mathbf{g}'\mathbf{H}^{-1}\mathbf{g}$, where **g** = gradient and **H** = Hessian.

We use the last of these as the convergence criterion. This is a scale free measure, which is invariant to the sample size. (See Chapter R9.) Since the likelihood is globally concave, convergence will be fast and monotonic. Moreover, the entire optimization is based on the moments of **X** and **Y** (sums of squares and cross products), so the amount of computation is independent of the sample size. The one pass through the data to obtain the moment matrices will be the only significant amount of time spent.

Remaining output is a display of the pattern matrix and a list of the coefficient estimates, standard errors, and t ratios. The final display is the maximum likelihood estimator of **E**. Results kept by this estimator are the same as for the GLS estimator. The coefficient vector kept is the unconstrained parameter vector, *b* and the estimated asymptotic covariance matrix, *varb*. As before, *sigma* is the estimate of the disturbance covariance matrix. For example, in the first illustration above, the parameter matrix has eight nonzero cells, but only six free parameters. The parameter vector, *b*, would be the six values $[a_1,a_2,a_3,a_4,b_1,c_1]$. Finally, the *Last Model* labels are the ones you provide in your **; Labels** list. A coefficient matrix, *b_sur_ml* is constructed from the estimated parameter vector in the same fashion as described earlier for the GLS estimator.

E24.3.3 Application

The data below will be used to fit a translog cost/demand system. The data are from Berndt and Wood (1975). The authors estimated a model of production in the U.S. manufacturing sector for 1947-1971. The four factors are capital (*K*), labor (*L*), energy (*E*) and materials (*M*). Quantities are denoted '*Q*' while price indices are denoted '*P*.' The output quantity is denoted '*Y*' in the model below. The reader is referred to Greene (2011) or Berndt and Wood (1975) for details on the translog model. For convenience, denote by *k*, *l*, and *e*, the logs of the normalized prices, as in $k = \log(PK/PM)$, and so on. Let *y* denote ln*Y*, *c* denote log(*Total Cost*/*PM*), and S_i denote the cost shares. The equations of the full model are:

$$c = \alpha + \beta_k k + \beta_l l + \beta_e e + \beta_y y + \theta_{yy} y^2 + \delta_k ky + \delta_l ly + \delta_e ey$$
$$+ \gamma_{kk} k^2/2 + \gamma_{kl} kl + \gamma_{ke} ke + \gamma_{ll} l^2/2 + \gamma_{le} le + \gamma_{ee} e^2/2 + \varepsilon_c,$$
$$S_k = \beta_k + \gamma_{kk} k + \gamma_{kl} l + \gamma_{ke} e + \delta_k y + \varepsilon_k,$$
$$S_l = \beta_l + \gamma_{kl} \mathrm{k} + \gamma_{ll} l + \gamma_{le} e + \delta_l y + \varepsilon_l,$$
$$S_k = \beta_k + \gamma_{ke} k + \gamma_{le} l + \gamma_{ee} e + \delta_e y + \varepsilon_e.$$

There are a total of 30 parameters in the model, but 15 constraints leave only 15 free parameters to be estimated.

quantity	*qk*	*ql*	*qe*	*qm*	*pk*	*pl*	*pe*	*pm*
196.205	9.3130	45.0961	7.75697	120.207	1.00000	1.00000	1.00000	1.00000
182.829	10.6264	43.9693	7.20873	106.468	1.00270	1.15457	1.30259	1.05526
191.077	11.5423	41.8166	7.91134	113.107	0.74371	1.15584	1.19663	1.06225
217.532	11.9624	44.4985	8.40976	129.378	0.92497	1.23535	1.21442	1.12430
235.289	12.2972	48.7602	9.16439	136.689	1.04877	1.33784	1.25180	1.21694
244.086	13.0450	51.1402	9.22766	141.135	0.99744	1.37949	1.27919	1.19961
269.111	13.6777	54.4577	9.97689	156.706	1.00654	1.43458	1.27505	1.19044
247.312	14.2198	51.2944	10.07850	142.018	1.08757	1.45362	1.30357	1.20612
277.789	14.7225	54.0984	10.39200	159.050	1.10315	1.51121	1.34277	1.23835
281.382	15.1736	55.7854	10.95190	162.295	0.99607	1.58187	1.37155	1.29336
282.153	16.0311	55.9122	11.82740	163.127	1.06321	1.64641	1.38010	1.30703
262.425	16.8214	52.6973	11.22090	150.735	1.15619	1.67389	1.39338	1.32700
291.418	16.9557	56.4288	11.95920	169.792	1.30758	1.73430	1.36756	1.30774
296.644	16.9042	56.9827	12.16510	169.226	1.25413	1.78280	1.38025	1.33946
297.000	17.1108	56.0163	12.34530	167.971	1.26329	1.81977	1.37631	1.34319
320.884	17.2227	58.5997	12.85290	178.634	1.26525	1.88531	1.37689	1.34745
337.855	17.4505	59.6128	13.67400	191.822	1.32294	1.93379	1.34737	1.33144
359.146	17.8079	61.1658	13.70810	198.323	1.32798	2.00998	1.38969	1.35197
389.238	18.4595	64.6947	14.09460	215.563	1.40659	2.05539	1.38635	1.37543
417.185	19.6165	69.2726	14.93410	228.398	1.45100	2.13441	1.40102	1.41879
425.702	21.2163	70.1610	15.81500	234.596	1.38618	2.20617	1.39197	1.42428
451.210	22.4894	72.3024	16.21180	250.484	1.49901	2.33869	1.43389	1.43481
466.830	23.5281	74.2756	17.05680	253.226	1.44957	2.46412	1.46481	1.53356
446.710	24.7325	71.2039	18.57820	244.296	1.32465	2.60532	1.45907	1.54758
457.986	25.6062	68.9305	17.90340	263.076	1.20178	2.76026	1.64689	1.54979

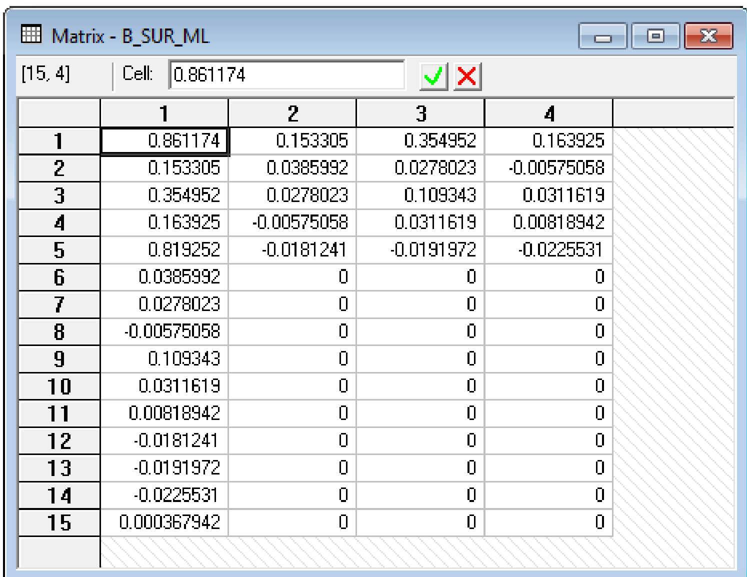
The following results provide estimates of the full model first. We then test the hypothesis of the cross equation restrictions in the share equations by estimating them as a system without the cost equation, and with and without the cross equation equality restrictions. A likelihood ratio test is used to test the hypothesis (such as it is – without the assumption of the restrictions, the share equations have no theoretical basis).

```
CREATE        ; cost      =  pk*qk + pl*ql + pe*qe + pm*qm
              ; sk        =  pk*qk/cost
              ; sl        =  pl*ql/cost
              ; se        =  pe*qe/cost
              ; sm        =  pm*qm/cost
              ; c         =  Log(cost/pm)
              ; k         =  Log(pk/pm) ; l = Log(pl/pm) ; e = Log(pe/pm)
              ; kk        =  .5*k*k ; ll = .5*l*l
              ; ee        =  .5*e*e ; kl = k*l  ; ke = k*e ; le = l*e
              ; y         =  Log(quantity) ; yy = .5*y*y ; ky = k*y
              ; ly        =  l*y  ; ey = e*y $
SURE          ; Lhs       =  c,sk,sl,se
              ; Labels    =  a,bk,bl,be,by,ckk,ckl,cke,cll,cle,cee,dky,dly,dey,byy
              ; Rhs       =  one,k,l,e,y,kk,kl,ke,ll,le,ee,ky,ly,ey,yy
              ; Pattern   =  a,bk,bl,be,by,ckk,ckl,cke,cll,cle,cee,dky,dly,dey,byy,
                             bk,ckk,ckl,cke,dky,10*0,
                             bl,kl,cll,cle,dly,10*0,
                             be,cke,cle,cee,dey,10*0 $
SURE          ; Lhs       =  sk,sl,se
              ; Labels    =  bk,bl,be,ckk,ckl,cke,cll,cle,cee,dky,dly,dey
              ; Rhs       =  one,k,l,e,y
              ; Pattern   =  bk,ckk,ckl,cke,dky,
                             bl,ckl,cll,cle,dly,
                             be,cke,cle,cee,dey $
CALC          ; lc        =  logl $
SURE          ; Lhs       =  sk,sl,se
              ; Labels    =  bk,bl,be,ckk,ckl,cke,cll,cle,cee,dky,dly,dey,clk,cel,cek
              ; Rhs       =  one,k,l,e,y
              ; Pattern   =  bk,ckk,ckl,cke,dky,
                             bl,clk,cll,cle,dly,    ? no constraint clk = ckl
                             be,cek,cel,cee,dey  $ same: cek & cke, cel & cle
CALC          ; List ;       lu = logl
              ; lrt       =  2*(lu-lc) $
```

The first model is the full four equation model with all constraints. The second is the three share equations with constraints imposed. The third estimates the three share equations without restrictions.

```
-----------------------------------------------------------------------------
Constrained MLE for Multivariate Regression Model
First iter.  0 F=  215.3786 log|W|= -28.5818 g<H>g=  3.0083
Last  iter.  7 F=  436.9910 log|W|= -46.3108 g<H>g=   .0000
Number of observations used in estimation =    25
Model specification is given in run log
--------+--------------------------------------------------------------------
        |                 Standard             Prob.      95% Confidence
 SUR_MLE|  Coefficient      Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       A|     .86117        .89696      .96  .3370     -.89684   2.61919
      BK|     .15331***     .02558     5.99  .0000      .10316    .20345
      BL|     .35495***     .07169     4.95  .0000      .21445    .49546
      BE|     .16392***     .01730     9.48  .0000      .13002    .19783
      BY|     .81925***     .31496     2.60  .0093      .20195   1.43655
     CKK|     .03860***     .00529     7.30  .0000      .02823    .04897
     CKL|     .02780***     .00820     3.39  .0007      .01174    .04387
     CKE|    -.00575**      .00224    -2.57  .0101     -.01013   -.00137
     CLL|     .10934***     .02479     4.41  .0000      .06075    .15794
     CLE|     .03116***     .00530     5.88  .0000      .02078    .04154
     CEE|     .00819*       .00475     1.72  .0849     -.00113    .01750
     DKY|    -.01812***     .00482    -3.76  .0002     -.02757   -.00868
     DLY|    -.01920        .01369    -1.40  .1609     -.04603    .00764
     DEY|    -.02255***     .00325    -6.93  .0000     -.02893   -.01618
     BYY|     .00037        .05521      .01  .9947     -.10785    .10858
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

   SIGMA|           1              2              3              4
--------+------------------------------------------------------------
       1|    .121761E-03    .656679E-05   -.350106E-06    .171339E-05
       2|    .656679E-05    .638215E-05    .356703E-05    .193327E-05
       3|   -.350106E-06    .356703E-05    .226717E-04    .201951E-05
       4|    .171339E-05    .193327E-05    .201951E-05    .114032E-05
```


Matrix - B_SUR_ML

[15, 4] Cell: 0.861174

	1	2	3	4
1	0.861174	0.153305	0.354952	0.163925
2	0.153305	0.0385992	0.0278023	-0.00575058
3	0.354952	0.0278023	0.109343	0.0311619
4	0.163925	-0.00575058	0.0311619	0.00818942
5	0.819252	-0.0181241	-0.0191972	-0.0225531
6	0.0385992	0	0	0
7	0.0278023	0	0	0
8	-0.00575058	0	0	0
9	0.109343	0	0	0
10	0.0311619	0	0	0
11	0.00818942	0	0	0
12	-0.0181241	0	0	0
13	-0.0191972	0	0	0
14	-0.0225531	0	0	0
15	0.000367942	0	0	0

Figure E24.3 Coefficient Matrix for MLE

```
-----------------------------------------------------------------------------
Constrained MLE for Multivariate Regression Model
First iter.  0 F=  199.9708 log|W|= -24.5113 g<H>g=  2.6613
Last  iter.  5 F=  359.0365 log|W|= -37.2365 g<H>g=   .0000
Number of observations used in estimation =    25
Model:    ONE   K     L     E     Y
--------   ----- ----- ----- ----- -----
SK        BK    CKK   CKL   CKE   DKY
SL        BL    CKL   CLL   CLE   DLY
SE        BE    CKE   CLE   CEE   DEY
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 SUR_MLE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
      BK|     .15241***      .02567     5.94  .0000       .10210     .20272
      BL|     .35725***      .07185     4.97  .0000       .21642     .49808
      BE|     .16386***      .01729     9.48  .0000       .12998     .19774
     CKK|     .03713***      .00542     6.85  .0000       .02650     .04776
     CKL|     .02798***      .00826     3.39  .0007       .01180     .04416
     CKE|    -.00615***      .00226    -2.73  .0064      -.01057    -.00173
     CLL|     .10992***      .02488     4.42  .0000       .06116     .15869
     CLE|     .03126***      .00529     5.90  .0000       .02088     .04163
     CEE|     .00808*        .00475     1.70  .0893      -.00124     .01739
     DKY|    -.01800***      .00484    -3.72  .0002      -.02748    -.00852
     DLY|    -.01962         .01373    -1.43  .1528      -.04653     .00728
     DEY|    -.02255***      .00325    -6.94  .0000      -.02892    -.01618
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
   SIGMA|            1              2              3
--------+---------------------------------------------
       1|   .630167E-05   .338359E-05   .190080E-05
       2|   .338359E-05   .226901E-04   .196970E-05
       3|   .190080E-05   .196970E-05   .112857E-05

-----------------------------------------------------------------------------
Constrained MLE for Multivariate Regression Model
First iter.  0 F=  199.9708 log|W|= -24.5113 g<H>g=  2.6646
Last  iter.  2 F=  361.4197 log|W|= -37.4272 g<H>g=   .0000
Number of observations used in estimation =    25
Model:    ONE   K     L     E     Y
--------   ----- ----- ----- ----- -----
SK        BK    CKK   CKL   CKE   DKY
SL        BL    CLK   CLL   CLE   DLY
SE        BE    CEK   CEL   CEE   DEY
```

```
--------+----------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
 SUR_MLE|  Coefficient       Error       z     |z|>Z*         Interval
--------+----------------------------------------------------------------
      BK|      .22250***      .05134     4.33  .0000      .12188    .32312
      BL|      .36031***      .09969     3.61  .0003      .16492    .55571
      BE|      .18247***      .02237     8.15  .0000      .13861    .22632
     CKK|      .03570***      .00516     6.92  .0000      .02560    .04581
     CKL|      .05256***      .01534     3.43  .0006      .02249    .08262
     CKE|     -.01076         .01567     -.69  .4924     -.04148    .01996
     CLL|      .11574***      .02979     3.89  .0001      .05736    .17412
     CLE|      .03379         .03044     1.11  .2669     -.02586    .09345
     CEE|      .00701         .00683     1.03  .3050     -.00638    .02040
     DKY|     -.03148***      .00964    -3.26  .0011     -.05037   -.01258
     DLY|     -.02069         .01872    -1.11  .2691     -.05739    .01600
     DEY|     -.02616***      .00420    -6.22  .0000     -.03439   -.01792
     CLK|      .01513         .01001     1.51  .1309     -.00450    .03475
     CEL|      .03805***      .00669     5.69  .0000      .02494    .05115
     CEK|     -.00724***      .00225    -3.22  .0013     -.01165   -.00284
--------+----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-------------------------------------------------------------------------
   SIGMA|             1             2             3
--------+-----------------------------------------
       1|   .557408E-05   .290208E-05   .168248E-05
       2|   .290208E-05   .210204E-04   .174947E-05
       3|   .168248E-05   .174947E-05   .105881E-05

[CALC] LU      =    361.4196710
[CALC] LRT     =      4.7664414
Calculator: Computed   2 scalar results
```

E24.3.4 Technical Details

The maximum likelihood estimator uses Newton's method. Let **Y** denote the $n \times M$ matrix of data on the M Lhs variables specified and let **X** denote the counterpart for the K Rhs variables. All computations are based on moments of the data, so after a single pass through the data set to accumulate $\mathbf{Y}'\mathbf{Y}$, $\mathbf{X}'\mathbf{Y}$, and $\mathbf{X}'\mathbf{X}$, iterations are extremely rapid. Let $\mathbf{\Pi}$ denote the $K \times M$ parameter matrix. $\mathbf{\Pi}$ is the transpose of the matrix defined in the **; Pattern** specification of the **SURE** command. Let **E** denote the $n \times M$ matrix of disturbances. Each row of **E** is the M disturbances for the M equations for the ith observation, $\boldsymbol{\varepsilon}_t$. Then,

$$\mathbf{\Sigma} = E[(1/n)\mathbf{E}'\mathbf{E}].$$

The model is

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{E}$$

Let **P** denote any estimate of $\mathbf{\Pi}$. Then, the residuals are

$$\mathbf{U} = \mathbf{Y} - \mathbf{XP}.$$

The sample estimate of $\mathbf{\Sigma}$ will always be

$$\mathbf{W} = (1/n)\mathbf{U}'\mathbf{U}.$$

The concentrated log likelihood function for this model is

$$\begin{aligned}\log L^* &= \text{a constant} - ½ \log \det[(1/n)(\mathbf{Y}\text{-}\mathbf{X\Pi})'(\mathbf{Y}\text{-}\mathbf{X\Pi})] \\ &= \text{a constant} - ½ \log \det(\mathbf{\Omega}).\end{aligned}$$

Note that $\mathbf{\Omega}$ is not equal to $\mathbf{\Sigma}$, though $E[\mathbf{\Omega}]$ equals $\mathbf{\Sigma}$. Finally, define the following matrices:

$$\mathbf{S_{xx}} = (1/n)\mathbf{X}'\mathbf{X}, \ \mathbf{S_{xy}} = (1/n)\mathbf{X}'\mathbf{Y}, \text{ and } \mathbf{S_{yy}} = (1/n)\mathbf{Y}'\mathbf{Y}.$$

For the model being estimated, note that $\mathbf{\Pi}$ has a number of zeros in it, and many of the elements are equal to each other. We will impose these constraints later.

$$\partial \log L^*/\partial\mathbf{\Pi} = (1/n)\mathbf{X}'\mathbf{E}\mathbf{\Omega}^{\mathbf{-1}} = \mathbf{G}^*.$$

Defining $\boldsymbol{\pi}$ to be the column vector obtained by stacking the columns of $\mathbf{\Pi}$, we have

$$\partial^2 \log L^*/\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}' = \mathbf{\Omega}^{-1}\otimes\mathbf{S_{xx}} = \mathbf{H}^*.$$

Let $\boldsymbol{\gamma}$ = vector of Q nonzero elements in $\boldsymbol{\pi}$.

In order to implement Newton's method, we assemble a column vector from $\mathbf{G}^*$ by extracting the elements corresponding to nonzero elements of $\mathbf{\Pi}$. Denote this vector $\mathbf{g}$. Likewise, extract the relevant elements from $\mathbf{H}^*$ into a matrix $\mathbf{H}$ of much smaller dimension. (Rows and columns of $\mathbf{H}^*$ corresponding to zeros in $\mathbf{\Pi}$ are discarded. To impose the equality constraints in $\mathbf{\Pi}$, define $\boldsymbol{\beta}$ to be the unique, free parameters in the model. Thus, there are, say, J elements in $\boldsymbol{\beta}$. There are, say, Q elements in $\boldsymbol{\gamma}$, and $Q \geq J$. Several elements in $\boldsymbol{\gamma}$ may equal the same element of $\boldsymbol{\beta}$. Define the matrix $\mathbf{K}$ such that

$$\mathbf{K}_{ij} = 1 \text{ if } \gamma_i = \beta_j \text{ and 0 otherwise.}$$

Therefore, $\mathbf{K}$ has L rows and J columns. Every row contains a single one and J-1 zeros. Finally

$$\partial \log L^*/\partial\boldsymbol{\beta} = \mathbf{K}'\mathbf{g}$$

and

$$\partial^2 \log L^*/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \mathbf{K}'\mathbf{HK}.$$

These give the necessary elements for the Newton iterations.

E24.4 Instrumental Variables (3SLS) Estimation of a Set of Linear Equations

The setup for three stage least squares (3SLS) is identical to that for the GLS (not the maximum likelihood) estimator for the SURE model as shown in Section E25.2. The only change is the addition of a set of instrumental variables to the command. The command is

```
3SLS    ; Lhs  = y1,y2,...,ym
        ; Eq1  = list of Rhs variables in first equation
        ; Eq2  = list of Rhs variables in second equation
        ...
        ; EqM  = list of Rhs variables in Mth equation (up to 20 equations)
        ; Inst = complete list of exogenous variables $
```

The list of exogenous variables should include all of the exogenous variables appearing on the right hand sides of the equations of the model (only once in the list) and may include any other variables as well.

This estimator is obtained by first regressing all variables on the right hand side of each equation on all of the variables in the list of instruments and retaining the fitted values. Any variable in an Eqn list which appears in the Inst list as well is reproduced exactly since in this event, this first stage regression produces a perfect fit with a coefficient of one on that variable and zeros for all the others. Thereafter, the procedure is identical to the SURE procedure. Note, the fitted variables are not actually created; only the necessary sample moments using the fitted values where appropriate are physically retained for the computations.

After estimation, the disturbance covariance matrix is estimated using the original variables, not the fitted ones. This procedure can be allowed to iterate by specifying

```
        ; Maxit = maximum
```

If you do not provide this, the default is one iteration. To obtain Zellner's three stage least squares estimator, use

```
        ; Maxit = 0
```

Do note, iterated 3SLS does not bring gains in efficiency and does not produce an MLE. Moreover, iterated 3SLS frequently differs dramatically from 2SLS.

Application to Klein's Model I

We continue the example of Section E21.2 with

```
3SLS    ; Lhs = c,i,wp
        ; Eq1 = cons ; Eq2 = invs ; Eq3 = wage
        ; Inst = exog
        ; Maxit = 0 $
```

The iteration produces the trace:

```
Iteration    0, maximum |Δb/b|=    1.000000
Iteration    1, maximum |Δb/b|=    6.218180
```

The model output is as follows:

```
Criterion function is max(abs(%chg in b(i))).
Iteration    0, 3SLS          =    1.000000
Iteration    1, 3SLS          =    6.218180

-----------------------------------------------------------------------------
Estimates for equation: C........................
InstVar/GLS  least squares regression ...........
LHS=C        Mean                    =       53.99524
             Standard deviation      =        6.86087
             Number of observs.      =             21
Model size   Parameters              =              4
             Degrees of freedom      =             17
Residuals    Sum of squares          =        15.1599
             Standard error of e     =         .94433
Fit          R-squared               =         .98011
             Adjusted R-squared      =         .97660
Model test   F[  3,    17] (prob) =   279.2(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Durbin-Watson 1.425 Autocorrelation =          .2875
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       C|  Coefficient        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    16.4408***     1.30455    12.60  .0000     13.8839   18.9977
       P|      .12489        .10813     1.16  .2481     -.08704    .33682
    PLAG|      .16314        .10044     1.62  .1043     -.03371    .36000
       W|      .79008***     .03794    20.83  .0000      .71572    .86444
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Estimates for equation: I........................
InstVar/GLS  least squares regression ...........
LHS=I        Mean                    =        1.26667
             Standard deviation      =        3.55195
             Number of observs.      =             21
Model size   Parameters              =              4
             Degrees of freedom      =             17
Residuals    Sum of squares          =        35.5818
             Standard error of e     =        1.44674
Fit          R-squared               =         .82581
             Adjusted R-squared      =         .79507
Model test   F[  3,    17] (prob) =    26.9(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Durbin-Watson 1.996 Autocorrelation =          .0021
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       I|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    28.1778***     6.79377     4.15  .0000     14.8623   41.4934
       P|    -.01308         .16190     -.08  .9356     -.33039    .30423
    PLAG|     .75572***      .15293     4.94  .0000      .45598   1.05547
    KLAG|    -.19485***      .03253    -5.99  .0000     -.25861   -.13109
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Estimates for equation: WP........................
InstVar/GLS  least squares regression ...........
LHS=WP       Mean                  =       36.36190
             Standard deviation    =        6.30440
             Number of observs.    =             21
Model size   Parameters            =              4
             Degrees of freedom    =             17
Residuals    Sum of squares        =        8.84045
             Standard error of e   =         .72113
Fit          R-squared             =         .98626
             Adjusted R-squared    =         .98384
Model test   F[  3,    17] (prob) =   406.8(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Durbin-Watson 2.155 Autocorrelation =        -.0775
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      WP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.79722        1.11585     1.61  .1073     -.38982   3.98425
       X|     .40049***      .03181    12.59  .0000      .33814    .46285
    XLAG|     .18129***      .03416     5.31  .0000      .11434    .24824
       A|     .14967***      .02794     5.36  .0000      .09492    .20443
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E25: Nonlinear Systems of Regression Equations

E25.1 Introduction

This chapters present methods of estimating the parameters of the regression system

$$y_1 = f_1(\mathbf{x}_1,\boldsymbol{\beta}) + \varepsilon_1$$

$$y_2 = f_2(\mathbf{x}_2,\boldsymbol{\beta}) + \varepsilon_2$$

$$\ldots$$

$$y_M = f_M(\mathbf{x}_M,\boldsymbol{\beta}) + \varepsilon_M$$

or

$$\mathbf{y} = \mathbf{f}(\mathbf{X},\boldsymbol{\beta}) + \boldsymbol{\varepsilon},.$$

We assume $E[\boldsymbol{\varepsilon}|\text{all } \mathbf{x}] = \mathbf{0}$ and $E[\boldsymbol{\varepsilon\varepsilon}'|\text{all } \mathbf{x}] = \boldsymbol{\Sigma}$.

As stated, the model is a possibly nonlinear system of seemingly unrelated regressions. However, for some settings, the **x** vectors on the right hand sides of the equations may include endogenous variables, y_j, from other equations. That is, we also accommodate systems of simultaneous equations.

The system may contain up to 50 equations and up to 150 unique parameters. As defined, there is a single parameter vector, $\boldsymbol{\beta}$, to be estimated, though subsets of parameters can appear in each equation, and this is just a notational convenience. Estimates of the elements of $\boldsymbol{\Sigma}$ are also obtained.

The estimation procedures available for this model are:

- Nonlinear OLS, equation by equation (NLOLS): $\boldsymbol{\Sigma} = \text{diag}(\sigma_1,\ldots,\sigma_M)$
- Nonlinear equation by equation instrumental variables (NL2SLS)
- Nonlinear GLS (NLSUR): $\boldsymbol{\Sigma}$ = a full positive definite matrix
- Nonlinear GLS with instrumental variables (NL3SLS)
- Multiple equation GMM

The cases in which $\boldsymbol{\Sigma}$ is diagonal and there are no cross equation restrictions or equalities will replicate the nonlinear least squares and nonlinear instrumental variables equations estimators described earlier. The reasons that you might use this estimator in these cases are, first, estimating the equations jointly, even if uncorrelated, will be faster and, second, with this estimator, you can impose cross equation restrictions.

E25.2 Nonlinear Systems – The NLSUR Command

The essential command for the nonlinear system shown above is

NLSUR ; Lhs = ... the list of dependent variables
; Fn1 = ... the first equation
; Fn2 = ... the second equation
...
; FnM = ... the last equation (up to 20 equations)
; Labels = ... a list of labels for the parameters
; Start = ... the starting values for the iterations $

This setup is for a full, unrestricted Σ, which is estimated as part of the estimation process. The setup for the functions is exactly that shown in Section E14.3. All of the elements, functions, etc., shown there apply fully to these equation definitions. Note, as well that the parameters defined by the labels may appear anywhere in any equation, without restriction. That is, there is no presumption that any particular parameter applies to or belongs in any specific equation. Every equation is assumed to involve some or all of the parameters.

NOTE: The recursion feature and user supplied derivatives feature, both described in Section E14.5 are *not* supported for the **NLSUR** command.

The command shown above specifies a set of Lhs variables. The estimation criterion function will be based on the implied residuals,

$$e_j = y_j - Fn_j$$

(for example, the sums of squares). You may, instead, use the functions to define the 'residuals' directly, and omit the Lhs definition. This will be useful in specifying the GMM estimator, in which the orthogonality conditions may involve functions more complicated than a simple residual. For present purposes, then, let $\varepsilon_j(\boldsymbol{\beta})$ denote the residual defined above if you have included a **; Lhs** specification in your command. Otherwise, $\varepsilon_j(\boldsymbol{\beta})$ is $Fn_j(\boldsymbol{\beta})$. Let 't' index the T sample observations where needed. Let

$\boldsymbol{\varepsilon}_t(\boldsymbol{\beta})$ = the column vector of M residuals for observation t.

Let $\mathbf{E}_j(\boldsymbol{\beta})$ = the column vector of T residuals for equation j.

To specify the different estimation criteria, your command should appear as follows: (Since the criteria are all quadratic, the multiplication of each by ½ removes an inconvenient 2 from the derivatives. There is no other significance to this scaling.)

NOTE: Any of the following may include **; Wts = a weighting variable**, in which case, all sums of observations are computed using this weighting variable.

E25.2.1 OLS Estimation, Equation by Equation (NLOLS)

Include **; Sigma = I** in your command. In this case, the estimation criterion is

$$F = \tfrac{1}{2}\,\Sigma_i \boldsymbol{\varepsilon}_i{}'\boldsymbol{\varepsilon}_i = \tfrac{1}{2}\,\Sigma_i\Sigma_m \varepsilon_{im}{}^2 = \tfrac{1}{2}\,\Sigma_m \mathbf{E}_m(\boldsymbol{\beta})'\mathbf{E}_m(\boldsymbol{\beta}).$$

This is the sum of the sums of squares for each equation. If there are no cross equation equalities (i.e., no parameter name appears in more than one equation), then this is the same as using **NLSQ** once for each equation. If, in addition, all equations are linear, this will be the same as **REGRESS**, equation by equation.

```
NLSUR    ; Lhs    = ... the list of dependent variables
         ; Fn1    = ... the first equation
         ; Fn2    = ... the second equation
         ...
         ; FnM    = ... the last equation (up to 20 equations)
         ; Labels = ... a list of labels for the parameters
         ; Start  = ... the starting values for the iterations
         ; Sigma  = I  $
```

E25.2.2 Weighted Least Squares, Equation by Equation (NLWLS)

Include **; Sigma = D** in your command for a 'diagonal' disturbance covariance matrix. The estimation criterion is the weighted sum of squares,

$$F = \tfrac{1}{2}\,\Sigma_m (1/\sigma_m{}^2)\mathbf{E}_m(\boldsymbol{\beta})'\mathbf{E}_m(\boldsymbol{\beta}).$$

This is a groupwise heteroscedastic regression model. If there are no cross equation equalities, this will, once again, be the same as **NLSQ** equation by equation. If, in addition, the equations are all linear, this will be the same as model (S1,R0) in the **TSCS** model.

```
NLSUR    ; Lhs    = ... the list of dependent variables
         ; Fn1    = ... the first equation
         ; Fn2    = ... the second equation
         ...
         ; FnM    = ... the last equation (up to 20 equations)
         ; Labels = ... a list of labels for the parameters
         ; Start  = ... the starting values for the iterations
         ; Sigma  =  D  $
```

E25.2.3 IV Estimation, Equation by Equation (NL2SLS)

Include **; Sigma = I** and **; Inst = list of instrumental variables** in the command. Then,

$$F = ½ \Sigma_m \mathbf{E}_m(\beta)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}_m(\beta).$$

This is the sum of the 2SLS criteria for the *M* equations. If there are no cross equation equalities, this will be the same as using **NLSQ ; Inst = ...** once for each equation. If, in addition, all equations are linear, this is the same as 2SLS equation by equation.

```
NLSUR    ; Lhs    = ... the list of dependent variables
         ; Fn1    = ... the first equation
         ; Fn2    = ... the second equation
         ...
         ; FnM    = ... the last equation (up to 20 equations)
         ; Labels = ... a list of labels for the parameters
         ; Start  = ... the starting values for the iterations
         ; Sigma  = I
         ; Inst   = list of instrumental variables $
```

(There must be at least as many instrumental variables as there are parameters in the model.)

E25.2.4 Weighted IV Estimation, Equation by Equation (WNL2SLS)

Include **; Sigma = D** and **; Inst = list of instrumental variables** in the command. Then,

$$F = ½ \Sigma_m (1/\sigma_m^2)\mathbf{E}_m(\beta)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}_m(\beta).$$

This is the sum of the 2SLS criteria for the *M* equations, each weighted by its own variance. This is, once again, a groupwise heteroscedastic model. If there are no cross equation equalities, this would be the same as using **NLSQ ; Inst = ...** once for each equation. If all equations are linear, it is the same as 2SLS, equation by equation.

```
NLSUR    ; Lhs    = ... the list of dependent variables
         ; Fn1    = ... the first equation
         ; Fn2    = ... the second equation
         ...
         ; FnM    = ... the last equation (up to 20 equations)
         ; Labels = ... a list of labels for the parameters
         ; Start  = ... the starting values for the iterations
         ; Sigma  = D
         ; Inst   = list of instrumental variables $
```

E25.2.5 Nonlinear GLS Estimation (NLSURE)

The command is the one shown at the beginning of this section. I.e., no specification of *sigma*. (This is the default model.) The estimation criterion is

$$F = \tfrac{1}{2}\, \Sigma_m \Sigma_n \sigma^{mn} \mathbf{E}_m(\boldsymbol{\beta})' \mathbf{E}_n(\boldsymbol{\beta}),$$

where σ^{ij} is the ijth element of $\boldsymbol{\Sigma}^{-1}$. This is the nonlinear counterpart to the SURE estimator of Sections E24.2 and E24.3. If all equations are linear and there are no constraints, this will be the same as the SURE estimator in Section E24.2. If there are cross equation equality constraints, it is the MLE of Section E24.3.

```
NLSUR     ; Lhs    = ... the list of dependent variables
          ; Fn1    = ... the first equation
          ; Fn2    = ... the second equation
          ...
          ; FnM    = ... the last equation (up to 20 equations)
          ; Labels = ... a list of labels for the parameters
          ; Start  = ... the starting values for the iterations $
```

(There must be at least as many instrumental variables as there are parameters in the model.)

E25.2.6 Nonlinear IV Systems Estimation (NL3SLS)

Specify only **; Inst = list of instrumental variables** but do not specify **; Sigma**. In this case, the estimation rule is

$$F = \tfrac{1}{2}\, \Sigma_m \Sigma_n \sigma^{mn} \mathbf{E}_m(\boldsymbol{\beta})' \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \mathbf{E}_n(\boldsymbol{\beta}).$$

This is the nonlinear counterpart to the 3SLS estimator described in Section E24.4.

```
NLSUR     ; Lhs    = ... the list of dependent variables
          ; Fn1    = ... the first equation
          ; Fn2    = ... the second equation
          ...
          ; FnM    = ... the last equation (up to 20 equations)
          ; Labels = ... a list of labels for the parameters
          ; Start  = ... the starting values for the iterations
          ; Inst   = list of instrumental variables $
```

E25.2.7 GMM Estimation (GMM)

Include **; Inst = list of instrumental variables** and **; Pds = number for weighting matrix** in the command. The number of periods is needed to compute the weighting matrix for GMM estimation. The estimation criterion for GMM estimation is

$$F = ½ \Sigma_m \Sigma_n \mathbf{E}_m(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}_{mn}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}_n(\boldsymbol{\beta})$$

where $$\boldsymbol{\Omega}_{mn} = \mathbf{E}[(1/T)\Sigma_t \varepsilon_{tm}\varepsilon_{tn}\mathbf{z}_t\mathbf{z}_t'].$$

Note that the full $\boldsymbol{\Omega}$ contains M^2 blocks, none of which are assumed to be empty. This matrix must be estimated using the starting values.

```
NLSUR   ; Lhs    = ... the list of dependent variables
        ; Fn1    = ... the first equation
        ; Fn2    = ... the second equation
        ...
        ; FnM    = ... the last equation (up to 20 equations)
        ; Labels = ... a list of labels for the parameters
        ; Start  = ... the starting values for the iterations
        ; Inst   = list of instrumental variables
        ; Pds    = number of periods for Newey-West (may be 0) $
```

E25.2.8 Weighting Observations in Equation Systems

You may, if you wish, superimpose a weighting scheme on all of the preceding with

; Wts = weighting variable

This is the usual weighting procedure, but there is no assumption that the weights are observation specific variances; they may just be replication factors, or any other form of weight that you wish to apply. In any event, weights are still scaled to sum to N unless you suppress this with **; Wts = ...,Noscale**. Note, however, that the way that weights will be applied depends on the estimation criterion. In all cases, the weight is applied to the term in a sum. Thus, in NLOLS, with **; Wts** in use, the criterion becomes $½ \Sigma_i w_i \Sigma_m \varepsilon_{im}^2$, whereas in the various IV procedures, which are not simple sums of terms such as this, the weights are applied to the summations in the moment matrices. To consider an example, let **W** denote a diagonal matrix with your weights on the diagonal. In NL2SLS, the estimation criterion becomes

$$F = ½\Sigma_m \mathbf{E}_m(\boldsymbol{\beta})'\mathbf{WZ}(\mathbf{Z}'\mathbf{WZ})^{-1}\mathbf{Z}'\mathbf{WE}_m(\boldsymbol{\beta}).$$

(Of course, we do not actually create the diagonal matrix internally.) The other estimators are constructed likewise.

E25.2.9 Model Specifications for the NLSUR Procedure

This is the full list of general specifications that are applicable to this model estimator.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, BFGS is the default.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays **S** in the output.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.
; Fix = list fixes the named parameters at the starting values.

Note that with **; Maxit = 0**, this is not necessarily an LM test at all, since the disturbances in these models are not assumed to be normally distributed, and, even if they were, the estimation criteria listed above are not the log likelihood functions in most cases. As such, **; Maxit = 0** is best viewed as a useful descriptive device that allows you to examine your model for a fixed set of parameters. (Note, as well, that **; Maxit = 0** is the same as **; Fix All**, but provides more useful information.)

E25.3 Output and Saved Results from NLSUR

The output from this procedure is largely the same as that from **NLSQ**. The saved results are:

Matrices:	*b*	= the estimated parameter vector
	varb	= asymptotic covariance matrix
	sigma	= estimate of **Σ**
Scalars:	*kreg*	= the number of parameters in the model
	nreg	= the number of observations
	logl	= the value of the criterion function

Last Model: The labels are those in your **; Labels** list

Last Function: None

Results from the procedure will include the initial table reporting the procedure used, the number of iterations completed, and so on. *The 'log likelihood' reported is actually the minimized criterion function, not a true log likelihood.* This is followed by the table of estimates, estimated standard errors, and so on. If you have specified a **; Lhs** list, an additional table will give a listing of the Lhs variables, means and standard deviations, and sum of squared residuals and an R^2 for each equation.

NOTE: This R^2 is not bounded in [0,1] because the fitting criterion is not linear ordinary least squares with a constant term.

When you provide a set of Lhs variables for an NLSURE model, the diagnostic output will also include McElroy's R^2 measure for the system. This is computed as

$$R_m^2 = \sum_s \sum_t \sigma^{st} (1/n) \sum_{i=1}^{n} (y_{is} - \bar{y}_s)(y_{it} - \bar{y}_t) = \text{tr}(\mathbf{S}^{-1}\mathbf{V}_{\mathbf{y}})$$

where $\mathbf{V}_y$ is the sample covariance matrix for the Lhs variables.

There are no residuals or fitted values from this procedure. The parameters are retrievable, however, so you can construct these with **CREATE**.

E25.4 Application

To illustrate use of this estimator, we will estimate a system of linear equations. The procedure does not differentiate between linear and nonlinear systems, so this illustrates the full procedure. We have in hand the first 100 observations in the Grunfeld data used in Chapter E15.

```
SAMPLE      ; 1-100$
CREATE      ; igm= i        ; fgm= f        ; cgm = c
            ; ich  = i[+20] ; fch  = f[+20] ; cch  = c[+20]
            ; ige  = i[+40] ; fge  = f[+40] ; cge  = c[+40]
            ; iwe = i[+60] ; few = f[+60] ; cwe  = c[+60]
            ; ius  = i[+80] ; fus  = f[+80] ; cus  = c[+80] $
NAMELIST    ; xgm = one,fgm,cgm
            ; xch  = one,fch,cch
            ; xge  = one,fge,cge
            ; xwe = one,fwe,cwe
            ; xus  = one,fus,cus
            ; y = igm,ich,ige,iwe,ius $
SAMPLE      ; 1-20 $
NLSUR       ; Lhs = y
            ; Fn1 = b0'xgm ; Fn2=b0'xch ; Fn3=b0'xge ; Fn4=b0'xwe ; Fn5=b0'xus
            ; Start = 0,0,0 ; Labels = b0,b1,b2 ; Sigma = I $
NLSUR       ; Lhs = y
            ; Fn1 = b0'xgm ; Fn2=b0'xch ; Fn3=b0'xge ; Fn4=b0'xwe ; Fn5=b0'xus
            ; Start = 0,0,0 ; Labels = b0,b1,b2 ; Sigma = D $
NLSUR       ; Lhs = y
            ; Fn1 = b0'xgm ; Fn2=b0'xch ; Fn3=b0'xge ; Fn4=b0'xwe ; Fn5=b0'xus
            ; Start = 0,0,0 ; Labels = b0,b1,b2 $
```

The equations are linear, with cross equation equality restrictions. The three commands will reestimate models (S0,R0), (S1,R0), and (S2,R0), from Section E15.3, respectively. Note that, as discussed in the next section, the estimated standard errors differ from those given previously. Moreover, in the more complex models, the parameter estimates differ slightly as well. This is due, in part to the convergence rule used for **NLSURE**, which does not use the actual second derivatives and, thus, does not find the exact minimizer of the criterion, as **TSCS** does and, second, because **NLSUR** does not necessarily converge to exactly the same estimate of $\mathbf{\Sigma}$.

The first set of results is equivalent to pooled least squares. The OLS results are shown as well. Note that although the parameter estimates are identical, the standard errors are noticeably different. The reason for this difference is the method of computation of the covariance matrix in the NLSUR case. The routine minimizes

$$F = \Sigma_i \mathbf{e}(\boldsymbol{\beta})'\mathbf{e}(\boldsymbol{\beta})/2 = (1/2)\Sigma_i \Sigma_m e_{im}(\boldsymbol{\beta})^2$$

where each derivative vector is five (for this case, M in general) by one. The covariance matrix used is then based on

$$\text{BHHH} = \{\Sigma_i \Sigma_m [e_{im}(\boldsymbol{\beta})]^2 \mathbf{x}_{im}\mathbf{x}_{im}'\}^{-1}.$$

This matrix will be approximately equal to $(1/\sigma^2)(\mathbf{X'X})^{-1}$ under the assumption that the disturbances have the common variance σ^2 and that they are independent of the pseudoregressors. Since it is known from the specification that you have specified **I** as the covariance matrix, this matrix is then scaled by the square of an estimator of this common variance; in this case that will be $(\sigma^2)^2$. In large samples, this will give the same answer as the more familiar estimator. But, in a finite sample, such as the one of 20 observations here, the results will differ noticeably.

Note, as well, the diagnostic about unusually fast convergence that appears with the output. The reason that this estimation converged so quickly is that the equations are linear. The routine has not examined the equation specifications to discover this, so the warning is the routine one that shows up when a nonlinear optimization problem reaches convergence more quickly than expected.

```
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    831549.4

-----------------------------------------------------------------------------
Nonlinear minimization over  5 equations.
Dependent variable             MultEqns
Log likelihood function    831549.37585
Estimation based on N =     20, K =   0
Inf.Cr.AIC  =********* AIC/N = ********
Model estimated: Jun 20, 2011, 19:41:38
Disturbances are uncorrelated
Pooled variance is   16630.9875171
Covariance matrix used is s-sqrd*I
Number of iterations over S is     0
Used equation by equation nonlinear OLS  .
McElroy R-squared for the system =  .99995
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|   -63.6112       53.97581    -1.18  .2386   -169.4018   42.1795
      B1|     .11844***      .01487     7.96  .0000      .08928    .14759
      B2|     .25648***      .04073     6.30  .0000      .17666    .33630
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+-----------------------------------------------------------------------+
| __ Equation   Mean of LHS   S.D. of LHS   R-squared   Sum of squares |
|  1 IGM         608.020000    309.574628     .865383  .2451222113D+06 |
|  2 ICH         410.475000    125.399429   -1.402341  .7177601842D+06 |
|  3 IGE         102.290000     48.584499  -12.685050  .6137555637D+06 |
|  4 IWE          86.123500     42.725555     .111538  .3081541124D+05 |
|  5 IUS          61.802500     15.166932  -11.731514  .5564538125D+05 |
| Note, R-squared can be negative if not using unconstrained OLS.      |
+-----------------------------------------------------------------------+
```

The second set of results is based on an assumption that the system is groupwise heteroscedastic. In this case, the results differ from the results of the TSCS approach only in that there is a slight difference in the diagonal covariance matrix used in estimation. Note, to make these comparable, the TSCS procedure must be iterated. The results are shown below.

```
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    831549.4

-----------------------------------------------------------------------------
Nonlinear minimization over  5 equations.
Dependent variable              MultEqns
Log likelihood function         44.71621
Estimation based on N =     20, K =   0
Inf.Cr.AIC  =  -89.432 AIC/N =   -4.472
Model estimated: Jun 20, 2011, 19:41:38
Disturbances are uncorrelated
Pooled variance is      12.8662025
Covariance matrix used is Diagonal
Number of iterations over S is     0
Used equation by equation nonlinear OLS  .
McElroy R-squared for the system =  .45187
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|   -38.0503*     22.03449    -1.73  .0842    -81.2371    5.1365
      B1|     .12341***     .00574    21.52  .0000      .11217     .13465
      B2|     .18989***     .02099     9.05  .0000      .14875     .23102
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+------------------------------------------------------------------------+
| __ Equation    Mean of LHS   S.D. of LHS   R-squared   Sum of squares |
|  1 IGM          608.020000    309.574628     .786930  .3879781258D+06 |
|  2 ICH          410.475000    125.399429   -1.105324  .6290188463D+06 |
|  3 IGE          102.290000     48.584499  -13.745632  .6613212375D+06 |
|  4 IWE           86.123500     42.725555     .671647  .1138859314D+05 |
|  5 IUS           61.802500     15.166932   -6.198258  .3146128437D+05 |
| Note, R-squared can be negative if not using unconstrained OLS.       |
+------------------------------------------------------------------------+
```

The final set of results corresponds to the fully general nonlinear seemingly unrelated regressions model.

```
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    831549.4
```

```
-----------------------------------------------------------------------------
Nonlinear minimization over  5 equations.
Dependent variable              MultEqns
Log likelihood function         48.57176
Estimation based on N =      20, K =   0
Inf.Cr.AIC  =  -97.144 AIC/N =   -4.857
Model estimated: Jun 20, 2011, 19:41:38
Disturbances are   correlated
Pooled variance is      17.2452502
Covariance matrix used is (1/N)E'E
Number of iterations over S is     0
Used multiple    equation nonlinear GLS  .
McElroy R-squared for the system =  .80507
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|   -54.1071***     5.21198   -10.38  .0000    -64.3224   -43.8918
      B1|      .11149***     .00551    20.25  .0000      .10070     .12228
      B2|      .25113***     .01259    19.95  .0000      .22645     .27580
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
+------------------------------------------------------------------------+
| __ Equation    Mean of LHS   S.D. of LHS   R-squared   Sum of squares |
|  1 IGM          608.020000    309.574628     .853176  .2673502576D+06 |
|  2 ICH          410.475000    125.399429   -1.540064  .7589083079D+06 |
|  3 IGE          102.290000     48.584499  -11.661628  .5678565035D+06 |
|  4 IWE           86.123500     42.725555     .273733  .2518984879D+05 |
|  5 IUS           61.802500     15.166932  -12.587489  .5938657396D+05 |
| Note, R-squared can be negative if not using unconstrained OLS.        |
+------------------------------------------------------------------------+
```

E25.5 Technical Details

The various estimation criteria listed above will replicate other settings when certain restrictions are in place. For example, NLOLS with no cross equation restrictions in place and linear equations is the same as TSCS. In these cases, the estimator will usually produce the same parameter estimates, but may produce slightly, or in a small sample, noticeably different standard errors. The reason is that the linear estimators (**REGRESS, TSCS, 2SLS**) use (in principle) the actual second derivatives matrices of their estimation criteria. But, **NLSUR** always uses the outer products of the first derivatives to accumulate its estimate of the asymptotic covariance matrix. These will normally be reasonably close to each other, but, as noted, in a finite sample, they can differ.

For estimation of the systems in which $\boldsymbol{\Sigma}$ is not $\sigma^2\mathbf{I}$, we use a straightforward two level iteration. The procedure is as follows:

Step 1. At entry, set $\boldsymbol{\Sigma}$ either to $\mathbf{I}$ or to the matrix you supply with **; Sigma = name**.

Step 2. Obtain the parameter estimates conditioned on this estimate of $\boldsymbol{\Sigma}$.

Step 3. Use the parameter estimates to recompute $\boldsymbol{\Sigma}$.

Step 4. Assess convergence based on the log determinant of the estimated $\boldsymbol{\Sigma}$. If the change is less than 10^{-4}, exit. If there are more than 20 iterations on $\boldsymbol{\Sigma}$, exit on maximum iterations. Else, return to Step 2.

E26: Models for Binary Choice

E26.1 Introduction

We define models in which the response variable being described is inherently discrete as qualitative response (QR) models. This and the next several chapters will describe *LIMDEP*'s qualitative dependent variable model estimators. The simplest of these are the binomial choice models, which are the subject of this chapter and Chapters E27-E29. This will be followed by the progressively more intricate formulations such as bivariate and multivariate probit, multinomial logit, ordered choice and models for count data. *LIMDEP* supports a large variety of models and extensions for the analysis of binary choice. The parametric model formulations, probit, logit, extreme value (complementary log log) etc. are treated in Chapter E27. Panel data models for binary choice appear in Chapters E30 and E31. Semi- and nonparametric models appear in Chapter E32.

There are numerous references for practitioners using the binary choice modeling framework. Four which are widely used are Maddala (1983), Greene (2011), Long (1997) and DeMaris (2004). Another recent source is for binary choice modeling is Greene and Hensher (2010, Chapters 1-4).

E26.2 Modeling Binary Choice

A binomial response may be the outcome of a decision or the response to a question in a survey. Consider, for example, survey data which indicate political party choice, mode of transportation, occupation, or choice of location. We model these in terms of probability distributions defined over the set of outcomes. There are a number of interpretations of an underlying data generating process that produce the binary choice models we consider here. All of them are consistent with the models that *LIMDEP* estimates, but the exact interpretation is a function of the modeling framework.

E26.2.1 Underlying Processes

Consider a process with two possible outcomes indicated by a *dependent variable*, y, labeled for convenience, $y = 0$ and $y = 1$. We assume, as well, that there is a set of measurable *covariates*, x, which will be used to help explain the occurrence of one outcome or the other. Most models of binary choice set up in this fashion will be based upon an *index function*, $\boldsymbol{\beta}'\mathbf{x}$, where $\boldsymbol{\beta}$ is a vector of parameters to be estimated. The modeling of discrete, binary choice in these terms, is typically done in one of the following frameworks:

Random Utility Approach

The respondent derives utility

$$U_0 = \boldsymbol{\beta}_0'\mathbf{x} + \varepsilon_0 \text{ from choice 0, and } U_1 = \boldsymbol{\beta}_1'\mathbf{x} + \varepsilon_1 \text{ from choice 1,}$$

in which ε_0 and ε_1 are the individual specific, random components of the individual's utility that are unaccounted for by the measured covariates, **x**. The choice of alternative 1 reveals that $U_1 > U_0$, or that

$$\varepsilon_0 - \varepsilon_1 < \boldsymbol{\beta}_1'\mathbf{x} - \boldsymbol{\beta}_0'\mathbf{x}.$$

Let $\varepsilon = \varepsilon_0 - \varepsilon_1$ and let $\boldsymbol{\beta}'\mathbf{x}$ represent the difference on the right hand side of the inequality – **x** is the union of the two sets of covariates, and $\boldsymbol{\beta}$ is constructed from the two parameter vectors with zeros in the appropriate locations if necessary. Then, the binary choice model applies to the probability that $\varepsilon \leq \boldsymbol{\beta}'\mathbf{x}$, which is the familiar sort of model shown in the next paragraph. This is a convenient way to view migration behavior and survey responses to questions about economic issues.

Latent Regression Approach

A latent regression is specified as

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon.$$

The observed counterpart to y^* is

$$y = 1 \text{ if and only if } y^* > 0.$$

This is the basis for most of the binary choice models in econometrics, and is described in further detail below. It is the same model as the reduced form in the previous paragraph. Threshold models, such as labor supply and reservation wages lend themselves to this approach.

Conditional Mean Function Approach

We assume that y is a binary variable, taking values 0 and 1, and formulate a priori that $\text{Prob}[y=1] = F(\boldsymbol{\beta}'\mathbf{x})$, where F is any function of the index that satisfies the axioms of probability,

$$0 \leq F(\boldsymbol{\beta}'\mathbf{x}) \leq 1$$

$$F'(\boldsymbol{\beta}'\mathbf{x}) \geq 0,$$

$$\lim_{z\downarrow -\infty} F(z) = 0, \lim_{z\uparrow +\infty} F(z) = 1.$$

It follows that, $$F(\boldsymbol{\beta}'\mathbf{x}) = 0 \times \text{Prob}[y = 0 \mid \mathbf{x}] + 1 \times \text{Prob}[y = 1 \mid \mathbf{x}]$$

is the conditional mean function for the observed binary y. This may be treated as a nonlinear regression or as a binary choice model amenable to maximum likelihood estimation. This is a useful departure point for less parametric approaches to binary choice modeling.

E26.2.2 Modeling Approaches

This and the next several chapters document three approaches to formulating the binary choice models described above:

Parametric Models – Probit, Logit, Extreme Value, Gompertz, Burr, Arctangent

Most of the material below (and the received literature) focuses on models in which the full functional form, including the probability distribution, are defined a priori. Thus, the probit model which forms the basis of most of the results in econometrics, is based on a latent regression model in which the disturbances are assumed to have a normal distribution. The logit model, in contrast, can be construed as a random utility model in which it is assumed that the random parts of the utility functions are distributed as independent extreme value. The complementary log log model arises as the natural distribution in a setting of counts of occurrences (such as part failures or numbers of arrivals of messages at a receiving center) in which the analyst is interest in modeling not the number of occurrences, but whether none or any events have occurred. The Burr distribution allows asymmetry in the logit framework. Finally, the Arctangent model provides a flexible, interesting functional form.

Semiparametric Models – Maximum Score, Semiparametric Analysis

A semiparametric approach to modeling the binary choice steps back one level from the previous model in that the specific distributional assumption is dropped, while the covariation (index function) nature of the model is retained. Thus, the semiparametric approach analyzes the common characteristics of the observed data which would arise regardless of the specific distribution assumed. Thus, the semiparametric approach is essentially the conditional mean framework without the specific distribution assumed. For the models that are supported in *LIMDEP*, MSCORE and Klein and Spady's framework, it is assumed only that $F(\boldsymbol{\beta}'\mathbf{x})$ exists and is a smooth continuous function of its argument which satisfies the axioms of probability. The semiparametric approach is more general (and more robust) than the parametric approach, but it provides the analyst far less flexibility in terms of the types of analysis of the data that may be performed. In a general sense, the gain to formulating the parametric model is the additional precision with which statements about the data generating process may be made. Hypothesis tests, model extensions, and analysis of, e.g., interactions such as marginal effects, are difficult or impossible in semiparametric settings.

Nonparametric Analysis – NPREG

The nonparametric approach, as its name suggests, drops the formal modeling framework. It is largely a bivariate modeling approach in which little more is assumed than that the probability that y equals one depends on some x. (It can be extended to a latent regression, but this requires prior specification and estimation, at least up to scale, of a parameter vector.) The nonparametric approach to analysis of discrete choice is done in *LIMDEP* with a kernel density (largely based on the computation of histograms) and with graphs of the implied relationship. Nonparametric analysis is, by construction, the most general and robust of the techniques we consider, but, as a consequence, the least precise. The statements that can be made about the underlying DGP in the nonparametric framework are, of necessity, very broad, and usually provide little more than a crude overall characterization of the relationship between a y and an x.

E26.2.3 The Linear Probability Model

One approach to modeling binary choice has been to ignore the special nature of the dependent variable, and use conventional least squares. The resulting model,

$$\text{Prob}[y_i = 1] = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$$

has been called the linear probability model (LPM). The LPM is known to have several problems, most importantly that the model cannot be made to satisfy the axioms of probably independently of the particular data set in use. Some authors have documented approaches to forcing the LPM on the data, e.g., Fomby, et al., (1984), Long (1997) and Angrist and Pischke (2009). These computations can easily be done with the other parts of *LIMDEP*, but will not be pursued here.

E26.3 Grouped and Individual Data for Binary Choice Models

There are two types of data which may be analyzed. We say that the data are *individual* if the measurement of the dependent variable is physically discrete, consisting of individual responses. The familiar case of the probit model with measured 0/1 responses is an example. The data are *grouped* if the underlying model is discrete but the observed dependent variable is a proportion. In the probit setting, this arises commonly in bioassay. A number of respondents have the same values of the independent variables, and the observed dependent variable is the proportion of them with individual responses equal to one. Voting proportions are a common application from political science.

All of the qualitative response models estimated by *LIMDEP* can be estimated with either individual or grouped data. You do not have to inform the program which type you are using; if necessary, the data are inspected to determine which applies. The differences arise only in the way starting values are computed and, occasionally, in the way the output should be interpreted. Cases sometimes arise in which grouped data contain cells which are empty (proportion is zero) or full (proportion is one). This does not affect maximum likelihood estimation and is handled internally in obtaining the starting values. No special attention has to be paid to these cells in assembling the data set.

E26.4 Variance Normalization

In the latent regression formulation of the model, the observed data are generated by the underlying process

$$y = 1 \text{ if and only if } \boldsymbol{\beta}'\mathbf{x} + \varepsilon > 0.$$

The random variable, ε, is assumed to have a zero mean (which is a simple normalization if the model contains a constant term). The variance is left unspecified. The data contain no information about the variance of ε. Let σ denote the standard deviation of ε. The same model and data arise if the model is written as

$$y = 1 \text{ if and only if } (\boldsymbol{\beta}/\sigma)'\mathbf{x} + \varepsilon/\sigma > 0.$$

which is equivalent to

$$y = 1 \text{ if and only if } \boldsymbol{\gamma}'\mathbf{x} + w > 0.$$

where the variance of w equals one. Since only the sign of y is observed, no information about overall scaling is contained in the data. Therefore, the parameter σ is not estimable; it is assumed with no loss of generality to equal one. (In some treatments (Horowitz (1993)), the constant term in β is assumed to equal one, instead, in which case, the 'constant' in the model is an estimator of $1/\sigma$. This is simply an alternative normalization of the parameter vector, not a substantive change in the model.)

E26.5 The Constant Term in Index Function Models

A question that sometimes arises is whether the binary choice model should contain a constant term. The answer is yes, unless the underlying structure of your model specifically dictates that none be included. There are a number of useful features of the parametric models that will be subverted if you do not include a constant term in your model:

- Familiar fit measures will be distorted. Indeed, omitting the constant term can seriously degrade the fit of a model, and will never improve it.

- Certain useful test statistics, such as the overall test for the joint significance of the coefficients, may be rendered noncomputable if you omit the constant term.

- Some properties of the binary choice models, such as their ability to reproduce the average outcome (sample proportion) will be lost.

Forcing the constant term to be zero is a linear restriction on the coefficient vector. Like any other linear restriction, if imposed improperly, it will induce biases in the remaining coefficients. (Orthogonality with the other independent variables is not a salvation here. Thus, putting variables in mean deviation form does not remove the constant term from the model as it would in the linear regression case.)

E27: Probit and Logit Models: Estimation

E27.1 Introduction

We define models in which the response variable being described is inherently discrete as qualitative response (QR) models. This and the next several chapters will describe *LIMDEP*'s qualitative dependent variable model estimators. The simplest of these are the binomial choice models, which are the subject of this chapter and Chapters E28 and E29. This will be followed by the progressively more intricate formulations such as bivariate and multivariate probit, multinomial logit, ordered choice and models for count data.

LIMDEP supports a large variety of models and extensions for the analysis of binary choice. The parametric model formulations, probit, logit, extreme value (complementary log log) etc. are treated in this chapter. Several model extensions such as models with endogenous variables, and sample selection, are treated in Chapter E29. Panel data models for binary choice appear in Chapters E30 and E31. Semi- and nonparametric models are documented in Chapter E32.

E27.2 Parametric Models for Binary Choice

LIMDEP supports six parametric functional forms for binary choice models. The basic model commands for the six models are:

$$\left.\begin{array}{c}\textbf{PROBIT}\\ \textbf{LOGIT}\\ \textbf{ARCTANGENT}\\ \textbf{GOMPERTZ}\\ \textbf{COMPLOGLOG}\\ \textbf{BURR}\end{array}\right\}\ \textbf{; Lhs = dependent variable ; Rhs = regressors \$}$$

Data on the dependent variable may be either individual or proportions for all six cases.

E27.2.1 Functional Forms for Parametric Models

Six specific parametric model formulations are provided as internal procedures in *LIMDEP* for binary choice models. The probabilities and density functions are as follows:

Probit

$$F = \int_{-\infty}^{\beta'\mathbf{x}_i} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt \;=\; \Phi(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \phi(\boldsymbol{\beta}'\mathbf{x}_i)$$

Logit

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} \;=\; \Lambda(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

Arctangent

$F = (2/\pi)\ \arctan(\exp(\boldsymbol{\beta}'\mathbf{x}_i)),\qquad f = (2/\pi)\{1/[1+(\exp(\boldsymbol{\beta}'\mathbf{x}_i))^2]\}$

Complementary log log

$F = 1 - \exp(-\exp(\boldsymbol{\beta}'\mathbf{x}_i)) = \mathrm{C}(\boldsymbol{\beta}'\mathbf{x}_i),\qquad f = \exp(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \mathrm{C}(\boldsymbol{\beta}'\mathbf{x}_i)]$

Gompertz, or type 1 extreme value

$F = \exp(-\exp(-\boldsymbol{\beta}'\mathbf{x}_i)) = \mathrm{G}(\boldsymbol{\beta}'\mathbf{x}_i),\qquad f = \exp(-\boldsymbol{\beta}'\mathbf{x}_i)\mathrm{G}(\boldsymbol{\beta}'\mathbf{x}_i)$

Burr or Scobit

$$F = \left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right]^{\gamma} = [\Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]^{\gamma},\ \gamma > 0,\qquad f = \gamma[\Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]^{\gamma}\ [1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

None of these is obviously best for any situation. (The advantage of the probit model becomes overwhelming when the binary choice model is part of a more elaborate, possibly multiple equation structure.) The complementary log log distribution does arise naturally from a complete censoring of the positive values of the Poisson regression model. The first two listed above are symmetric while the latter four are not. The Burr distribution is an extension of the logistic model. The logistic model is the special case of $\gamma = 1$. Plots of the CDFs and PDFs appear below. Since the shape of the Burr distribution depends on γ, we have chosen an intermediate value of 1.5 for purposes of illustration. The program used to produce the figures is shown below as well.

In the upper figure, the two symmetric distributions, probit and logit, cross at zero in the center of the figure. The complementary log log is the higher one; it assigns a smaller probability to the right tail. As the figure at the right shows, the other asymmetric distributions assign higher probability to the right tail. The same effects can be seen in the lower figures, which plot the densities.

```
SAMPLE      ; 1-101 $
CREATE      ; z = Trn(-3,.06)
            ; probit = Phi(z) ; logit  = Lgp(z)
            ; cloglog = 1 - Exp(-Exp(z))
            ; gompit = Exp(-Exp(-z))
            ; burr = Logit ^ 1.5
            ; arctan = 2/pi*atn(Exp(z))$
CREATE      ; dprobit =  N01(z)
            ; dlogit  = logit*(1-logit)
            ; dcloglog = Exp(z) * Exp(-Exp(z))
            ; dgompit = -Log(gompit)*gompit
            ; dburr = 1.5*burr*(1-logit)
            ; darctan = 2/pi*Exp(z)/(1+Exp(z)*Exp(z))$
```

```
PLOT    ; Lhs = z
        ; Rhs = probit,logit,cloglog,arctan,gompit
        ; Endpoints = -3,3 ; Fill ; Bars = .5 ; Spikes = 0
        ; Yaxis = CDF
        ; Title = Probability Functions  $
PLOT    ; Lhs = z
        ; Rhs = logit,cloglog,gompit,burr
        ; Endpoints = -3,3 ; Fill ; Bars = .5 ; Spikes = 0
        ; Yaxis = CDF
        ; Title = Asymmetric Probability Functions vs. Logit $
PLOT    ; Lhs = z
        ; Rhs = dprobit,dlogit,dcloglog,darctan,dgompit
        ; Endpoints = -3,3 ; Fill ; Spikes = 0
        ; Yaxis = PDF
        ; Title = Density Functions $
PLOT    ; Lhs = z
        ; Rhs = dlogit,dcloglog,dgompit,dburr
        ; Endpoints = -3,3 ; Fill ; Spikes = 0
        ; Yaxis = PDF
        ; Title = Asymmetric Density Functions vs. Logit $
```

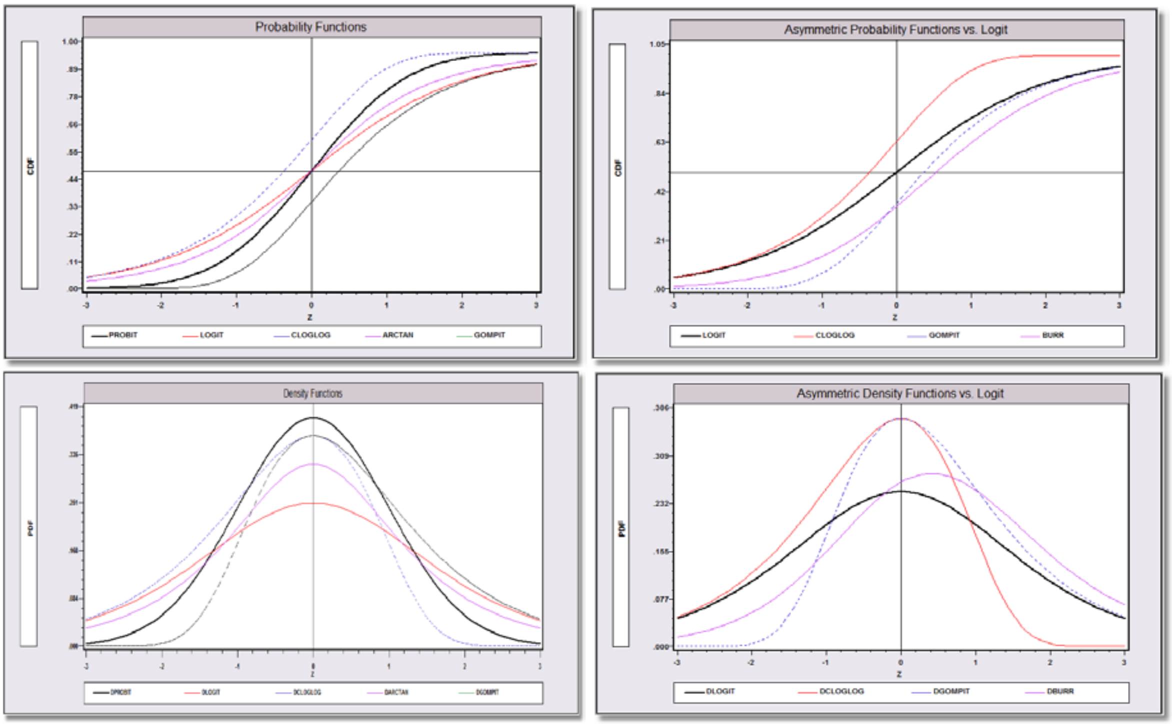


Figure E27.1 Densities and CDFs for Binary Choice Models

E27.2.2 Data Used in Estimation of Parametric Models

The Dependent Variable – Individual or Proportions

Data on the dependent variable for these models may be individual or grouped. The estimation program will check internally, and adjust accordingly where necessary. The log likelihood function is the same for either case. The only special consideration concerns the computation of the starting values for the iterations. If you do not provide your own starting values, they are determined for the individual data case by simple least squares. The OLS estimator is not useful in itself, but it does help to adjust the scale of the coefficient vector for the first iteration. For the grouped data case, however, the initial values are determined by the minimum chi squared, weighted least squares computation. Since this will generally involve logarithms or other transformations which become noncomputable at zero or one, they are not computed for individual data.

Problems with the Independent Variables

There is a special consideration for the independent variables in the binary choice model. If a variable x_k is such that the range of x_k can be divided into two parts and within the two parts, the value of the dependent variable is always the same, then this variable becomes a perfect predictor for the model. The estimator will break down, sometimes by iterating endlessly as the coefficient vector drifts to extreme values. The following program illustrates the effect: The variable *z* is positive when *y* equals one and negative when it equals zero. The estimator exited after 100 iterations, but appears actually to have converged normally – note the derivatives are extremely small. But, a probit model should take less than 10 iterations. Second, note that the log likelihood function is essentially zero, indicative of a perfect fit. The coefficient on *z* is nonsensical, and the standard errors are essentially infinite. All are indicators of a bad data set and/or model. The extreme (perfect) values for the fit measures on the next page underscore the point.

```
SAMPLE      ; 1-100 $
CALC        ; Ran(12345) $
CREATE      ; x = Rnn(0,1)
            ; d = Rnu(0,1) > .5 $
CREATE      ; y = (-.5 + x + d + Rnn(0,1)) > 0 $
CREATE      ; If(y = 1)z =  Rnu(0,1)
            ; If(y = 0)z = -Rnu(0,1) $
PROBIT      ; Lhs = y ; Rhs = one,x,z
            ; Output = 4 ; Summarize $
```

```
Nonlinear Estimation of Model Parameters
Method=NEWTON; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator= 100
Start values:   .45710D+00   .95098D-01   .68943D+00
1st derivs.     .30046D+02  -.22180D+02  -.27947D+02
Parameters:     .45710D+00   .95098D-01   .68943D+00
Itr  1 F=  .5148D+02 gtHg=  .7451D+01 chg.F=  .5148D+02 max|db|=  .2048D+01
```

```
1st derivs.      .75382D+01  -.74206D+01  -.87858D+01
Parameters:      .62796D-01   .28522D+00   .21013D+01
Itr  2 F=  .1817D+02 gtHg=  .3623D+01 chg.F=  .3331D+02 max|db|=  .2973D+01
1st derivs.      .22778D+01  -.25169D+01  -.34737D+01
Parameters:     -.12389D+00   .45652D+00   .32904D+01
 (Iterations 3 - 98 omitted)
Itr 99 F=  .2155D-11 gtHg=  .2231D-06 chg.F=  .3664D-13 max|db|=  .8482D-03
1st derivs.     -.13675D-11  -.73224D-12  -.10213D-11
Parameters:     -.98485D+00   .14753D+00   .14438D+03
Itr100 F=  .2119D-11 gtHg=  .2204D-06 chg.F=  .3553D-13 max|db|=  .8477D-03
Maximum of   100 iterations. Exit iterations with status=1.
Function=  .51483973128D+02, at entry,  .20847767956D-11 at exit
------------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                 Y
Log likelihood function       .00000
Restricted log likelihood  -69.13461
Chi squared [   2 d.f.]    138.26922
Significance level            .00000
McFadden Pseudo R-squared  1.0000000
Estimation based on N =    100, K =   3
Inf.Cr.AIC  =    6.000 AIC/N =    .060
--------+---------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+---------------------------------------------------------------------
        |Index function for probability
Constant|    -.98505      148462.2      .00 1.0000 ***********  ***********
       X|     .14766      120032.6      .00 1.0000 ***********  ***********
       Z|    144.424      345728.4      .00  .9997 -677470.698  677759.546
--------+---------------------------------------------------------------------
+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable Y          |
+----------------------------------------+
|                  Y=0       Y=1     Total|
| Proportions  .53000    .47000   1.00000|
| Sample Size      53        47       100|
+----------------------------------------+
| Log Likelihood Functions for BC Model  |
|              P=0.50    P=N1/N   P=Model|
| LogL =       -69.31    -69.13       .00|
+----------------------------------------+
| Fit Measures based on Log Likelihood   |
| McFadden = 1-(L/L0)           =  1.00000|
| Estrella = 1-(L/L0)^(-2L0/n) =  1.00000|
| R-squared (ML)                =   .74910|
| Akaike Information Crit.      =   .06000|
| Schwartz Information Crit.    =   .13816|
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                         =  1.00000|
| Ben Akiva and Lerman          =  1.00000|
| Veall and Zimmerman           =  1.00000|
| Cramer                        =  1.00000|
+----------------------------------------+
```

```
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |      53 ( 53.0%)|       0 (    .0%)|      53 ( 53.0%)|
|  1   |       0 (    .0%)|      47 ( 47.0%)|      47 ( 47.0%)|
+------+----------------+----------------+----------------+
|Total |      53 ( 53.0%)|      47 ( 47.0%)|     100 (100.0%)|
+------+----------------+----------------+----------------+
+----------------------------------------------------------+
|Crosstab for Binary Choice Model.  Predicted probability |
|vs. actual outcome. Entry = Sum[Y(i,j)*Prob(i,m)] 0,1.   |
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|      Predicted Probability      |                |
|Value |    Prob(y=0)        Prob(y=1)   | Total Actual   |
+------+----------------+----------------+----------------+
| y=0  |      52 ( 52.0%)|       0 (    .0%)|      53 ( 52.0%)|
| y=1  |       0 (    .0%)|      46 ( 46.0%)|      47 ( 46.0%)|
+------+----------------+----------------+----------------+
|Total |      53 ( 52.0%)|      46 ( 46.0%)|     100 ( 98.0%)|
+------+----------------+----------------+----------------+
```

In general, for every Rhs variable, *x*, the minimum *x* for which *y* is one must be less than the maximum *x* for which *y* is zero, and the minimum *x* for which *y* is zero must be less than the maximum *x* for which *y* is one. If either condition fails, the estimator will break down. This is a more subtle, and sometimes less obvious failure of the estimator. Unfortunately, it does not lead to a singularity and the eventual appearance of collinearity in the Hessian. You might observe what appears to be convergence of the estimator on a set of parameter estimates and standard errors which might look reasonable. The main indication of this condition would be an excessive number of iterations – the probit model will usually reach convergence in only a handful of iterations – and a suspiciously large standard error is reported for the coefficient on the offending variable, as in the preceding example. You can check for this condition with the command

CALC ; Chk (names of independent variables to check,
name of dependent variable) $

The offending variable in the previous example would be tagged by this check;

CALC ; Chk(x,z,y) $

```
Error   462: 0/1 choice model is inestimable. Bad variable = Z
Error   463: Its values predict 1[Y       = 1] perfectly.
```

This computation will issue warnings when the condition is found in any of the variables listed. (Some computer programs will check for this condition automatically, and drop the offending variable from the model. In keeping with *LIMDEP*'s general approach to modeling, this program does not automatically make functional form decisions. This is up to the analyst.)

Dummy Variables with Empty Cells

A problem similar to the one noted above arises when your model includes a dummy variable that has no observations in one of the other cells of the dependent variable. An example appears in Greene (1993, p. 673) in which the Lhs variable is always zero when the variable ‘Southwest’ is zero. Professor Terry Seaks has used this example to examine a number of econometrics programs. He found that no program which did not specifically check for the failure – only one did – could detect the failure in some other way. All iterated to apparent convergence, though with very different estimates of this coefficient and differing numbers of iterations because of their use of different convergence rules. This form of incomplete matching of values likewise prevents estimation, though the effect is likely to be more subtle. In this case, a likely outcome is that the iterations will fail to converge, though the parameter estimates will not necessarily become extreme.

Here is an example of this effect at work. The probit model looks excellent in the full sample. In the restricted sample, *d* never equals zero when *y* equals zero. The estimator appears to have converged, the derivatives are zero, but the standard errors are huge:

```
SAMPLE    ; 1-100 $
CALC      ; Ran(12345) $
CREATE    ; x = Rnn(0,1) ; d = Rnu(0,1) > .5 $
CREATE    ; y = (-.5 + x + d + Rnn(0,1)) > 0 $
PROBIT    ; Lhs = y ; Rhs = one,x,d $
REJECT    ; y = 0 & d = 0 $
PROBIT    ; Lhs = y ; Rhs = one,x,d $
```

```
Normal exit:   6 iterations. Status=0, F=    42.82216
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                   Y
Log likelihood function      -42.82216
Restricted log likelihood    -69.13461
Chi squared [   2 d.f.]       52.62490
Significance level              .00000
McFadden Pseudo R-squared     .3805974
Estimation based on N =    100, K =   3
Inf.Cr.AIC  =    91.644 AIC/N =    .916
Hosmer-Lemeshow chi-squared =   6.83600
P-value=  .33628 with deg.fr. =       6
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.93918***      .23374    -4.02  .0001    -1.39729   -.48106
       X|    1.17177***      .24254     4.83  .0000      .69639   1.64715
       D|    1.53192***      .35304     4.34  .0000      .83997   2.22386
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y
Log likelihood function       -16.60262
Restricted log likelihood     -32.85957
Chi squared [   2 d.f.]        32.51388
Significance level               .00000
McFadden Pseudo R-squared      .4947400
Estimation based on N =     61, K =   3
Inf.Cr.AIC  =    39.205 AIC/N =    .643
Hosmer-Lemeshow chi-squared =   4.91910
P-value=  .08547 with deg.fr. =       2
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    14.0462       607774.4      .00 1.0000 ***********  ***********
       X|    1.41264***      .39338     3.59  .0003      .64163   2.18365
       D|   -13.3995       607774.4      .00 1.0000 ***********  ***********
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

You can check for this condition if you suspect it is present by using a crosstab. The command is

CROSSTAB ; Lhs = dependent variable
; Rhs = independent dummy variable $

The 2×2 table produced should contain four nonempty cells. If any cells contain zeros, as in the table below, then the model will be inestimable.

```
+-------------------------------------------------------------------+
|Cross Tabulation                                                   |
|Row variable is Y         (Out of range 0-49:       0)             |
|Number of Rows =  2       (Y          =  0 to  1)                  |
|Col variable is D         (Out of range 0-49:       0)             |
|Number of Cols =  2       (D          =  0 to  1)                  |
|Chi-squared independence tests:                                    |
|Chi-squared[   1] =    6.46052   Prob value =  .01103              |
|G-squared  [   1] =    9.92032   Prob value =  .00163              |
+-------------------------------------------------------------------+
|                      D                                            |
+--------+--------------+------+                                    |
|       Y|      0      1| Total|                                    |
+--------+--------------+------+                                    |
|       0|      0     14|    14|                                    |
|       1|     16     31|    47|                                    |
+--------+--------------+------+                                    |
|   Total|     16     45|    61|                                    |
+-------------------------------------------------------------------+
```

Missing Values

Missing values in the current sample will always impede estimation. In the case of the binary choice models, if your sample contains missing observations for the dependent variable, you will receive a warning about improper coding of the values of the Lhs variable. This message will be given whenever values of the dependent variable appear to be neither binary (0/1) nor a proportion, strictly between 0 and 1.

```
Probit: Data on Y are badly coded. (<0,1> and <=0 or >= 1).
```

Missing values for the independent variables will also badly distort the estimates. Since the program assumes you will be deciding what observations to use for estimation, and -999 (the missing value code) is a valid value, missing values on the right hand side of your model are not flagged as an error. But, it is obvious that they can seriously affect the results. The second model is computed without the missing values. The true values of the coefficients are both one, which is reflected in the much more reasonable second set of results.

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-1000 $
CREATE      ; x1 = Rnn(0,1) ; x2 = (Rnu(0,1) > .5) ; e=Rnn(0,1) $
CREATE      ; y = (-.5 + x1 +x2 + e) > 0 $
CREATE      ; If(_obsno > 900)x2 = -999 $
PROBIT      ; Lhs = y ; Rhs = one,x1,x2 $
SKIP $
PROBIT      ; Lhs = y ; Rhs = one,x1,x2 $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y
Log likelihood function      -524.80744
Restricted log likelihood    -693.13918
Chi squared [   2 d.f.]       336.66349
Significance level               .00000
McFadden Pseudo R-squared      .2428542
Estimation based on N =   1000, K =   3
Inf.Cr.AIC  = 1055.615 AIC/N =    1.056
Hosmer-Lemeshow chi-squared =  10.15008
P-value=  .25465 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .02186         .04709      .46  .6425     -.07043    .11414
      X1|     .92513***      .05870    15.76  .0000      .81008   1.04018
      X2|     .00018         .00015     1.23  .2176     -.00011    .00047
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
---------------------------------------------------------------
Deleted    100 observations with missing data. N is now    900
---------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                   Y
Log likelihood function      -416.47674
Restricted log likelihood    -623.79691
Chi squared [   2 d.f.]       414.64034
Significance level               .00000
McFadden Pseudo R-squared      .3323520
Estimation based on N =    900, K =   3
Inf.Cr.AIC  =  838.953 AIC/N =     .932
Hosmer-Lemeshow chi-squared =    .56208
P-value=  .99979 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.48950***       .07072    -6.92  .0000     -.62811   -.35090
      X1|    1.03767***       .06903    15.03  .0000      .90238   1.17297
      X2|    1.05649***       .10443    10.12  .0000      .85181   1.26117
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

You should use either **SKIP** or **REJECT** to remove the missing data from the sample. (See Chapter R7 for details on skipping observations with missing values.)

E27.3 Model Commands

The model commands for the six binary choice models listed above are largely the same:

PROBIT
LOGIT
ARCTANGENT
GOMPERTZ
COMPLOGLOG
BURR

; Lhs = dependent variable ; Rhs = regressors $

Data on the dependent variable may be either individual or proportions for all six cases. If the data are proportions, the dependent variable gives the proportion of ones. The program deduces the proportion of zeros as one minus this value. You need not make any special note of which. *LIMDEP* will inspect the data to determine which type of data you are using. In either case, you provide only a single dependent variable. As usual, you should include a constant term in the model unless your application specifically dictates otherwise.

The command builder dialog boxes may also be used to construct these commands. The probit, complementary log log, Gompertz and arctangent models are found in Models:Binary Choice/Probit. The Lhs and Rhs variables are specified on the Main page of the dialog box. Then, the Options page offers the various model choices shown in Figure E27.2.

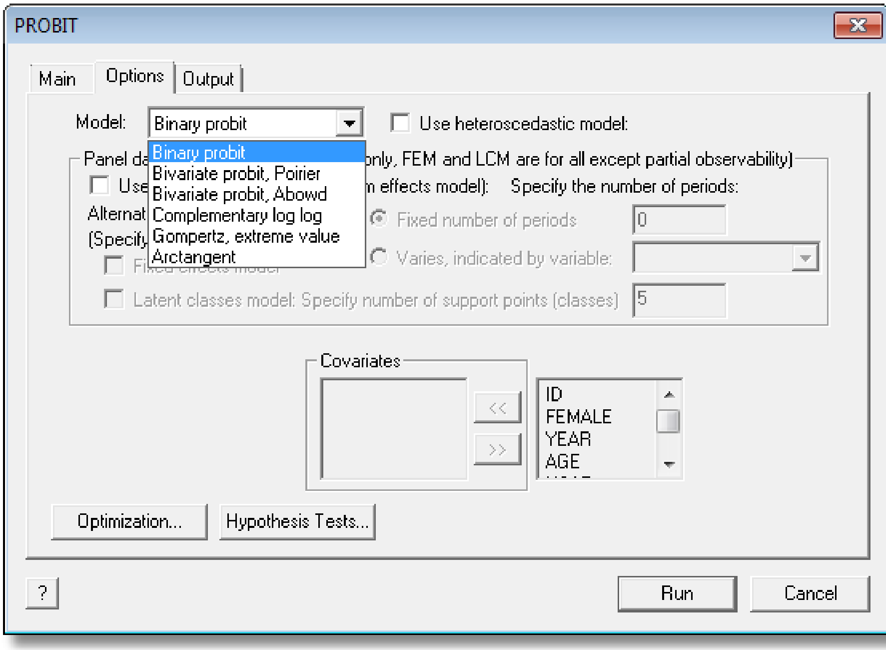


Figure E27.2 Command Builder Options Page for Probit and Other Models

The probit model is the default – no menu selection is necessary. The complementary log log, Gompertz and arctangent models are the last two options in this menu. Note, the command builder generates a probit command of the form

PROBIT	**; Lhs = ... ; Rhs = ...**
and optionally,	**; Model = Comploglog**
or	**; Model = Gompertz**
or	**; Model = Arctangent $**

which is equivalent to the separate commands shown above. The logit model is specified in Models:Binary Choice/Logit. The Burr model is a modification of the logit model and can be selected on the logit model Options page, as shown in Figure E27.3 – the check box is above the Optimization button.

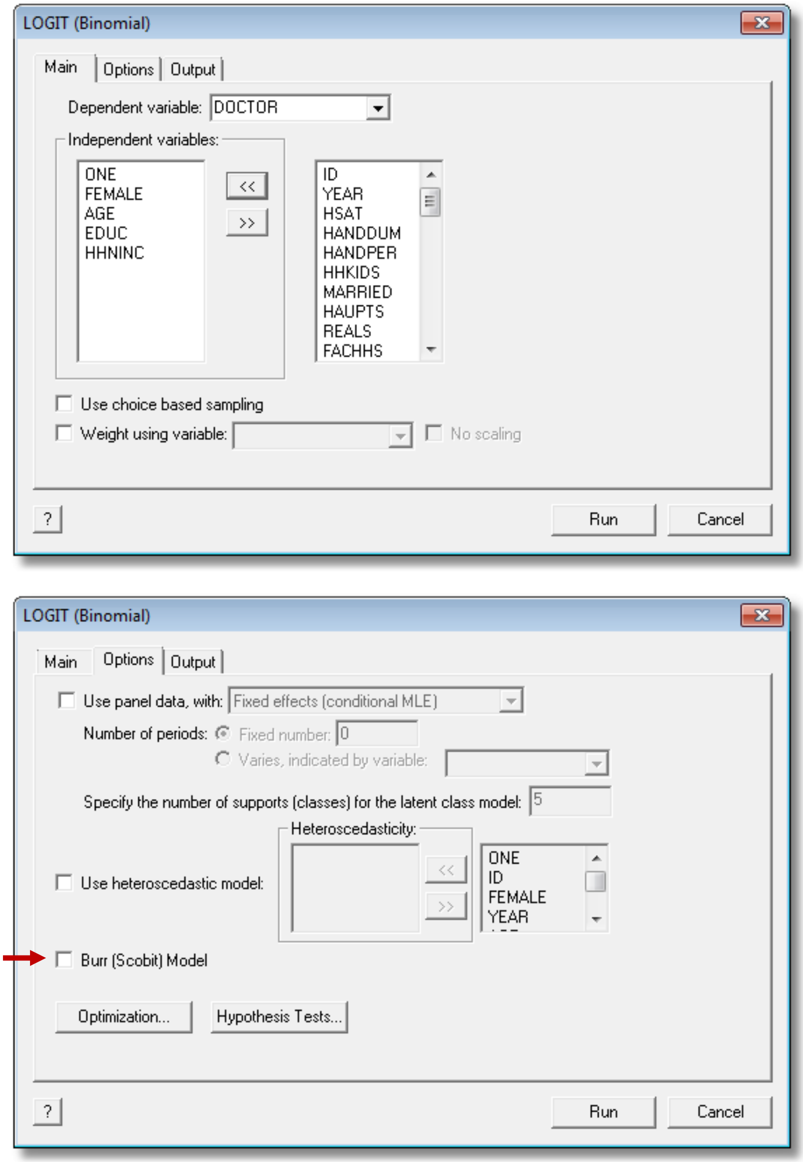


Figure E27.3 Command Builder for Logit Model and Burr Models

All of the standard options for optimization are available. These are discussed in Chapter R26. To reiterate, these are as follows (and operate the same in all model settings):

; Maxit = n	to set maximum iterations (may be 0 to compute LM statistics)
; Start = list	to give starting values (see Section E27.12)
; Tlf [= value]	to control convergence on the function value
; Tlb[= value]	to control convergence on change in parameters
; Tlg[= value]	to control convergence on derivatives weighted by inverse Hessian
; Alg = name	to request a particular algorithm, Newton, DFP, BFGS, etc.
; Output = value	to control the technical output in the displayed results

E27.4 Output

The binary choice models can produce a very large amount of optional output. Computation begins with some type of least squares estimation in order to obtain starting values. With ungrouped data, we simply use OLS of the binary variable on the regressors. If requested, the usual regression results are given, including diagnostic statistics, e.g., sum of squared residuals, and the coefficient 'estimates.' The OLS estimates based on individual data are known to be inconsistent. They will be visibly different from the final maximum likelihood estimates. For the grouped data case, the estimates are GLS, minimum chi squared estimates, which are consistent and efficient. Full GLS results will be shown for this case.

NOTE: The OLS results will not normally be displayed in the output. To request the display, use **; OLS** in any of the model commands.

E27.4.1 Reported Estimates

Final estimates include:

- $\log L$ = the log likelihood function at the maximum,
- $\log L_0$ = the log likelihood function assuming all slopes are zero. If your Rhs variables do not include *one*, this statistic will be meaningless. It is computed as

$$\log L_0 = n[P\log P + (1\text{-}P)\log(1\text{-}P)]$$

where P is the sample proportion of ones.

- McFadden's pseudo R^2 - 1 - $\log L/\log L_0$.
- The chi squared statistic for testing H_0: $\boldsymbol{\beta} = \mathbf{0}$ (not including the constant) and the significance level = probability that χ^2 exceeds test value. The statistic is

$$\chi^2 = 2(\log L - \log L_0).$$

- Akaike's information criterion, $-2(\log L - K)$ and the normalized AIC, = $-2(\log L - K)/n$.
- The sample and model sizes, n and K.
- Hosmer and Lemeshow's fit statistic and associated chi squared and p value. (The Hosmer and Lemeshow statistic is documented in Section E27.8.)

The standard statistical results, including coefficient estimates, standard errors, t ratios, p values and confidence intervals appear next. A complete listing is given below with an example. After the coefficient estimates are given, two additional sets of results can be requested, an analysis of the model fit and an analysis of the model predictions.

We will illustrate with binary logit and probit estimates of a model for visits to the doctor using the German health care data described in Chapter E2. The first model command is

LOGIT ; Lhs = doctor
; Rhs = one,age,hhninc,hhkids,educ,married
; OLS ; Summary
; Output = IC $ (Display all variants of information criteria)

Note that the command requests the optional listing of the OLS starting values and the additional fit and diagnostic results. The results for this command are as follows. With the exception of the table noted below, the same results (with different values, of course) will appear for all five parametric models. Some additional optional computations and results will be discussed later.

```
-----------------------------------------------------------------------------
Binomial Logit Model for Binary Choice
There are  2 outcomes for LHS variable DOCTOR
These are the OLS estimates based on the
binary variables for each outcome Y(i)=j.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     .63280***      .05584    11.33  .0000      .52335    .74224
     AGE|     .00387***      .00082     4.73  .0000      .00226    .00547
  HHNINC|    -.08338**       .03967    -2.10  .0356    -.16114   -.00563
  HHKIDS|    -.08456***      .01943    -4.35  .0000    -.12264   -.04647
    EDUC|    -.00804**       .00355    -2.27  .0234    -.01500   -.00109
 MARRIED|     .03209         .02131     1.51  .1321    -.00968    .07387
--------+--------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function      -2121.43961
Restricted log likelihood    -2169.26982
Chi squared [   5 d.f.]         95.66041
Significance level                .00000
McFadden Pseudo R-squared       .0220490
Estimation based on N =   3377, K =   6
Inf.Cr.AIC  = 4254.879 AIC/N =    1.260
FinSmplAIC  = 4254.904 FIC/N =    1.260
Bayes IC    = 4291.628 BIC/N =    1.271
HannanQuinn = 4268.018 HIC/N =    1.264
Hosmer-Lemeshow chi-squared =  17.65094
P-value=  .02400 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     .52240**       .24887     2.10  .0358      .03463   1.01018
     AGE|     .01834***      .00378     4.85  .0000      .01092    .02575
  HHNINC|    -.38750**       .17760    -2.18  .0291    -.73559   -.03941
  HHKIDS|    -.38161***      .08735    -4.37  .0000    -.55282   -.21040`
    EDUC|    -.03581**       .01576    -2.27  .0230    -.06669   -.00493
 MARRIED|     .14709         .09727     1.51  .1305    -.04357    .33774
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E27.4.2 Fit Measures

The model results are followed by a cross tabulation of the correct and incorrect predictions of the model using the rule

$$\hat{y} = 1 \text{ if } F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i) > .5, \text{ and 0 otherwise.}$$

For the models with symmetric distributions, probit and logit, the average predicted probability will equal the sample proportion. If you have a quite unbalanced sample – high or low proportion of ones – the rule above is likely to result in only one value, zero or one, being predicted for the Lhs variable. You can choose a threshold different from .5 by using

; Limit = the value you wish

in your command. There is no direct counterpart to an R^2 in regression. Authors very commonly report the

$$Pseudo - R^2 = 1 - \frac{\log L(\text{model})}{\log L(\text{constants only})}.$$

We emphasize, this is not a proportion of variation explained. Moreover, as a fit measure, it has some peculiar features. Note, for our example above, it is 1 - (-17673.10)/(-18019.55) = 0.01923, yet with the standard prediction rule, the estimated model predicts almost 63% of the outcomes correctly.

```
+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Logit    model for variable DOCTOR     |
+----------------------------------------+
|                 Y=0       Y=1     Total|
| Proportions  .34202    .65798   1.00000|
| Sample Size    1155      2222      3377|
+----------------------------------------+
| Log Likelihood Functions for BC Model  |
|              P=0.50    P=N1/N   P=Model|
| LogL =     -2340.76  -2169.27  -2121.44|
+----------------------------------------+
| Fit Measures based on Log Likelihood   |
| McFadden = 1-(L/L0)          =   .02205|
| Estrella = 1-(L/L0)^(-2L0/n) =   .02824|
| R-squared (ML)               =   .02793|
| Akaike Information Crit.     =  1.25996|
| Schwartz Information Crit.   =  1.27084|
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                        =   .02693|
| Ben Akiva and Lerman         =   .56223|
| Veall and Zimmerman          =   .04899|
| Cramer                       =   .02735|
+----------------------------------------+
```

The fit measures are documented in Section E27.8.

The next set of results examines the success of the prediction rule

$$\text{Predict } y_i = 1 \text{ if } P_i > P^* \text{ and 0 otherwise}$$

where P^* is a defined threshold probability. The default value of P^* is 0.5, which makes the prediction rule equivalent to 'Predict $y_i = 1$ if the model says the predicted event $y_i = 1 \mid \mathbf{x}_i$ is more likely than the complement, $y_i = 0 \mid \mathbf{x}_i$.' You can change the threshold from 0.5 to some other value with

; Limit = your *P**

Predictions for Binary Choice Model. Predicted value is 1 when probability is greater than .500000, 0 otherwise. Note, column or row total percentages may not sum to 100% because of rounding. Percentages are of full sample.

Actual Value	Predicted Value 0	Predicted Value 1	Total Actual
0	21 (.6%)	1134 (33.6%)	1155 (34.2%)
1	12 (.4%)	2210 (65.4%)	2222 (65.8%)
Total	33 (1.0%)	3344 (99.0%)	3377 (100.0%)

Crosstab for Binary Choice Model. Predicted probability vs. actual outcome. Entry = Sum[Y(i,j)*Prob(i,m)] 0,1. Note, column or row total percentages may not sum to 100% because of rounding. Percentages are of full sample.

Actual Value	Predicted Probability Prob(y=0)	Predicted Probability Prob(y=1)	Total Actual
y=0	415 (12.3%)	739 (21.9%)	1155 (34.2%)
y=1	739 (21.9%)	1482 (43.9%)	2222 (65.8%)
Total	1155 (34.2%)	2221 (65.8%)	3377 (99.9%)

This table computes a variety of conditional and marginal proportions based on the results using the defined prediction rule. For examples, the 66.697% equals (1482/2222)100% while the 66.727% is (1482/2221)100%.

```
-----------------------------------------------------------------------
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                     66.697%
Specificity = actual 0s correctly predicted                     35.931%
Positive predictive value = predicted 1s that were actual 1s    66.727%
Negative predictive value = predicted 0s that were actual 0s    35.931%
Correct prediction = actual 1s and 0s correctly predicted       56.174%
-----------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s          63.983%
False neg. for true pos. = actual 1s predicted as 0s          33.258%
False pos. for predicted pos. = predicted 1s actual 0s        33.273%
False neg. for predicted neg. = predicted 0s actual 1s        63.983%
False predictions = actual 1s and 0s incorrectly predicted     43.767%
-----------------------------------------------------------------------
```

E27.4.3 Covariance Matrix

The estimated asymptotic covariance matrix of the coefficient estimator is not automatically displayed – it might be huge. You can request a display with

; Covariance Matrix (or ; Printvc)

If the matrix is not larger than 5×5, it will be displayed in full. If it is larger, an embedded object that holds the matrix will show, instead. By double clicking the object, you can display the matrix in a window. An example appears in Figure E27.4 below.

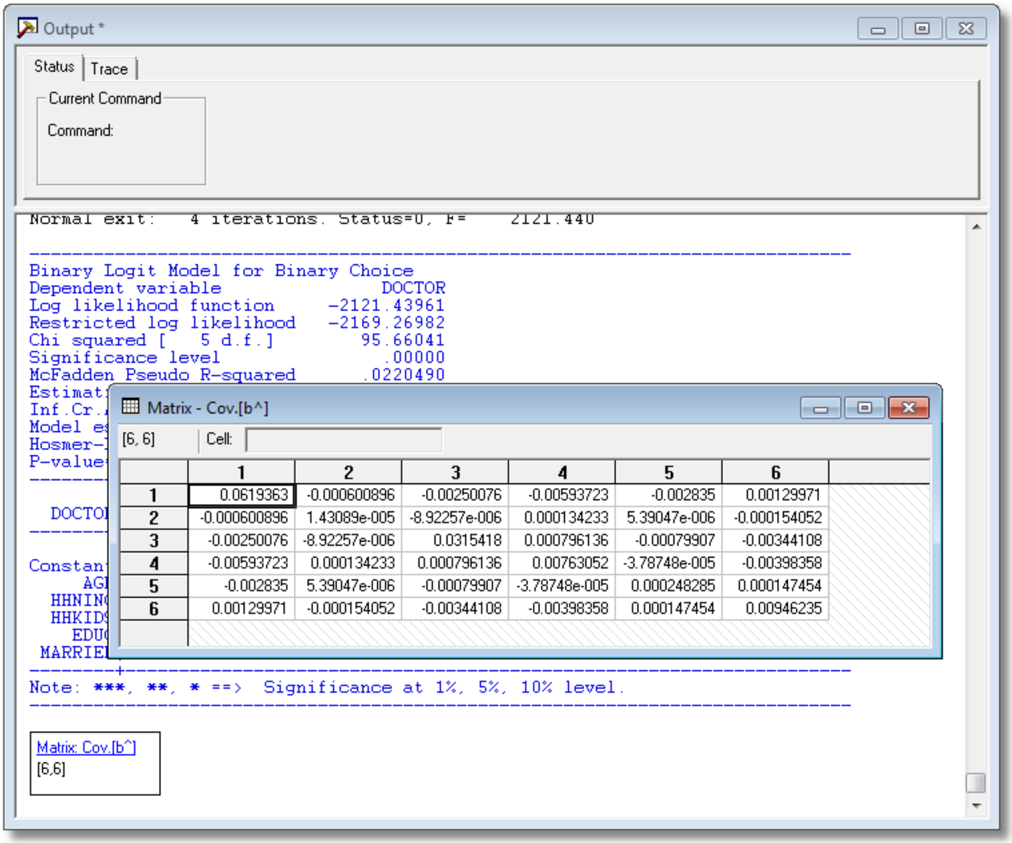


Figure E27.4 Embedded Matrix

E27.4.4 Retained Results and Generalized Residuals

The results saved by the binary choice models are:

Matrices: *b* = estimate of $\boldsymbol{\beta}$ (also contains γ for the Burr model)
varb = asymptotic covariance matrix

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Variables: *logl_obs* = individual contribution to log likelihood
score_fn = generalized residual. See Section E27.9.

Last Model: *b_variables*

Last Function: Prob(y = 1 | $\mathbf{x}$) = $F(\mathbf{b}'\mathbf{x})$. This varies with the model specification.

Models that are estimated using maximum likelihood automatically create a variable named *logl_obs*, that contains the contribution of each individual observation to the log likelihood for the sample. Since the log likelihood is the sum of these terms, you could, in principle, recover the overall log likelihood after estimation with

```
CALC        ; List ; Sum(logl_obs) $
```

The variable can be used for certain hypothesis tests, such as the Vuong test for nonnested models. The following is an example (albeit, one that appears to have no real power) that applies the Vuong test to discern whether the logit or probit is a preferable model for a set of data:

```
LOGIT       ; … $
CREATE      ; lilogit = logl_obs $
PROBIT      ; … $
CREATE      ; liprobit = logl_obs ; di = liprobit - lilogit $
CALC        ; List ; vtest = Sqr(n) * Xbr(di) / Sdv(di) $
```

The 'generalized residuals' in a parametric binary choice model are the derivatives of the log likelihood with respect to the constant term in the model. These are sometimes used to check the specification of the model (see Chesher and Irish (1987)). These are easy to compute for the models listed above – in each case, the generalized residual is the derivative of the log of the probability with respect to $\boldsymbol{\beta}'\mathbf{x}$. This is computed internally as part of the iterations, and kept automatically in your data area in a variable named *score_fn*. The formulas for the generalized residuals are provided in Section E27.12 with the technical details for the models. For example, you can verify the convergence of the estimator to a maximum of the log likelihood with the instruction

```
CALC        ; List ; Sum(score_fn) $
```

E27.5 Robust Covariance Matrix Estimation

The preceding describes a covariance estimator that accounts for a specific, observed aspect of the data. The concept of the 'robust' covariance matrix is that it is meant to account for hypothetical, unobserved failures of the model assumptions. The intent is to produce an asymptotic covariance matrix that is appropriate even if some of the assumptions of the model are not met. (It is an important, but infrequently discussed issue whether the estimator, itself, remains consistent in the presence of these model failures – that is, whether the so called robust covariance matrix estimator is being computed for an inconsistent estimator.) (Chapter R10 provides general discussion of robust covariance matrix estimation.)

E27.5.1 The Sandwich Estimator

A robust covariance matrix estimator adjusts the estimated asymptotic covariance matrix for possible misspecification in the model which leaves the MLE consistent but the estimated asymptotic covariance matrix incorrectly computed. One example would be a binary choice model with unspecified latent heterogeneity. A frequent adjustment for this case is the 'sandwich estimator,' which is the choice based sampling estimator suggested above with weights equal to one. (This suggests how it could be computed.) The desired matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log F_i}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'}\right)\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{\partial \log F_i}{\partial\hat{\boldsymbol{\beta}}}\right)\left(\frac{\partial \log F_i}{\partial\hat{\boldsymbol{\beta}}'}\right)'\right]\left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log F_i}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'}\right)\right]^{-1}$$

Three ways to obtain this matrix are

; Wts = one ; Choice based sampling
or **; Robust**
or **; Cluster = 1**

The computation is identical in all cases. (As noted below, the last of them will be slightly larger, as it will be multiplied by $n/(n-1)$.)

E27.5.2 Clustering

A related calculation is used when observations occur in groups which may be correlated. This is rather like a panel; one might use this approach in a random effects kind of setting in which observations have a common latent heterogeneity. The parameter estimator is unchanged in this case, but an adjustment is made to the estimated asymptotic covariance matrix. The calculation is done as follows: Suppose the n observations are assembled in G clusters of observations, in which the number of observations in the ith cluster is n_i. Thus,

$$\sum_{i=1}^{G} n_i = n.$$

Let the observation specific gradients and Hessians be

$$\mathbf{g}_{ij} = \frac{\partial \log L_{ij}}{\partial \boldsymbol{\beta}}$$

$$\mathbf{H}_{ij} = \frac{\partial^2 \log L_{ij}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

The uncorrected estimator of the asymptotic covariance matrix based on the Hessian is

$$\mathbf{V}_H = \mathbf{-H}^{-1} = \left(-\sum_{i=1}^{G} \sum_{j=1}^{n_i} \mathbf{H}_{ij}\right)^{-1}$$

Estimators for some models such as the Burr model will use the BHHH estimator, instead. In general,

$$\mathbf{V}_B = \left(\sum_{i=1}^{G} \sum_{j=1}^{n_i} \mathbf{g}_{ij}\mathbf{g}'_{ij}\right)^{-1}$$

Let **V** be the estimator chosen. Then, the corrected asymptotic covariance matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{V}\,\frac{G}{G-1}\left[\sum_{i=1}^{G} \left(\sum_{j=1}^{n_i} \mathbf{g}_{ij}\right)\left(\sum_{j=1}^{n_i} \mathbf{g}_{ij}\right)'\right]\mathbf{V}$$

Note that if there is exactly one observation per cluster, then this is $G/(G\text{-}1)$ times the sandwich estimator discussed above. Also, if you have fewer clusters than parameters, then this matrix is singular – it has rank equal to the minimum of G and K, the number of parameters.

This procedure is described in greater detail in Section E27.5.3. To request the estimator, your command must include

; Cluster = specification

where the specification is either the fixed value if all the clusters are the same size, or the name of an identifying variable if the clusters vary in size. Note, this is not the same as the variable in the Pds function that is used to specify a panel. The cluster specification must be an identifying code that is specific to the cluster. For example, our health care data used in our examples is an unbalanced panel. The first variable is a family *id*, which we will use as follows

; Cluster = id

The results below demonstrate the effect of this estimator. Three sets of estimates are given. The first are the original logit estimates that ignore the cross observation correlations. The second use the correction for clustering. The third is a panel data estimator – the random effects estimator described in Chapter E30 – that explicitly accounts for the correlation across observations. It is clear that the different treatments change the results noticeably.

Uncorrected covariance matrix

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    -.20205**       .09397    -2.15  .0315     -.38622   -.01787
     AGE|     .01935***      .00130    14.90  .0000      .01681    .02190
    EDUC|    -.02477***      .00578    -4.28  .0000     -.03611   -.01344
 MARRIED|     .12023***      .03376     3.56  .0004      .05405    .18640
  HHNINC|    -.21388***      .07580    -2.82  .0048     -.36245   -.06532
  HHKIDS|    -.24879***      .02983    -8.34  .0000     -.30726   -.19032
  FEMALE|     .58305***      .02620    22.26  .0000      .53171    .63439
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

Cluster corrected covariance matrix

```
+---------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of  27326 observations contained   7293 clusters defined by |
| variable ID       which identifies by a value a cluster ID.         |
+---------------------------------------------------------------+
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    -.20205         .12997    -1.55  .1200     -.45678    .05269
     AGE|     .01935***      .00176    11.00  .0000      .01590    .02280
    EDUC|    -.02477***      .00811    -3.05  .0023     -.04067   -.00888
 MARRIED|     .12023***      .04556     2.64  .0083      .03093    .20953
  HHNINC|    -.21388**       .09276    -2.31  .0211     -.39568   -.03209
  HHKIDS|    -.24879***      .03842    -6.48  .0000     -.32409   -.17349
  FEMALE|     .58305***      .03744    15.57  .0000      .50967    .65644
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

Random effects estimates

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    -.70495***      .18028    -3.91  .0001    -1.05830   -.35160
     AGE|     .03656***      .00241    15.18  .0000      .03184    .04128
    EDUC|    -.03703***      .01132    -3.27  .0011     -.05923   -.01484
 MARRIED|     .05481         .05570      .98  .3251     -.05435    .16397
  HHNINC|     .00772         .11698      .07  .9474     -.22156    .23700
  HHKIDS|    -.23497***      .04727    -4.97  .0000     -.32763   -.14232
  FEMALE|     .77202***      .05357    14.41  .0000      .66702    .87702
     Rho|     .39909***      .00586    68.07  .0000      .38760    .41058
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

E27.5.3 Stratification and Clustering

The clustering estimator is extended to include stratum level grouping, where a stratum includes one or more clusters, and weighting to allow finite population correction. We suppose that there are a total of S strata in the sample. Each stratum, 's,' contains C_s clusters. The number of observations in a cluster is N_{cs}. Neglecting the weights for the moment,

$$\text{Variance estimator} = \mathbf{VGV}$$

$$\mathbf{V} = \text{the inverse of conventional estimator of the Hessian}$$

$$\mathbf{G} = \sum_{s=1}^{S} w_s \mathbf{G}_s$$

$$\mathbf{G}_s = \left(\sum_{c=1}^{C_s} \mathbf{g}_{cs}\mathbf{g}'_{cs}\right) - \tfrac{1}{C_s}\mathbf{g}_s\mathbf{g}'_s$$

$$\mathbf{g}_s = \sum_{c=1}^{C_s} \mathbf{g}_{cs}$$

$$\mathbf{g}_{cs} = \sum_{i=1}^{N_{cs}} w_{ics}\mathbf{g}_{ics}$$

where $\mathbf{g}_{ics}$ is the derivative of the contribution to the log likelihood of individual i in cluster c in stratum s. The remaining detail in the preceding is the weighting factor, w_s. The stratum weight is computed as

$$w_s = f_s \times h_s \times d$$

where f_s = 1 or a finite population correction, 1 - C_s/C_s^* where C_s^* is the true number of clusters in stratum s, where $C_s^* \geq C_s$.

h_s = 1 or $C_s/(C_s - 1)$

d = 1 or $(N-1)/(N-K)$ where N is the total number of observations in the entire sample and K is the number of parameters (rows in $\mathbf{V}$).

Use

; Cluster	=	the number of observations in a cluster (fixed) or the name of a stratification variable which gives the cluster an identification. This is the setup that is described above.
; Stratum	=	the number of observations in a stratum (fixed) or the name of a stratification variable which gives the stratum an identification
; Wts	=	the name of the usual weighting variable for model estimation if weights are desired. This defines w_{ics}.
; FPC	=	the name of a variable which gives the number of clusters in the stratum. This number will be the same for all observations in a stratum – repeated for all clusters in the stratum. If this number is the same for all strata, then just give the number.
; Huber		Use this switch to request h_s. If omitted, h_s = 1 is used.
; DFC		Use this switch to request the use of d given above. If omitted, d = 1 is used.

Further details on this estimator may be found in Section E30.3 and Section R10.3.

E27.6 Analysis of Partial Effects

Partial effects in a binary choice model are

$$\frac{\partial E[y\,|\,\mathbf{x}]}{\partial \mathbf{x}} = \frac{\partial F(\boldsymbol{\beta}'\mathbf{x})}{\partial \mathbf{x}} = \frac{dF(\boldsymbol{\beta}'\mathbf{x})}{d(\boldsymbol{\beta}'\mathbf{x})}\boldsymbol{\beta} = F'(\boldsymbol{\beta}'\mathbf{x})\boldsymbol{\beta} = f(\boldsymbol{\beta}'\mathbf{x})\boldsymbol{\beta}.$$

That is, the vector of marginal effects is a scalar multiple of the coefficient vector. The scale factor, $f(\boldsymbol{\beta}'\mathbf{x})$, is the density function, which is a function of $\mathbf{x}$. (The densities for the six binary choice models are listed in Section E27.2.1.) This function can be computed at any data vector desired. Average partial effects are computed by averaging the function over the sample observations. The elasticity of the probability is

$$\frac{\partial \log E[y\,|\,\mathbf{x}]}{\partial \log x_k} = \frac{x_k}{E[y\,|\,\mathbf{x}]}\frac{\partial E[y\,|\,\mathbf{x}]}{\partial x_k} = \frac{x_k}{E[y\,|\,\mathbf{x}]} \times \text{marginal effect}$$

When the variable in $\mathbf{x}$ that is changing in the computation is a dummy variable, the derivative approach to estimating the marginal effect is not appropriate. An alternative which is closer to the desired computation for a dummy variable, that we denote z, is

$$\begin{aligned}\Delta F_z &= \text{Prob}[y = 1\,|\,z = 1] - \text{Prob}[y = 1\,|\,z = 0]\\ &= F(\boldsymbol{\beta}'\mathbf{x} + \alpha z\,|\,z = 1) - F(\boldsymbol{\beta}'\mathbf{x} + \alpha z\,|\,z = 0)\\ &= F(\boldsymbol{\beta}'\mathbf{x} + \alpha) - F(\boldsymbol{\beta}'\mathbf{x}).\end{aligned}$$

LIMDEP examines the variables in the model and makes this adjustment automatically.

There are two programs in *LIMDEP* for obtaining partial effects for the binary choice (and most other) models, the built in computation provided by the model command and the **PARTIAL EFFECTS** command. Examples of both are shown below.

The **LOGIT**, **PROBIT**, etc. commands provide a built in, basic computation for partial effects. You can request the computation to be done automatically by adding

; Partial Effects (or ; Marginal Effects)

to your command. The results below are produced for logit model in the earlier example. The standard errors for the partial effects are computed using the delta method. See Section E27.12 for technical details on the computation. The results reported are the average partial effects.

```
--------+-------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
Average partial effects for sample obs.
--------+-------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+-------------------------------------------------------------
     AGE|     .00402***       .26013     4.92  .0000       .00242    .00562
  HHNINC|    -.08666**       -.05857    -2.22  .0267     -.16331   -.01001
  HHKIDS|    -.08524***      -.05021    -4.33  .0000     -.12382   -.04667   #
    EDUC|    -.00779**       -.13620    -2.24  .0252     -.01461   -.00097
 MARRIED|     .03279          .03534     1.52  .1288     -.00952    .07510   #
--------+-------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

The equivalent **PARTIAL EFFECTS** (or just **PARTIALS**) command, which would immediately follow the **LOGIT** command, would be

PARTIAL EFFECTS ; Effects: age / hhninc / hhkids / educ / married
; Summary $

```
-----------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
-----------------------------------------------------------------------
      AGE            .00402     .00082    4.92       .00242      .00562
      HHNINC       -.08666      .03911    2.22      -.16331     -.01001
   *  HHKIDS       -.08524      .01968    4.33      -.12382     -.04667
      EDUC         -.00779      .00348    2.24      -.01461     -.00097
   *  MARRIED        .03279     .02159    1.52      -.00952      .07510
-----------------------------------------------------------------------
```

The second method provides a variety of options for computing partial effects under various scenarios, plotting the effects, etc. See Chapter R11 for further details.

NOTE: If your model contains nonlinear terms in the variables, such as *age*^2 or interaction terms such as *age**female*, then you must use the **PARTIAL EFFECTS** command to obtain partial effects. The built in routine in the command, **; Partial Effects**, will not give the correct answers for variables that appear in nonlinear terms.

E27.6.1 The Krinsky and Robb Method

An alternative to the delta method described above that is sometimes advocated is the Krinsky and Robb method. By this device, we have our estimate of the model coefficients, **b**, and the estimated asymptotic covariance matrix, **V**. The marginal effects are computed as a function of **b** and the vector of means of the sample data, $\bar{\mathbf{x}}$, say $g_k(\mathbf{b}, \bar{\mathbf{x}})$ for the *k*th variable. The Krinsky and Robb technique involves sampling *R* draws from the asymptotic normal distribution of the estimator, computing the function with these *R* draws, then computing the empirical variance. This is not done automatically by the binary choice estimator, but you can easily do the computation using the **WALD** command. For an example, we will use this method to compute the marginal effects for two variables in the logit model estimated earlier. The program would be

```
NAMELIST    ; x = one,age,hhninc,hhkids,educ,married $
LOGIT       ; Lhs = doctor ; Rhs = x ; Partial Effects $
MATRIX      ; xbar = Mean(x) $
CALC        ; kx = Col(x) ; Ran(12345) $
WALD        ; Start = b ; Var = varb ; Labels = kx_b
            ; Fn1 = b2 * Lgd(b1'xbar)
            ; Fn2 = b3 * Lgd(b1'xbar)
            ; K&R ; Pts = 2000 $
```

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic             =     27.72506
Prob. from Chi-squared[ 2] =       .00000
Krinsky-Robb method used with 2000 draws
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
WaldFcns|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
 Fncn(1)|     .00409***      .00084     4.85  .0000      .00244    .00575
 Fncn(2)|    -.08694**       .03913    -2.22  .0263     -.16363   -.01025
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

```
---------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .00402     .00082    4.92      .00242       .00562
      HHNINC       -.08666     .03911    2.22     -.16331      -.01001
---------------------------------------------------------------------
```

There is a second sources of difference between the Krinsky and Robb estimates and the delta method results that follow: The Krinsky and Robb procedure is based on the means of the data while the delta method averages the partial effects over the observations. It is possible to perform the K&R iteration at every observation to reproduce the APE calculations by adding **; Average** to the **WALD** command. The results below illustrate.

```
--------+--------------------------------------------------------------------
 Fncn(1)|     .00407***      .00085     4.80  .0000      .00241    .00573
 Fncn(2)|    -.08673**       .03929    -2.21  .0273     -.16373   -.00973
--------+--------------------------------------------------------------------
```

We do not recommend this as a general procedure, however. It is enormously time consuming and does not produce a more accurate result.

Estimating Marginal Effects by Strata

Marginal effects may be calculated for indicated subsets of the data by using

; Margin = variable

where '*variable*' is the name of a variable coded 0,1,... which designates up to 10 subgroups of the data set, in addition to the full data set. For example, a common application would be

; Margin = sex

in which the variable *sex* is coded 0 for men and 1 for women (or vice versa). The variable used in this computation need not appear in the model; it may be any variable in the data set.

For example, using our logit model above, we now compute marginal effects separately for men and women:

LOGIT ; Lhs = doctor
; Rhs = one,age,hhninc,hhkids,educ,married
; Margin = female $

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function      -2121.43961
Restricted log likelihood    -2169.26982
Chi squared [   5 d.f.]         95.66041
Significance level                .00000
McFadden Pseudo R-squared       .0220490
Estimation based on N =   3377, K =   6
Inf.Cr.AIC  = 4254.879 AIC/N =    1.260
Hosmer-Lemeshow chi-squared =  17.65094
P-value=  .02400 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     .52240**       .24887     2.10  .0358      .03463   1.01018
     AGE|     .01834***      .00378     4.85  .0000      .01092    .02575
  HHNINC|    -.38750**       .17760    -2.18  .0291     -.73559   -.03941
  HHKIDS|    -.38161***      .08735    -4.37  .0000     -.55282   -.21040
    EDUC|    -.03581**       .01576    -2.27  .0230     -.06669   -.00493
 MARRIED|     .14709         .09727     1.51  .1305     -.04357    .33774
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used are FEMALE=0
--------+--------------------------------------------------------------------
        |     Partial                         Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00414***      .26343     4.84  .0000      .00247    .00582
  HHNINC|    -.08756**      -.06038    -2.18  .0291     -.16619   -.00893
  HHKIDS|    -.08714***     -.05161    -4.34  .0000     -.12645   -.04783   #
    EDUC|    -.00809**      -.14612    -2.27  .0234     -.01509   -.00109
 MARRIED|     .03351         .03549     1.50  .1334     -.01025    .07728   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used are FEMALE=1
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|     .00404***       .26337     4.88  .0000      .00242    .00567
  HHNINC|    -.08545**       -.05555    -2.18  .0290     -.16217   -.00873
  HHKIDS|    -.08519***      -.04911    -4.33  .0000     -.12379   -.04659   #
    EDUC|    -.00790**       -.13086    -2.28  .0225     -.01468   -.00111
 MARRIED|     .03279          .03550     1.50  .1345     -.01015    .07573   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used are All Obs.
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|     .00410***       .26352     4.86  .0000      .00244    .00575
  HHNINC|    -.08660**       -.05811    -2.18  .0291     -.16436   -.00884
  HHKIDS|    -.08626***      -.05044    -4.34  .0000     -.12524   -.04727   #
    EDUC|    -.00800**       -.13893    -2.27  .0230     -.01490   -.00110
 MARRIED|     .03318          .03551     1.50  .1339     -.01021    .07658   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+--------------------------------------------+
| Marginal Effects for Logit                 |
+----------+----------+----------+----------+
| Variable | FEMALE=0 | FEMALE=1 | All Obs. |
+----------+----------+----------+----------+
| AGE      |   .00414 |   .00404 |   .00410 |
| HHNINC   |  -.08756 |  -.08545 |  -.08660 |
| HHKIDS   |  -.08714 |  -.08519 |  -.08626 |
| EDUC     |  -.00809 |  -.00790 |  -.00800 |
| MARRIED  |   .03351 |   .03279 |   .03318 |
+----------+----------+----------+----------+
```

The computation using the built in estimator is done at the strata means of the data. The computation can be done by averaging across observations using the **PARTIAL EFFECTS** command. For example, the corresponding results for the income variable are obtained with

PARTIAL EFFECTS ; Effects: hhninc @ female=0,1$

```
-----------------------------------------------------------------------
Partial Effects  Analysis for Logit Probability Function
-----------------------------------------------------------------------
Effects on function with respect to HHNINC
Results are computed by average over sample observations
Partial effects for continuous HHNINC   computed by differentiation
Effect is computed as derivative     = df(.)/dx
-----------------------------------------------------------------------
df/dHHNINC          Partial    Standard
(Delta method)      Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
Subsample for this iteration is FEMALE   =  0   Observations:    1812
APE. Function       -.08585     .03925    2.19     -.16278     -.00892
-----------------------------------------------------------------------
Subsample for this iteration is FEMALE   =  1   Observations:    1565
APE. Function       -.08355     .03820    2.19     -.15841     -.00868
```

Examining the Effect of a Variable Over a Range of Values

Another useful device is a plot of the probability (conditional mean) over the range of a variable of interest either holding other variables at their means, or averaging over the sample values. The figure below does this for the income variable in the logit model for doctor visits. The figure is plotted for *hhkids* = 1 and *hhkids* = 0 to show the two effects. We see that the probability falls with increased income, and also for individuals in households in which there are children.

SIMULATE ; Scenario: & hhninc = 0(.05).5 | hhkids=0,1 ; plot$

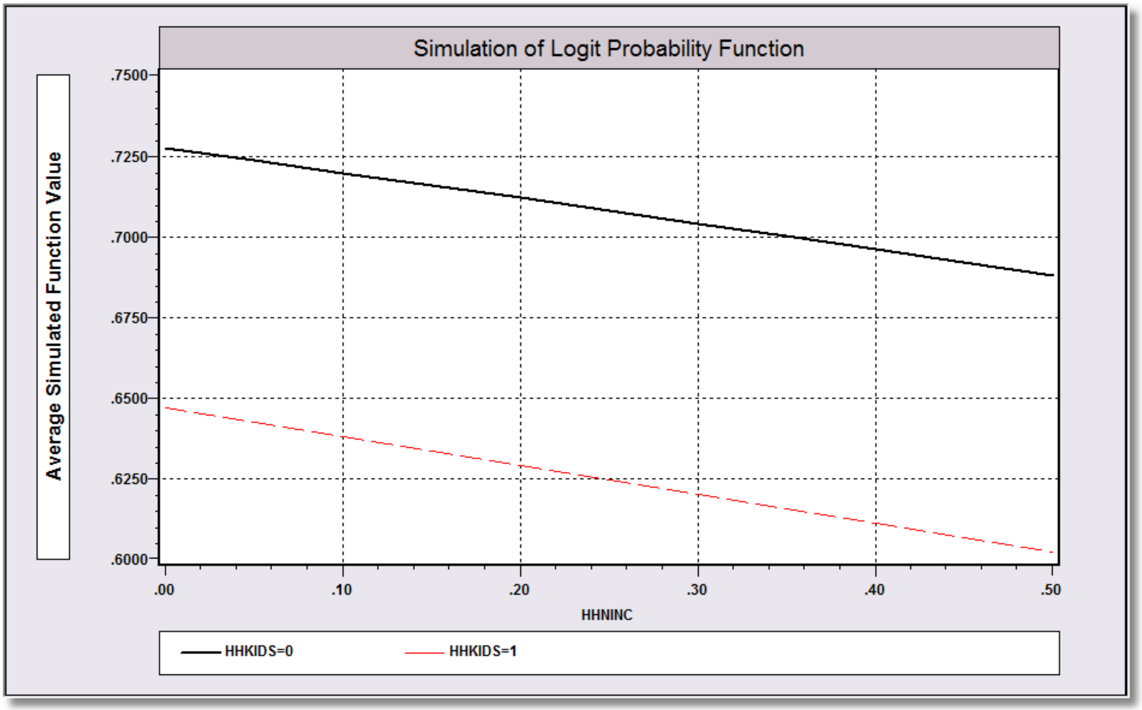


Figure E27.5 Probabilities Varying with Income

E27.7 Simulation and Analysis of a Binary Choice Model

This section describes a procedure that is used with all of the parametric models described above. It is used for two specific analyses. This procedure allows you to analyze the predictions made by a binary choice when the variables in the model are changed. It is similar to the **; Simulate** feature in *NLOGIT* 5. The analysis is provided in two parts:

- Change specific variables in the model by a prescribed amount, and examine the changes in the model predictions.
- Vary a particular variable over a range of values and examine the predicted probabilities when other variables are held fixed at their means.

This program is available for the six parametric binary choice models: *probit*, *logit*, *Gompertz*, *complementary log log*, *arctangent* and *Burr*. The *probit* and *logit* models may also be heteroscedastic. The routine is accessed as follows. First fit the model as usual. Then, use the identical model specification as shown below with the specifications indicated:

(MODEL) ; Lhs = ... ; Rhs = ... $

Then

```
BINARY CHOICE ; Lhs   = (the same) ; Rhs = (the same) ; ... (also the same)
              ; Model = Probit, Logit, Gompertz, Comploglog or Burr
              ; Start = B (from the preceding model)
```

(optional, the value to use for predicting Lhs = 1, default = .5)

; Threshold = P*

(optional) **; Scenario: variable operation = value /**
(variable operation = value) / ... (may be repeated)

(optional) **; Plot: variable (lower limit, upper limit) $**

In the **; Plot** specification, the limits part may be omitted, in which case the range of the variable is used. This will replicate for the one variable the computation of the program in the preceding section.

The **; Scenario** section computes all predicted probabilities for the model using the sample data and the estimated parameters. Then, it recomputes the probabilities after changing the variables in the way specified in the scenarios. (The actual data are not changed – the modification is done while the probabilities are computed.) The scenarios are of the form

variable operation = value

such as **hhkids + = 1** (effect of additional kids in the home)
or **hhninc * = 1.1** (effect of a 10% increase in income)

You may provide multiple scenarios. They are evaluated one at a time. This is an extension of the computation of marginal effects.

In the example below, we extend the analysis of marginal effect in the logit model used above. The scenario examined is the impact of every individual having one more child in the household then having a 50% increase in income. (Since *hhkids* is actually a dummy variable for the presence of kids in the home, increasing it by one is actually an ambiguous experiment. We retain it for the sake of a simple numerical example.) The plot shows the effect of income on the probability of visiting the doctor, according to the model.

```
NAMELIST    ; x = one,age,educ,married,hhninc,hhkids $
LOGIT       ; Lhs = doctor ; Rhs = x $
BINARY      ; Lhs = doctor ; Rhs = x
            ; Model = Logit ; Start = b
            ; Scenario:  hhkids + = 1 / hhninc * = 1.5  $
```

The model output is omitted for brevity.

```
+--------------------------------------------------------------+
|Scenario 1. Effect on aggregate proportions. Logit    Model  |
|Threshold T* for computing Fit = 1[Prob > T*] is  .50000     |
|Variable changing = HHKIDS  , Operation = +, value =   1.000 |
+--------------------------------------------------------------+
|Outcome           Base case       Under Scenario   Change    |
|      0        33 =    .98%       831 =   24.61%      798    |
|      1      3344 =  99.02%      2546 =   75.39%     -798    |
|  Total      3377 = 100.00%      3377 =  100.00%        0    |
+--------------------------------------------------------------+
+--------------------------------------------------------------+
|Scenario 2. Effect on aggregate proportions. Logit    Model  |
|Threshold T* for computing Fit = 1[Prob > T*] is  .50000     |
|Variable changing = HHNINC  , Operation = *, value =   1.500 |
+--------------------------------------------------------------+
|Outcome           Base case       Under Scenario   Change    |
|      0        33 =    .98%       106 =    3.14%       73    |
|      1      3344 =  99.02%      3271 =   96.86%      -73    |
|  Total      3377 = 100.00%      3377 =  100.00%        0    |
+--------------------------------------------------------------+
```

The **SIMULATE** command used in the example provides a greater range of scenarios that one can examine to see the effects of changes in a variable on the overall prediction of the binary choice model. The advantage of the **BINARY** command used here is that for straightforward scenarios, it can be used to provide useful tables such as the ones shown above.

E27.8 Measuring Fit in Binary Choice Models

A description of the ability of the binary choice model to predict the dependent variable is given by a 2×2 table which gives the success rate of the prediction rule

Predict $y_i = 1$ if fitted probability for $y_i = \hat{P}_i > P^*$, and 0 otherwise.

(This has been labeled a *confusion matrix* elsewhere in the literature.) This is the table produced by the logit model above.

```
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |      21 (    .6%)|   1134 ( 33.6%)|   1155 ( 34.2%)|
|  1   |      12 (    .4%)|   2210 ( 65.4%)|   2222 ( 65.8%)|
+------+----------------+----------------+----------------+
|Total |      33 (  1.0%)|   3344 ( 99.0%)|   3377 (100.0%)|
+------+----------------+----------------+----------------+
```

The value of P^* is reported with the table. This will normally be 0.5. But, if your sample is very unbalanced you may wish to change this with

; Limit = the desired value

In general, the better the model is, the larger will be the number of observations on the diagonals of this table. For example, by adding

; Limit = .6

in the model command, we obtain the following results:

```
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .600000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |     379 ( 11.2%)|    776 ( 23.0%)|   1155 ( 34.2%)|
|  1   |     555 ( 16.4%)|   1667 ( 49.4%)|   2222 ( 65.8%)|
+------+----------------+----------------+----------------+
|Total |     934 ( 27.7%)|   2443 ( 72.3%)|   3377 (100.0%)|
+------+----------------+----------------+----------------+
```

This actually worsens the fit of the model, based on the simple count of correct predictions. The change in the rule improves the ‘hit rate’ on the zeros, but at the cost of lowering the success at predicting the ones. This does say something about this criterion for model fit.

Hosmer and Lemeshow Diagnostic Statistic

Hosmer and Lemeshow have proposed a diagnostic measure for the probit and logit models (they focus on the latter) that assesses the match between actual and predicted values. To do the computation, we compute a fitted probability, F_i for each observation using the estimated model parameters. We then sort the fitted values in ascending order, carrying the actual y_i with them. The data are then divided into 10 percentiles based on the fitted values, and means of the predicted and actual data are computed within each group. The statistic is

$$H = \sum_{j=1}^{10} n_j \left[\frac{\left(\bar{y}_j - \bar{F}_j\right)^2}{\hat{\bar{P}}_j\left(1-\bar{F}_j\right)} \right]$$

(If the sample is not large, some groups at the high or low end may have insufficient variation to compute the denominator – the fitted values may all be very close to zero or one. The resulting statistic has a limiting chi squared distribution with eight (or fewer) degrees of freedom. Large values of the statistic suggest that the model is inappropriate. The example for the health care data below suggests this case.

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                 DOCTOR
Log likelihood function       -2121.43961
Restricted log likelihood     -2169.26982
Chi squared [    5 d.f.]         95.66041
Significance level                 .00000
McFadden Pseudo R-squared        .0220490
Estimation based on N =    3377, K =    6
Inf.Cr.AIC  = 4254.879 AIC/N =     1.260
Hosmer-Lemeshow chi-squared =  17.65094
P-value=  .02400 with deg.fr. =       8
--------+--------------------------------------------------------------------
```

Scalar Fit Measures for Binary Choice Models

Numerous other scalar fit measures have been proposed for binary choice models. They share the flaw that none satisfactorily mimic the true measure of proportion of variation explained given by R^2 in the linear regression context. *LIMDEP* reports several of these in a table with each set of estimates: (We are unable to recommend any of these as optimal. There is some discussion in Estrella (1998) which may be useful. See, also, Greene and Hensher (2010, Chapter 4).)

```
+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Logit    model for variable DOCTOR     |
+----------------------------------------+
|                  Y=0       Y=1     Total|
| Proportions  .34202    .65798   1.00000|
| Sample Size    1155      2222      3377|
+----------------------------------------+
| Log Likelihood Functions for BC Model  |
|              P=0.50    P=N1/N   P=Model|
| LogL =     -2340.76  -2169.27  -2121.44|
+----------------------------------------+
```

```
+----------------------------------------+
| Fit Measures based on Log Likelihood   |
| McFadden = 1-(L/L0)          =  .02205|
| Estrella = 1-(L/L0)^(-2L0/n) =  .02824|
| R-squared (ML)               =  .02793|
| Akaike Information Crit.     = 1.25996|
| Schwartz Information Crit.   = 1.27084|
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                        =  .02693|
| Ben Akiva and Lerman         =  .56223|
| Veall and Zimmerman          =  .04899|
| Cramer                       =  .02735|
+----------------------------------------+
```

The values in the table are computed as follows:

K	= number of coefficients in the model
N	= sample size
P_0	= proportion of zeros in the sample
P_1	$= 1 - P_0 = \overline{y}$
F_i	= Predicted probability that y_i equals 1 | $\mathbf{x}_i$
$\hat{P}_i$	= Predicted probability of observed $y_i = (1\text{-}y_i)(1\text{-}F_i) + y_iF_i$
L_0	= log likelihood with only a constant $= n\,(P_0 \log P_0 + P_1 \log P_1)$
L	= log likelihood $= \sum_{i=1}^{n} y_i \log F_i + (1-y_i)\log(1-F_i)$
McFadden	$= 1 - L/L_0$
Estrella	$= 1-(\log L/\log L_0)^{-2L_0/n}$
R^2 - ML	$= 1 - \exp[2(\log L_0 - \log L)/n]$
Akaike	$= (-2\log L + 2K)/n$
Schwarz	$= (-2\log L + K\log n)/n$
Efron	$= 1 - \sum_{i=1}^{n}\left(y_i - \hat{P}_i\right)^2 \Big/ \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$
Ben-Akiva	$= \dfrac{1}{n}\sum_i \hat{P}_i$
Veall	$= \left[(\delta-1)/(\delta-\text{McFadden})\right]\text{McFadden}, \delta = \text{N}/(2\log\text{L}_0)$
Cramer	= Average value of $\hat{P}_i \| y_i = 1$ - Average value of $\hat{P}_i \| y_i = 0$

You can obtain this same table of values for a binary variable y and any set of predicted probabilities contained in a variable with

CALC ; Fit (name of y variable, name of probabilities variable) $

This command always produces the output even if **; List** is not specified, but it does not produce any other results. The result of the **CALC** command is zero. (When you use this, the information criteria are not computed, as the degrees of freedom is not known.)

A Goodness of Fit Measure for the Probit Model Based on the Normal Distribution

This program computes a pseudo R squared for a probit model based on the formula given by Zavoina and McKelvey (1975) in their paper on the ordered probit model:

$$E[y_i^*|y_i] = yf_i = \mathbf{x}_i'\boldsymbol{\beta} + \lambda_i$$
$$\lambda_i = (2y_i - 1)\phi(\mathbf{x}_i'\boldsymbol{\beta}) / \Phi[(2y_i - 1)\mathbf{x}_i'\boldsymbol{\beta}]$$
$$R^2 = \text{Var}(yf) / [1 + \text{Var}(yf)]$$

where λ_i is the inverse Mill's ratio usually kept for **SELECT**. After setting up the sample for the problem, the commands are

```
NAMELIST   ; x = the Rhs variables for the probit model $
PROBIT     ; Lhs = y ; Rhs = x ; Hold(IMR = lambda) $
CREATE     ; yf = x'b + lambda $
CALC       ; zm = Var(yf) / (1 + Var(yf)) $
```

ROC Plots for Binary Choice Models

ROC (receiver operating characteristic) plots provide a loose descriptive measure of fit in a binary choice model, and can be used to some extent to compare models. You may obtain these for all parametric binary choice models: *logit* (with or without heteroscedasticity), *probit* (with or without heteroscedasticity), *complementary log log*, *Gompertz* and *Burr*. The request is simply

; ROC

added to any binary choice model command. An example appears below. The curve is constructed by computing for the range of values of P^* from zero to one,

Sensitivity(P^*) = proportion of observations for which estimated and actual values of y_i are both equal to one when the estimated y_i equals one if the the predicted probability is greater than or equal to P^*.

and

Specificity(P^*) = the proportion of values for which predicted and actual zeros match.

The graph is constructed by plotting Sensitivity(P^*) against 1 - Specificity(P^*). The 'fit measure' is then computed as the area under the ROC curve. A greater area implies a greater model fit. (The field is a unit rectangle.) A model with no fit has an area of 0.5. The request for the ROC plot also produces a plot of the ability of the model to predict zeros and ones, again as a function of P^*. This figure is produced by plotting the Specificity(P^*) and Sensitivity(P^*) against P^*. An example based on the earlier logit model appears below.

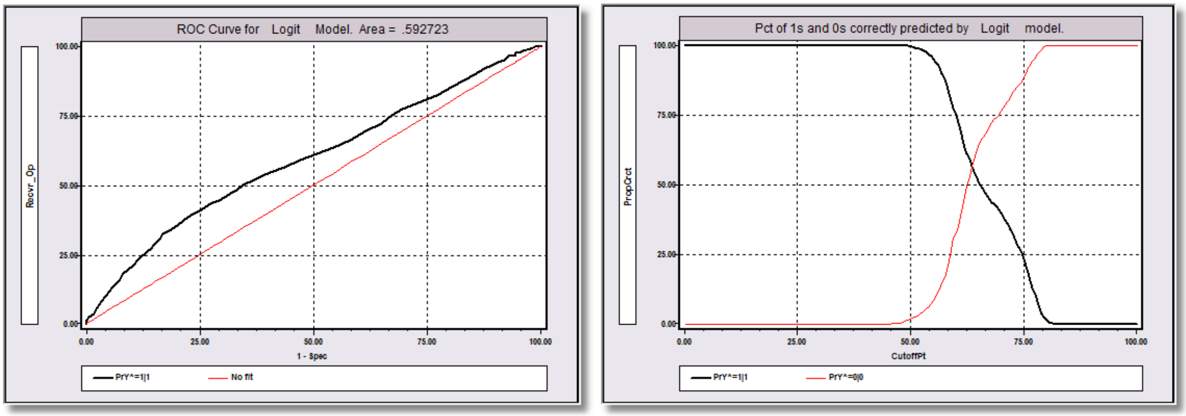


Figure E27.6 Analysis of Model Fit

E27.9 Saving Predictions and Residuals

Predictions for the binary choice models are created by adding

; Keep = the name of the variable.

Predictions are computed using the rule,

Predict y_i = 1 if fitted probability > P^*, and 0 otherwise.

Predictions retained with **; Keep = name** are the samples of ones and zeros produced by the prediction rule above. You may also keep the predicted probabilities, $F(\hat{\beta}'\mathbf{x}_i)$ with

; Prob = name

Residuals are requested with

; Res = name

They are the difference between actual and predicted values; residuals may be -1, 0, or 1. These results may be displayed with the model results by adding

; List

to the model command. The listing for probit model based on a small data set is shown below. (Our health care data set contains over 27,000 observations. We would not want to list a sample this large.) We do note, these ‘residuals’ are unlikely to be useable in this form. The generalized residuals for the model discussed below are likely to be more useable as a diagnostic tool.

Observation	Observed Y	Predicted Y	Residual	x(i)b	Prob[Y=1]
1	.00000	.00000	.0000	-2.0931	.0182
2	.00000	.00000	.0000	-1.6157	.0531
3	.00000	.00000	.0000	-.8782	.1899
4	.00000	.00000	.0000	-2.0842	.0186
5	1.0000	1.0000	.0000	.1372	.5546
(rows omitted)					
21	.00000	.00000	.0000	-1.5388	.0619
22	1.0000	1.0000	.0000	1.3079	.9045
23	.00000	.00000	.0000	-.6032	.2732
24	.00000	1.0000	-1.0000	1.0256	.8475
25	1.0000	1.0000	.0000	.9709	.8342

E27.10 Using Weights and Choice Based Sampling

The **; Wts** option can always be used in the usual fashion for the probit and logit models. However, in the grouped data case, a somewhat different treatment may be desired. The observations may consist of p_i, $\mathbf{x}_i$ and n_i, where n_i is the number of replications used to obtain p_i. The usual treatment assumes that p_i is a sample of one from a distribution with variance $p_i(1-p_i)$. But p_i is more precise than this. Its unconditional variance is $p_i(1-p_i)/n_i$. Thus, the efficiency of the estimator of $\boldsymbol{\beta}$ is underestimated. There is also an inherent heteroscedasticity which must be accounted for. The heteroscedasticity due to p_i is built into the likelihood function. But if your proportions are based on different numbers of observations, the variances will differ correspondingly. This can be accounted for by including n_i as a weighting variable. Since the weighting procedure automatically scales the weights so that they sum to the sample size, which would be inappropriate here, it is necessary to modify the specification. Use

; Wts = variable, Noscale

or just

; Wts = variable, N

to prevent the automatic scaling. This produces a replication of the observations, which is what is needed for grouped data.

This usage often has the surprising side effect of producing implausibly small standard errors. Consider, for example, using unscaled weights for statewide observations on election outcomes. The implication of the **Noscale** parameter is that each proportion represents millions of observations. Once again, this is an issue that must be considered on a case by case basis.

Choice Based Sampling

In some individual data cases, the data are deliberately sampled so that one or the other outcome is overrepresented in the sample. For example, suppose that in a binary response setting, the true proportion of ones in the population is .05 and the true proportion of zeros is .95. One might over sample the ones in order to learn more about the decision process. However, some account must be taken of this fact in the estimation since it obviously will impart some biases. The following assumes that these population proportions are known, which must be true to apply the technique. We use the assumed values to demonstrate the technique; other values would be substituted in the analogous manner.

The general principle involved is as follows: Suppose that the sample is deliberately drawn so that it contains 50% ones and 50% zeros while it is known that the true proportions in the population are .05 and .95. Then, the ones are overrepresented by a factor of .50/.05 = 10 while the zeros are underrepresented by a factor of .50/.95 = .5263. To obtain the right 'mix' in the sample, it is necessary to scale down the ones by a factor of .05/.50 = .1 and scale up the zeros by a factor of .95/.50 = 1.9. This can be handled simply by using a weighting variable during estimation to reweight the observations. The precise method of doing so is discussed below. (See, also, Manski and McFadden (1981).)

An additional change must be made in order to obtain the correct asymptotic covariance matrix for the estimates. Let **H** be the Hessian of the (weighted) log likelihood, i.e., the usual estimator for the variance matrix of the estimates, and let **G′G** be the summed outer products of the first derivatives of the (weighted) log likelihood. (This is the inverse of the BHHH estimator.) Manski and McFadden (1981) show that the appropriate covariance matrix for the estimates is

$$\mathbf{V} = (\mathbf{-H})^{-1}\ \mathbf{G'G}\ (\mathbf{-H})^{-1}.$$

The computation of the weighted estimator and the corrected asymptotic covariance is handled automatically in *LIMDEP* by the following estimation programs:

- univariate probit, logit, extreme value and Gompertz model,
- bivariate probit model with and without sample selection,
- binomial and multinomial logit models,
- discrete choice (conditional logit).

With the exception of the last of these, you request the estimator with

```
; Wts = name of weighting variable
; Choice Based
```

The weighting variable can usually be created with a single command. For example, the weighting variable suggested in the example used above would be specified as follows:

```
CREATE      ; wt = (.95/.50)*(y = 0)  +  (.05/.50)*(y = 1) $
```

For models that do not appear in the list above, there is a general way to do this kind of computation. How the weights are obtained will be specific to your application if you wish to do this. To compute the counterpart to **V** above, you can do the following:

```
CREATE      ; wt = the desired weighting variable $
Model name  ; ... specification of the model
            ; Wts = the weighting variable
            ; Cluster = 1 $
```

Since the 'cluster' estimator computes a sandwich estimator, we need only 'trick' the program by specifying that each cluster contains one observation. The observations in the parts will be weighted by the variable given, so this is exactly what is needed.

E27.11 Heteroscedasticity in Probit and Logit Models

The univariate choice model with multiplicative heteroscedasticity is

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

$$y_i = 1 \text{ if } y_i^* > 0 \text{ and } y_i = 0 \text{ if } y_i^* \leq 0,$$

$$\varepsilon_i \sim \text{Normal or Logistic with mean 0, and variance} \propto [\exp(\boldsymbol{\gamma}'\mathbf{w}_i)]^2$$

(In the logistic case, the true variance is scaled by $\pi^2/3$.)

NOTE: These heteroscedasticity models require individual data.

Request the model with heteroscedasticity with

```
PROBIT      ; Lhs  = dependent variable
or LOGIT    ; Rhs  = regressors in x
            ; Rh2 = list of variables in w
            ; Heteroscedasticity (or just ; Het) $
```

Other options and specifications for this model are the same as the basic model. (See Section E10.3 for discussion of variants of heteroscedasticity which can be accommodated with this model.) Two general options that are likely to be useful are

; Keep = name to retain predictions
; Prob = name to retain fitted probabilities

and the controls of the iterations and the amount of output.

NOTE: Do not include *one* in the Rh2 list. A constant in $\boldsymbol{\gamma}$ is not identified.

This model differs from the basic model only in the presence of the variance term. The output for this model is also the same, with the addition of the coefficients for the variance term. The initial OLS results are computed without any consideration of the heteroscedasticity, however.

Since the log likelihood for this model, unlike the basic model, is not globally concave, the default algorithm is BFGS, not Newton's method.

For purposes of hypothesis testing and imposing restrictions, the parameter vector is

$$\boldsymbol{\theta} = [\beta_1,...,\beta_K,\gamma_1,...,\gamma_L].$$

If you provide your own starting values, give the right number of values in exactly this order.

You can also use **WALD** and **; Test:** to test hypotheses about the coefficient vector. Finally, you can impose restrictions with

; Rst = ...

or **; CML: restrictions...**

NOTE: In principle, you can impose equality restrictions across the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with **; Rst = ...**, (i.e., force an element in $\boldsymbol{\beta}$ to equal one in $\boldsymbol{\gamma}$), but the results are unlikely to be satisfactory. Implicitly, the variables involved are of different scales, and this will place a rather stringent restriction on the model.

Use

; Robust

or **; Cluster = id variable or group size**

to request the sandwich style robust covariance matrix estimator or the cluster correction.

NOTE: There is no 'robust' covariance matrix for the logit or probit model that is robust to heteroscedasticity, in the form of the White estimator for the linear model. In order to accommodate heteroscedasticity in a binary choice model, you must model it explicitly..

NOTE: **; Maxit = 0** provides an easy way to test for heteroscedasticity with an LM test.

To test the hypothesis of homoscedasticity against the specification of this more general model, the following template can be used: (The model may be **LOGIT** if desired.)

```
NAMELIST    ; x  = ... the Rhs of the probit model
            ; w  = ... the Rh2 of the heteroscedasticity model $
CALC        ; m = Col(w) $
PROBIT      ; Lhs  = ...
            ; Rhs  = x $
PROBIT      ; Lhs  = ...
            ; Rhs  = x
            ; Rh2 = w ; Het
            ; Start = b, m_0
            ; Maxit = 0 $
```

This produces an LM statistic and (superfluously) reproduces the restricted model.

The results that are saved automatically are the same as for the basic model, that is, *b*, *varb*, and the scalars. In this case, *b* will contain the full set of estimates, with the slopes followed by the variance parameters, i.e., [**b**,**c**]. The *Last Model* labels for the **WALD** command are [*b_variable*, *c_variable*].

We note, this model may be rather weakly identified by the observed data, unless they are plentiful and the model is sharply consistent with the data. In fact, identification is not a problem, and the model is straightforward to estimate. But, one could argue that the specification problem addressed by this model is one of functional form rather than heteroscedasticity. That is, the model specification is arguably indistinguishable from one with a peculiar kind of conditional mean function, which, in turn, could be standing in for some other, perhaps reasonable, albeit nonlinear model. In addition, it is common for the estimated standard errors that are computed for this model to be quite large, as a result of a kind of multicollinearity – the high correlation of the derivatives of the log likelihood.

Application

To illustrate the model, we have refit the specification of the previous section with a variance term of the form Var[ε] = $[\exp(\gamma_1 \textit{female} + \gamma_2 \textit{working})]^2$. Since both of these are binary variables, this is equivalent to a groupwise heteroscedasticity model. The variances are 1.0, $\exp(2\gamma_1)$, $\exp(2\gamma_2)$ and $\exp(2\gamma_1+2\gamma_2)$ for the four groups. We have fit the original model without heteroscedasticity first. The second **LOGIT** command carries out the LM test of heteroscedasticity. The third command fits the full heteroscedasticity model.

```
INCLUDE     ; New ; year = 1994 $
NAMELIST    ; x = one,age,educ,married,hhninc,hhkids,female $
LOGIT       ; Lhs = doctor ; Rhs = x
            ; Partial Effects $
NAMELIST    ; w = female,working $
CALC        ; m = Col(w) $
LOGIT       ; Lhs = doctor ; Rhs = x
            ; Heteroscedasticity ; Rh2 = w
            ; Start = b,m_0
            ; Maxit = 0 $
LOGIT       ; Lhs  = doctor ; Rhs = x
            ; Heteroscedasticity ; Rh2 = w
            ; Partial Effects $
PARTIALS    ; Effects: female $
```

The model results have been rearranged in the listing below to highlight the differences in the models. Also, for convenience, some of the results have been omitted.

```
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function      -2085.33796
```

The LM statistic is included in the initial diagnostic statistics for the second model estimated.

```
LM Stat. at start values         3.11867
LM statistic kept as scalar     LMSTAT
```

These are the results for the model with homoscedastic disturbances.

```
Inf.Cr.AIC  = 4184.676 AIC/N =    1.239
Restricted log likelihood    -2169.26982
McFadden Pseudo R-squared       .0386913
```

These are the coefficient estimates for the two models.

Homoscedastic disturbances

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Characteristics in numerator of Prob[Y = 1]					
Constant	.14726	.25460	.58	.5630	-.35173	.64626
AGE	.01643***	.00384	4.28	.0000	.00891	.02395
EDUC	-.01965	.01608	-1.22	.2219	-.05117	.01188
MARRIED	.15536	.09904	1.57	.1167	-.03875	.34947
HHNINC	-.39474**	.17993	-2.19	.0282	-.74739	-.04208
HHKIDS	-.41534***	.08866	-4.68	.0000	-.58911	-.24157
FEMALE	.64274***	.07643	8.41	.0000	.49295	.79253

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Heteroscedastic disturbances

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Characteristics in numerator of Prob[Y = 1]					
Constant	.12927	.30739	.42	.6741	-.47320	.73174
AGE	.02036***	.00501	4.06	.0000	.01053	.03018
EDUC	-.02913	.01984	-1.47	.1421	-.06803	.00976
MARRIED	.19969	.12639	1.58	.1141	-.04803	.44742
HHNINC	-.36965*	.22169	-1.67	.0954	-.80414	.06485
HHKIDS	-.53029***	.12783	-4.15	.0000	-.78083	-.27974
FEMALE	1.24685***	.45754	2.73	.0064	.35009	2.14361
	Disturbance Variance Terms					
FEMALE	.44128*	.25946	1.70	.0890	-.06725	.94982
WORKING	.08459	.10082	.84	.4014	-.11300	.28219

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

These are the marginal effects for the two models. Note that the effects are also computed for the terms in the variance function. The explanatory text indicates the treatment of variables that appear in both the linear part and the exponential part of the probability.

```
+-------------------------------------------+
| Partial derivatives of probabilities with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Effects are the sum of the mean and var-  |
| iance term for variables which appear in  |
| both parts of the function.               |
+-------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Elasticity

Homoscedastic disturbances

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
Average partial effects for sample obs.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00352***     -.00205     4.29  .0000       .00191    .00512
    EDUC|    -.00421         .00058    -1.22  .2218     -.01096    .00254
 MARRIED|     .03357        -.00031     1.56  .1194     -.00868    .07582   #
  HHNINC|    -.08452**       .00044    -2.20  .0282     -.16000   -.00905
  HHKIDS|    -.09058***      .00027    -4.65  .0000     -.12876   -.05240   #
  FEMALE|     .13842***     -.00119     8.60  .0000      .10687    .16997   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Heteroscedastic disturbances

```
-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effects are the sum of the mean and var-
iance term for variables which appear in
both parts of the function.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
     AGE|     .00337***      .20980     3.84  .0001       .00165    .00509
    EDUC|    -.00482        -.08104    -1.47  .1404     -.01123    .00159
 MARRIED|     .03306         .03424     1.59  .1119     -.00769    .07380
  HHNINC|    -.06119        -.03975    -1.63  .1038     -.13492    .01254
  HHKIDS|    -.08778***     -.04969    -4.45  .0000     -.12640   -.04916
  FEMALE|     .20639***      .13969     5.09  .0000      .12687    .28592
        |Disturbance Variance Terms
  FEMALE|    -.07388        -.05000    -1.08  .2784     -.20747    .05972
 WORKING|    -.01416        -.01493     -.71  .4801     -.05347    .02514
        |Sum of terms for variables in both parts
  FEMALE|     .13252***      .08969     3.52  .0004      .05875    .20629
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The partial effects for the heteroscedasticity model are computed at the means of the variables. It is possible to obtain average partial effects by using the **PARTIAL EFFECTS** program rather than the built in marginal effects routine. The following shows the results for *female*, which appears in both parts of the model.

PARTIAL EFFECTS ; Effects: female $

```
---------------------------------------------------------------------
Partial Effects  Analysis for Heteros. Logit Prob.Function
---------------------------------------------------------------------
Effects on function with respect to FEMALE
Results are computed by average over sample observations
Partial effects for binary var FEMALE   computed by first difference
---------------------------------------------------------------------
df/dFEMALE         Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
APE. Function       .13430     .01653    8.12      .10190      .16669
```

These are the summaries of the predictions of the two estimated models. The performance of the two models in terms of the simple count of correct predictions is almost identical – the heteroscedasticity model correctly predicts three observations more than the homoscedasticity model. The mix of correct predictions is very different, however.

Homoscedastic disturbances

```
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |     82 (  2.4%)|   1073 ( 31.8%)|   1155 ( 34.2%)|
|  1   |     85 (  2.5%)|   2137 ( 63.3%)|   2222 ( 65.8%)|
+------+----------------+----------------+----------------+
|Total |    167 (  4.9%)|   3210 ( 95.1%)|   3377 (100.0%)|
+------+----------------+----------------+----------------+
```

Heteroscedastic disturbances

```
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |    131 (  3.9%)|   1024 ( 30.3%)|   1155 ( 34.2%)|
|  1   |    139 (  4.1%)|   2083 ( 61.7%)|   2222 ( 65.8%)|
+------+----------------+----------------+----------------+
|Total |    270 (  8.0%)|   3107 ( 92.0%)|   3377 (100.0%)|
+------+----------------+----------------+----------------+
```

E27.12 Estimation Methods and Technical Details

This section will document the estimation methods used for fitting the binary choice models, and some options available for controlling these. We also lay out some of the technical background for the models.

E27.12.1 Maximum Likelihood Estimation

With only a few exceptions, the estimation technique used for fitting the binary choice models is maximum likelihood. For the parametric models, let y_i denote the observed individual outcome, and p_i denote an observed proportion in the grouped data case. The log likelihood functions for the two cases are

$$\log L = \Sigma_i\, w_i[(1 - y_i)\log(1 - F_i) + y_i \log F_i]$$

and

$$\log L = \Sigma_i\, w_i[n_i(1 - p_i)\log F_i + n_i\, p_i \log F_i]$$

where n_i is usually one, but may be the number of observations in the 'group,' w_i is a general weight, which may always be applied in estimation, and $F_i = F(\boldsymbol{\beta}'\mathbf{x}_i)$. Estimates of the model parameters are obtained by maximizing the log likelihood. In most cases, Newton's method is the most effective algorithm, though all others provided by *LIMDEP* may be used. The probit, logit, Gompertz and complementary log log models have globally concave log likelihoods, and estimation is generally routine. Unless the data are very badly conditioned, all of the estimators should converge uniformly and quite rapidly; none present particularly difficult problems of computation. The Burr model is typically more difficult to estimate because the log likelihood is not globally concave.

Asymptotic standard errors may be computed in a variety of ways. In most cases, the estimated asymptotic covariance matrix will be the negative inverse of the actual Hessian. For the models estimated by Newton's method, the covariance matrix for the coefficients is estimated with the second derivatives of the log likelihood. For the models computed with DFP, the summed outer products of the first derivatives of the log likelihood, the BHHH (Berndt, et al., (1974)), estimator is usually used instead. The Burr model is an example in which Newton's method is generally too crude without a line search.

The following results for binary choice models are widely known for the probit and logit models, but, it turns out, are completely general, and apply to the remainder as well. Some minor modification is required for models which contain ancillary parameters, such as the Burr model and the heteroscedasticity model discussed below, but nonetheless, the results are general. Also, the results below are extended to the grouped data case with only trivial modification. Denote by z_i the argument of F_i, $\boldsymbol{\beta}'\mathbf{x}_i$, and denote by f_i the derivative of F_i with respect to z_i – f_i will generally be the density function corresponding to CDF F_i. Then, for the individual data case,

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} w_i \left\{ -f_i \frac{1-y_i}{1-F_i} + f_i \frac{y_i}{F_i} \right\} \mathbf{x}_i \quad = \quad \sum_{i=1}^{n} w_i\, f_i \left\{ \frac{y_i}{F_i} - \frac{1-y_i}{1-F_i} \right\} \mathbf{x}_i \,.$$

Expressions for f_i for the various models estimated appear at the beginning of Section E26.2. Since $E[y_i] = F_i$, it follows obviously that the expected first derivative is zero, as would be required for a regular maximum likelihood problem.

The multiplier of $\mathbf{x}_i$ without the weight, w_i, is the generalized residual noted in Section E27.4.4. The specific forms of these terms are obtained as follows: Define

$$q_i = 2y_i - 1 = \text{-1 if } y_i \text{ equals 0, +1 if } y_i \text{ equals 1,}$$

$$a_i = \boldsymbol{\beta}'\mathbf{x}_i,$$

$$c_i = \exp(a_i),$$

$$d_i = \exp(-a_i)$$

Then, the generalized residuals are

Probit: $\dfrac{q_i\phi(q_i a_i)}{\Phi(q_i a_i)}$

Logit: $y_i - \Lambda(a_i)$

Comploglog: $c_i\left(\dfrac{y_i}{1+\exp(-c_i)} - 1\right)$

Gompertz: $\left(\dfrac{y_i - \exp(-d_i)}{1-\exp(-d_i)}\right)$

Arctangent: $\dfrac{q_i}{y_i F_i + (1-y_i)(1-F_i)}\left(\dfrac{2c_i}{\pi(1+c_i^2)}\right)$

Burr: $\left(y_i \;-\; \dfrac{(1-y_i)\Lambda_i^{\gamma}}{1-\Lambda_i^{\gamma}}\right)\gamma(1-\Lambda_i)$

The actual Hessian is

$$\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \sum_{i=1}^{n} w_i\left[\, f_i'\left\{\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}\right\} - f_i\left\{\frac{y_i f_i}{F_i^2} + \frac{(1-y_i) f_i}{(1-F_i)^2}\right\}\right]\mathbf{x}_i\mathbf{x}_i'$$

Computation of the Hessian for Newton's method requires expressions for f_i'. For the five models, not including the Burr model – this is considered below – these are

Probit: $-z_i\phi_i$

Logit: $\Lambda_i(1 - \Lambda_i)(1 - 2\Lambda_i)$

Extreme Value: $\lambda_i \exp(-\lambda_i)(1 - \lambda_i)$, with $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$

Gompertz: $\lambda_i \exp(-\lambda_i)(\lambda_i - 1)$, with $\lambda_i = \exp(-\boldsymbol{\beta}'\mathbf{x}_i)$.

Arctangent: $(2/\pi)[\lambda_i^2/(1+\lambda_i^2)^2](\lambda_i^2-2\lambda_i+1)$

The method of scoring can be used as well, by taking the expectation of the Hessian. Since y_i is the random variable of the expectation operator, and y_i enters the Hessian linearly, the surprisingly simple result which emerges is that for all the models,

$$E\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] = \sum_{i=1}^{n} -w_i \left[\frac{f_i^2}{F_i(1-F_i)}\right]\mathbf{x}_i\mathbf{x}_i'$$

which is the widely cited result. The third approach, for purposes of computing the BHHH estimator of the asymptotic covariance matrix is to use the outer product of gradients, or OPG estimator. This would be based on the inverse of

$$OPG = \sum_{i=1}^{n} w_i \left[f_i \left\{ \frac{y_i}{F_i} - \frac{1-y_i}{1-F_i} \right\} \right]^2 \mathbf{x}_i\mathbf{x}_i'.$$

Once again, this simplifies considerably. By expanding the square and using the results that y_i and $(1-y_i)$ both equal their squares, and $y_i\,(1-y_i) = 0$, the end result is simply

$$OPG = \sum_{i=1}^{n} w_i \left[\frac{f_i(y_i - F_i)}{F_i(1-F_i)} \right]^2 \mathbf{x}_i\mathbf{x}_i'.$$

As noted earlier, because of the extra parameter, γ, the Burr model does not fit into these neat simplifications. For this model, only the first derivatives are used in estimation by the BFGS algorithm and in computing the asymptotic covariance matrix by the OPG method. The first derivatives of the log likelihood for the Burr model are

$$\frac{\partial \log L}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} = \sum_{i=1}^{n} w_i \left(y_i - \frac{(1-y_i)\Lambda_i^{\gamma}}{1-\Lambda_i^{\gamma}} \right) \begin{bmatrix} \gamma(1-\Lambda_i)\mathbf{x}_i \\ \log \Lambda_i \end{bmatrix}$$

The OPG estimator is formed directly by using the summed outer products. No simplification is possible. The Hessian for this model is

$$\frac{\partial^2 \log L}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} \partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} = \sum_{i=1}^{n} w_i \left(y_i - \frac{(1-y_i)\Lambda_i^{\gamma}}{1-\Lambda_i^{\gamma}} \right) \begin{bmatrix} -\gamma\Lambda_i(1-\Lambda_i)\mathbf{x}_i\mathbf{x}_i' & (1-\Lambda_i)\mathbf{x}_i \\ (1-\Lambda_i)\mathbf{x}_i' & 0 \end{bmatrix}$$

$$-\sum_{i=1}^{n} w_i \left(\frac{(1-y_i)\Lambda_i^{\gamma}}{(1-\Lambda_i^{\gamma})^2} \right) \begin{bmatrix} \gamma(1-\Lambda_i^{\gamma})\mathbf{x}_i \\ \log \Lambda_i \end{bmatrix} \begin{bmatrix} \gamma(1-\Lambda_i^{\gamma})\mathbf{x}_i \\ \log \Lambda_i \end{bmatrix}'$$

The expected Hessian is considerably simpler as the first term has expectation zero. Once again, y_i enters linearly in the second. Using its expectation, Λ_i^γ, the second term reduces the expression to

$$E\left[\frac{\partial^2 \log L}{\partial\begin{pmatrix}\boldsymbol{\beta}\\ \gamma\end{pmatrix}\partial\begin{pmatrix}\boldsymbol{\beta}\\ \gamma\end{pmatrix}}\right] = -\sum_{i=1}^{n} w_i\left(\frac{\Lambda_i^\gamma}{1-\Lambda_i^\gamma}\right)\begin{bmatrix}\gamma(1-\Lambda_i^\gamma)\mathbf{x}_i\\ \log\Lambda_i\end{bmatrix}\begin{bmatrix}\gamma(1-\Lambda_i^\gamma)\mathbf{x}_i\\ \log\Lambda_i\end{bmatrix}'$$

(Note that for the special case of $\gamma = 1$, this reduces to the familiar result for the logit model.)

Ancillary parameters such as the slopes in the heteroscedasticity models and γ in the Burr model are started at zero and one, respectively. In general, where there is no natural data/model based starting value, we use a value which as a restriction produces a simpler model. Thus, the choices noted for the heteroscedastic and the Burr models produce the homoscedastic models and the binary logit model.

E27.12.2 Minimum Chi Squared Estimation with Grouped Data

In the grouped data cases, weighted least squares is an alternative estimation strategy. The approach uses the inverse transformation of the probability function. Let π_i be the true value of F_i. Then, we write

$$F^{-1}(\pi_i) = \boldsymbol{\beta}'\mathbf{x}_i \text{ and } F^{-1}(p_i) = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i.$$

Expand the former in a linear Taylor series to obtain

$$F^{-1}(p_i) = F^{-1}(\pi_i) + (p_i - \pi_i)\frac{dF^{-1}(\pi_i)}{d\pi_i}.$$

The latter derivative is just the reciprocal of the density,$(1/f_i)$. The variance of the right hand side is, therefore, $\text{Var}[p_i - \pi_i]/(f_i)^2$, which suggests a generalized least squares approach. In each case, $\text{Var}[p_i - \pi_i]$ is $\pi_i(1 - \pi_i)/n_i$, which in the context of our model gives, finally,

$$\text{Var}[F^{-1}(p_i)] = \frac{F_i(1-F_i)}{n_i f_i^2}$$

With grouped data, then, one might use an iterative strategy. Given starting values for $\boldsymbol{\beta}$, compute the weights implied by the variance function above, then compute weighted least squares regression of $F^{-1}(p_i)$ on $\mathbf{x}_i$. The iteration can be reentered if desired. Once again, the Burr model does not lend itself to this approach, but for the other five, it is straightforward using the inverse transformations

Probit:	$F^{-1}(p_i)$	$= \Phi^{-1}(p_i)$ (must be approximated)
Logit:	$F^{-1}(p_i)$	$= \log[p_i/(1-p_i)]$
Extreme Value:	$F^{-1}(p_i)$	$= -\log(\log(1 - p_i))$
Gompertz:	$F^{-1}(p_i)$	$= -\log(-\log(p_i)$
Arctangent	$F^{-1}(p_i)$	$= \pi/2 \tan(p_i)$

Minimum chi squared estimators have the same properties as, but are numerically different from the maximum likelihood estimators.

Starting Values

The precise set of values to be provided for **; Start** varies from one model to another. Starting values may be provided for the model. Those used by the program if you do not provide your own are as follows:

- **Individual data:** Simple least squares regression of y on **X**
- **Grouped data:** The minimum chi squared, weighted least squares estimates.

The first round of these weighted least squares estimators, using p_i as F_i in the weights, is computed to obtain the starting values for the MLEs for each of the four models mentioned. The least squares results at the beginning of the output (when requested) will contain an indication that this has been the computation. Iteration of the minimum chi squared estimator is not continued, as the starting values are simply used to continue the maximum likelihood estimation. (Note, as well, that the MCS estimator has a problem not shared by the MLE. If any of the proportions are zero or one, the weights will not be computable. Authors have suggested various fixes; the most common is simply to use a small value such as $1/n$ or $1 - 1/n$ as appropriate.

Standard Errors for Marginal Effects Based on the Delta Method

Standard errors are computed using the delta method. Let $\boldsymbol{\delta}$ denote the marginal effects, and **d** denote the sample estimate. The asymptotic covariance matrix is estimated with

$$\text{Asy.Var}[\mathbf{d}] = \mathbf{G} \times \text{Asy.Var}\left[\left[\hat{\boldsymbol{\beta}}\right]\right] \times \mathbf{G}'$$

where **G** is the matrix of derivatives,

$$\mathbf{G} = \frac{\partial\boldsymbol{\delta}}{\partial\boldsymbol{\beta}'} = f(\boldsymbol{\beta}'\mathbf{x})\mathbf{I} + [\mathrm{d}f(\boldsymbol{\beta}'\mathbf{x})/\mathrm{d}(\boldsymbol{\beta}'\mathbf{x})]\boldsymbol{\beta}\mathbf{x}'$$

evaluated at $\hat{\boldsymbol{\beta}}$ and the particular vector (the vector of sample means). (In the Burr model, there is an extra column in **G** to account for the estimate of γ.) For a dummy variable, the asymptotic standard error must be changed slightly. This is accomplished simply by changing the appropriate row of **G** to

$$\mathbf{G}_z = [f(\boldsymbol{\beta}'\mathbf{x} + \alpha z)]\begin{pmatrix}\mathbf{x}\\1\end{pmatrix}' - [f(\boldsymbol{\beta}'\mathbf{x} + \alpha z)]\begin{pmatrix}\mathbf{x}\\0\end{pmatrix}'$$

NOTE: We are frequently asked about the hypothesis test that the marginal effects equal zero, and in particular, about the fairly common case in which a marginal effect is ‘insignificant’ when the corresponding coefficient is ‘significant.’ Our own assessment is that significance tests of the influence of a variable should be based on the coefficients, not the marginal effects. The latter is a function – and a highly nonlinear one at that – of all the coefficients in the model, and the hypothesis that this function equals zero is not equivalent to the hypothesis that the coefficient is zero or that the variable in question is not a significant determinant of the outcome.

Average Partial Effects vs. Partial Effects at the Means

Some authors (e.g., Wooldridge (1999)) argue that one *should* compute marginal effects by averaging the individual estimates, rather than computing the partial effects once at the means of the variables. Save for small sample variation, the difference in these two results is likely to be small, as suggested by the example below.

```
-----------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
      AGE           .00402     .00082    4.92      .00242       .00562
      HHNINC      -.08666     .03911    2.22     -.16331      -.01001
   *  HHKIDS      -.08524     .01968    4.33     -.12382      -.04667
      EDUC        -.00779     .00348    2.24     -.01461      -.00097
   *  MARRIED      .03279     .02159    1.52     -.00952       .07510
-----------------------------------------------------------------------

-----------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Computed at Data Means
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
      AGE           .00410     .00084    4.87      .00245       .00574
      HHNINC      -.08839     .03997    2.21     -.16673      -.01005
      HHKIDS      -.08556     .01959    4.37     -.12396      -.04717
      EDUC        -.00795     .00356    2.23     -.01492      -.00097
      MARRIED      .03319     .02172    1.53     -.00937       .07576
-----------------------------------------------------------------------
```

The built in **; Partial Effects** option for the binary choices uses the average partial effects in most cases, but partial effects at the means in a few cases such as the heteroscedastic models. The **PARTIAL EFECTS** command computes average partial effects by default in all cases but provides an option for you to choose to evaluate the effects at the means, instead.

E27.12.3 Binary Choice Models with Heteroscedasticity

The log likelihood function for the binary choice model with exponential heteroscedasticity is

$$\log L = \Sigma_i \log F(a_i),\ F = \Phi \text{ or } \Lambda,$$

where

$$a_i = (2y_i - 1)\boldsymbol{\beta}'\mathbf{x}_i \times \exp(-\boldsymbol{\gamma}'\boldsymbol{w}_i\ ).$$

We are taking advantage of the symmetry of the probit and logit functions to simplify the function. Let

$$\boldsymbol{\theta} = \text{the full parameter vector, } [\boldsymbol{\beta}',\boldsymbol{\gamma}']',\text{ in which } \boldsymbol{\gamma} \text{ may be } \mathbf{0}.$$

The derivatives are as follows, where we use the notation $\mathbf{a}_{i\theta}$ for $\partial a_i/\partial\boldsymbol{\theta}$:

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} = \Sigma_i [f(a_i)/F(a_i)]\partial a_i/\partial\boldsymbol{\theta} = \Sigma_i g_i \mathbf{a}_{i\theta}.$$

where $g_i = \phi(a_i)/\Phi(a_i)$ for the probit model ($f_i = \phi(a_i)$),

$g_i = (1 - \Lambda_i)$ for the logit model ($f_i = \Lambda_i(1-\Lambda_i)$),

and
$$\mathbf{a}_{i\theta} = (2y_i - 1)\exp(-\gamma'\mathbf{w}_i)\begin{bmatrix}\mathbf{x}_i \\ -(\boldsymbol{\beta}'\mathbf{x}_i)\mathbf{w}_i\end{bmatrix}$$

Using a similar subscripting notation for second derivatives, we have

$$\frac{\partial \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \Sigma_i h_i \mathbf{a}_{i\theta}\mathbf{a}_{i\theta'} + g_i \mathbf{a}_{i\theta\theta'}$$

where
$$h_i = \frac{\partial g_i}{\partial a_i} = \begin{cases} -a_i g_i - g_i^2 & \text{for the probit model} \\ -\Lambda_i(1-\Lambda_i) & \text{for the logit model}\end{cases}$$

and
$$a_{i\beta\beta'} = \mathbf{0},$$
$$a_{i\beta\gamma'} = (2y_i - 1)\exp(-\boldsymbol{\gamma}'\mathbf{w}_i)\mathbf{x}_i(-\mathbf{w}_i'),$$
$$a_{i\gamma\gamma'} = (2y_i - 1)(\boldsymbol{\beta}'\mathbf{x}_i)\exp(-\boldsymbol{\gamma}'\mathbf{w}_i)\mathbf{w}_i\mathbf{w}_i'.$$

There are two sets of marginal effects in this model:

$$\frac{\partial E[y_i\,|\,\mathbf{x}_i]}{\partial \mathbf{x}_i} = f(a_i)\frac{\partial a_i}{\partial \mathbf{x}_i} = f(a_i)\exp(-\boldsymbol{\gamma}'\mathbf{w}_i)\boldsymbol{\beta}$$

$$\frac{\partial E[y_i\,|\,\mathbf{x}_i]}{\partial \mathbf{w}_i} = f(a_i)\frac{\partial a_i}{\partial \mathbf{w}_i} = -f(a_i)\exp(-\boldsymbol{\gamma}'\mathbf{w}_i)(\boldsymbol{\beta}'\mathbf{x}_i)\boldsymbol{\gamma}$$

If any variables appear in both **x** and **w**, then the marginal effect of that variable on the conditional mean is the sum of the two parts.

E28: Tests and Restrictions in Models for Binary Choice

E28.1 Introduction

We define models in which the response variable being described is inherently discrete as qualitative response (QR) models. Chapters E26 and E27 presented the model formulation and estimation and analysis tools. This chapter will detail some aspects of hypothesis testing. Most of these results are generic, and will apply in other models as well. The hypothesis tests are general restrictions on parameters. Section E28.3 considers two broader specification tests. Section E28.4 documents how to impose restrictions on the maximum likelihood estimator.

E28.2 Testing Hypotheses

The full set of options is available for testing hypotheses and imposing restrictions on the binary choice models. In using these, the set of parameters is

$$\beta_1, ..., \beta_K \text{ plus } \gamma \text{ for the Burr model}$$

In the parametric models, hypotheses can be done with the standard trinity of tests: Wald, likelihood ratio and Lagrange Multiplier. All three are particularly straightforward for the binary choice models.

E28.2.1 Wald Tests

Wald tests are carried out in two ways, with the **; Test:** specification in the model command and by using the **WALD** command after fitting the model. The former is used for linear restrictions. The **WALD** command is more general and allows for tests of nonlinear restrictions on parameters.

The Wald statistic is computed using the estimates of an unrestricted model. The hypothesis implies a set of restrictions

$$H_0: \mathbf{c}(\boldsymbol{\beta}) = \mathbf{0}.$$

(This may involve linear distance from a constant, such as $2\beta_3 - 1.2 = 0$. The preceding formulation is used to achieve the full generality that *LIMDEP* allows.) The Wald statistic is computed by the formula

$$W = \mathbf{c}(\hat{\boldsymbol{\beta}})'\left[\mathbf{G}(\hat{\boldsymbol{\beta}})\left\{Est.Asy.Var(\hat{\boldsymbol{\beta}})\right\}\mathbf{G}(\hat{\boldsymbol{\beta}})'\right]^{-1}\mathbf{c}(\hat{\boldsymbol{\beta}})$$

where $$\mathbf{G}(\hat{\boldsymbol{\beta}}) = \frac{\partial \mathbf{c}(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}'}$$

and $\hat{\boldsymbol{\beta}}$ is the vector of estimated parameters.

You can request Wald tests of simple restrictions by including the request in the model command. For example:

```
PROBIT     ; Lhs = doctor
           ; Rhs = one,age,educ,married,hhninc,hhkids
           ; Test:  age + educ = 0,
                    married = 0 ,
                    hhninc + 2*hhkids = -.3  $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                DOCTOR
Log likelihood function    -17670.94233
Restricted log likelihood  -18019.55173
Chi squared [   5 d.f.]       697.21881
Significance level               .00000
McFadden Pseudo R-squared      .0193462
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =35353.885 AIC/N =    1.294
Hosmer-Lemeshow chi-squared = 105.22799
P-value=  .00000 with deg.fr. =       8
Wald test of  3 linear restrictions
Chi-squared =      26.06, P value   =    .00001
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .15500***      .05652     2.74  .0061      .04423    .26577
     AGE|     .01283***      .00079    16.24  .0000      .01129    .01438
    EDUC|    -.02812***      .00350    -8.03  .0000     -.03498   -.02125
 MARRIED|     .05226**       .02046     2.55  .0106      .01216    .09237
  HHNINC|    -.11643**       .04633    -2.51  .0120     -.20723   -.02563
  HHKIDS|    -.14118***      .01822    -7.75  .0000     -.17689   -.10548
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Note that the results reported are for the unrestricted model, and the results of the Wald test are reported with the initial header information. To fit the model subject to the restriction, we change **; Test:** in the command to **; CML:** with the following results:

```
PROBIT     ; Lhs =  doctor
           ; Rhs =  one,age,educ,married,hhninc,hhkids
           ; CML: age + educ = 0,
                  married = 0 ,
                  hhninc + 2*hhkids = -.3  $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                DOCTOR
Log likelihood function      -2125.57999
Restricted log likelihood    -2169.26982
Chi squared [   2 d.f.]         87.37966
Significance level                .00000
McFadden Pseudo R-squared       .0201403
Estimation based on N =   3377, K =   3
Inf.Cr.AIC  = 4257.160 AIC/N =     1.261
Linear constraints imposed             3
Hosmer-Lemeshow chi-squared =   20.93392
P-value=  .00733 with deg.fr. =        8
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .04583        .06144      .75  .4557     -.07458    .16624
     AGE|     .01427***     .00192     7.44  .0000      .01052    .01803
    EDUC|    -.01427***     .00192    -7.44  .0000     -.01803   -.01052
 MARRIED|        0.0    .....(Fixed Parameter).....
  HHNINC|    -.06304        .07079     -.89  .3731     -.20178    .07569
  HHKIDS|    -.11848***     .03539    -3.35  .0008     -.18785   -.04911
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

When the restrictions are built into the estimator with CML, the information reported is only that the restrictions were imposed. The results of the Wald or LR test cannot be reported because the unrestricted model is not computed.

E28.2.2 Likelihood Ratio Tests

Use the log likelihood functions from both restricted and unrestricted models. Log likelihood functions are saved automatically by the estimators. Do keep in mind that these are overwritten each time – the scalar *logl* gets replaced by each model command. Your general strategy for carrying out a likelihood ratio test would be

Model name ; ... - specifies the unrestricted model
CALC ; lu = logl $ Capture log likelihood function
Model name ; ... - specifies the restricted model
CALC ; lr = logl
; List ; chisq = 2*(lu - lr)
; 1 - Chi(chisq, degrees of freedom) $

You must supply the degrees of freedom. If the result of the last line is less than your significance level – usually 0.05 – then, the null hypothesis of the restriction would be rejected. Here are two examples: We continue to examine the German health care data. For purposes of these tests, just for the illustrations, we will switch to a probit model.

Simple Linear Restriction

The following tests the pair of linear restrictions suggested above. Looking at the unrestricted results from earlier, the restrictions don't look like they are going to pass. The results bear this out.

```
SAMPLE      ; All $
NAMELIST    ; x = one,age,educ,married,hhninc,hhkids $
LOGIT       ; Lhs = doctor ; Rhs = x $
CALC        ; lu = logl $
LOGIT       ; Lhs = doctor ; Rhs = x
            ; Rst = b0, b1, b1, 0, b2, b3 $
CALC        ; lr = logl
            ; List ; chisq = 2*(lu - lr) ; 1 - Chi(chisq,2) $
```

```
[CALC] CHISQ   =    158.9035080
[CALC] *Result*=       .0000000
Calculator: Computed   3 scalar results
```

Homogeneity Test

We are frequently asked about this. The sample can be partitioned into a number of subgroups. The question is whether it is valid to pool the subgroups. Here is a general strategy that is the maximum likelihood counterpart to the Chow test for linear models: Define a variable, say, *group*, that takes values 1,2,...,*G*, that partitions the sample. This is a stratification variable. The test statistic for homogeneity is

$$\chi^2 = 2[(\Sigma_{groups} \text{ log likelihood for the group}) - \text{log likelihood for the pooled sample}]$$

The degrees of freedom is G-1 times the number of coefficients in the model.

Create the group variable.

```
SAMPLE      ; Pooled sample ... however defined ... $
Model name  ; ... ; Quiet $  Specify the appropriate model.  Suppress the output.
CALC        ; chisq = -2*logl ; df = -kreg $
```

Automate the model fitting estimation, and accumulate the statistic.

```
PROC
  INCLUDE    ; New ; Group = i $
  Model name ; ... ; Quiet $  Specify the same model.  Suppress the output.
  CALC       ; chisq = chisq + 2*logl ; df = df + kreg $
ENDPROC
```

Determine the number of groups.

```
CALC        ; g = Max(group) $
```

Estimate the model once for each group.

```
EXEC        ; i = 1,g $
CALC        ; List ; chisq ; df ; 1 - Chi(chisq,df)  $
```

This procedure produces only the output of the last **CALC** command, which will display the test statistic, the degrees of freedom and the p value for the test.

To illustrate, we'll test the hypothesis that the same probit model for doctor visits applies to both men and women. This command suppresses all output save for the actual test of the hypothesis.

```
NAMELIST   ; x = one,age,educ,married,hhninc,hhkids $
PROBIT     ; If [ female = 0] ; Lhs = doctor ; Rhs = x ; Quiet $
CALC       ; l0 = logl $
PROBIT     ; If [ female = 1] ; Lhs = doctor ; Rhs = x ; Quiet $
CALC       ; l1 = logl $
PROBIT     ; Lhs = doctor ; Rhs = x ; Quiet $
CALC       ; l01 = logl ; List
           ; chisq = -2*(l01 - l0 - l1)
           ; df = 2*kreg ; pvalue = 1 - Chi(chisq,df) $
```

The results of the chi squared test strongly reject the homogeneity restriction.

```
[CALC] CHISQ   =    549.3141072
[CALC] DF      =     12.0000000
[CALC] PVALUE  =       .0000000
Calculator: Computed   4 scalar results
```

E28.2.3 Lagrange Multiplier Tests

The third procedure available for testing hypotheses is the Lagrange Multiplier, or LM approach. The Lagrange Multiplier statistic is computed as a Wald statistic for testing the hypothesis that the derivatives of the log likelihood are zero when evaluated at the restricted maximum likelihood estimator;

$$LM = \mathbf{g}(\hat{\boldsymbol{\beta}}_R)'\left[Est.Asy.Var\left\{\mathbf{g}(\hat{\boldsymbol{\beta}}_R)\right\}\right]^{-1}\mathbf{g}(\hat{\boldsymbol{\beta}}_R)$$

where $\hat{\boldsymbol{\beta}}_R$ = MLE of the parameters of the model, with restrictions imposed

$\mathbf{g}(\hat{\boldsymbol{\beta}}_R)$ = derivatives of log likelihood of full model, evaluated at $\hat{\boldsymbol{\beta}}_R$

The estimated asymptotic covariance matrix of the gradient is any of the usual estimators of the asymptotic covariance matrix of the coefficient estimator, negative inverse of the actual or expected Hessian, or the BHHH estimator based on the first derivatives only.

Your strategy for carrying out LM tests with *LIMDEP* is as follows:

Step 1. Obtain the restricted parameter vector. This may involve an unrestricted parameter vector in some restricted model, padded with some zeros, or a similar arrangement.

Step 2. Set up the full, unrestricted model as if it were to be estimated, but include in the command

```
; Start  = restricted parameter vector
; Maxit  = 0
```

The rest of the procedure is automated for you. The **; Maxit = 0** specification takes on a particular meaning when you also provide a set of starting values. It implies that you wish to carry out an LM test using the starting values.

To demonstrate, we will carry out the test of the hypothesis

$$\beta\_age + \beta\_educ = 0$$
$$\beta\_married = 0$$
$$\beta\_hhninc + \beta\_hhkids = -.3$$

that we tested earlier with a Wald statistic, now with the LM test. The commands would be as follows:

```
PROBIT    ; Lhs = doctor
          ; Rhs = one,age,educ,married,hhninc,hhkids
          ; CML: age+educ = 0, married = 0 , hhninc + 2*hhkids = -.3 $
PROBIT    ; Lhs = doctor
          ; Rhs = one,age,educ,married,hhninc,hhkids
          ; Maxit = 0 ; Start = b $
```

The results of the second model command provide the Lagrange multiplier statistic. The value of 26.06032 is the same as the Wald statistic computed earlier, 26.06.

```
Maximum of     0 iterations. Exit iterations with status=1.
Maxit = 0. Computing LM statistic at starting values.
No iterations computed and no parameter update done.
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable               DOCTOR
LM Stat. at start values       26.06032
LM statistic kept as scalar      LMSTAT
Log likelihood function    -17683.96508
Restricted log likelihood  -18019.55173
Chi squared [   5 d.f.]       671.17331
Significance level               .00000
McFadden Pseudo R-squared      .0186235
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =35379.930 AIC/N =    1.295
Model estimated: Jun 13, 2011, 19:40:02
Hosmer-Lemeshow chi-squared = 132.57086
P-value=  .00000 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.06593        .05655    -1.17  .2437     -.17678    .04491
     AGE|     .01484***     .00079    18.76  .0000      .01329    .01639
    EDUC|    -.01484***     .00351    -4.23  .0000     -.02171   -.00796
 MARRIED|       0.0         .02049      .00 1.0000 -.40156D-01  .40156D-01
  HHNINC|    -.09655**      .04636    -2.08  .0373     -.18741   -.00568
  HHKIDS|    -.10173***     .01821    -5.59  .0000     -.13742   -.06603
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

To complete the trinity of tests, we can carry out the likelihood ratio test, which we could do as follows:

```
PROBIT      ; Quiet ; Lhs = doctor
            ; Rhs = one,age,educ,married,hhninc,hhkids
            ; CML: b(2) + b(3) = 0, b(4) = 0, b(5) + b(6) = -.3 $
CALC        ; lr = logl $
PROBIT      ; Quiet ; Lhs = doctor
            ; Rhs = one,age,educ,married,hhninc,hhkids $
CALC        ; lu = logl ; List
            ; lrstat  = 2*(lu – lr) $
```

The result of the computation (which displays only the last statistic) is

```
[CALC] LRSTAT  =     26.0455042
Calculator: Computed   2 scalar results
```

The value of 26.0455 differs only trivially from the other values. This is actually not surprising, since they should all converge to the same statistic, and the sample in use here is very large.

E28.3 Two Specification Tests

The following are two specialized tests for the probit model, one for testing which of two competing models appears to be appropriate, and one test against the hypothesis of normality that underlies the probit model.

E28.3.1 A Test for Nonnested Probit Models

Davidson and MacKinnon (1993) present a test of the nonnested hypothesis that an alternative set of variables, $\mathbf{z}_i$, is the appropriate one for the structural equation of the probit model.

$$\text{Testing } y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon \text{ vs. } y^* = \mathbf{z}'\boldsymbol{\gamma} + u$$

```
NAMELIST    ; x  = the independent variables
            ; z  = the competing list of independent variables $
CREATE      ; y  = the dependent variable $
PROBIT      ; Quiet ; Lhs = y ; Rhs = x $
CREATE      ; xbeta  = x’b; fx = N01(xbeta) ; px = Phi(xbeta)
            ; v  = Sqr(px*(1-px)) ; dev = (y - px) / v
            ; xv = fx*xbeta / v $
PROBIT      ; Quiet ; Lhs = y ; Rhs = z $
CREATE      ; pz = Phi(z’b) ; test = (px - pz) / v $
REGRESS     ; Lhs = dev ; Rhs  = xv,test $
```

The test is carried out by referring the t ratio on *test* to the t table. A value larger than the critical value argues in favor of *z* as the correct specification. For example, the following tests for which of two specifications of the right hand side of the probit model is preferred.

```
NAMELIST   ; x = one,age,educ,married,hhninc,hhkids,self
           ; z = one,age,educ,married,hhninc,female,working $
CREATE     ; y = doctor $
```

The remaining commands are identical.

The essential regression results are as follows. We also reversed the roles of *x* and *z*. Unfortunately, as often happens in specifications, the results are contradictory.

```
--------+------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
     DEV|  Coefficient       Error         z   |z|>Z*          Interval
--------+------------------------------------------------------------------
      XV|     .04569**       .01985      2.30  .0214      .00678     .08459
    TEST|    -.79517***      .03995    -19.90  .0000    -.87348    -.71687
--------+------------------------------------------------------------------
      XV|     .04668**       .02033      2.30  .0217      .00684     .08652
    TEST|    -.26126***      .04273     -6.11  .0000    -.34500    -.17751
```

The t ratio of -19.9 in the first regression argues in favor of *z* as the appropriate specification. But, the also significant t ratio of -6.11 in the second argues in favor of *x*.

E28.3.2 A Test for Normality in the Probit Model

The second test is a Lagrange multiplier test against the null hypothesis of normality in the probit model. (The test was developed in Bera, Jarque and Lee (1984).) As usual in normality tests, the statistic is computed by comparing the third and fourth moments of an underlying variable to their expected value under normality. The computations are as follows, where *i* indicates the *i*th observation:

$$a_i = \mathbf{x}_i'\boldsymbol{\beta}$$

$$\phi_i = \phi(a_i)$$

$$\Phi_i = \Phi(a_i)$$

$$d_i = \phi_i\,(y_i - \Phi_i) \,/\, [\Phi_i(1 - \Phi_i)]$$

$$c_i = \phi_i^2 \,/\, [\Phi_i(1 - \Phi_i)]$$

$$m3_i = -1/2(a_i^2 - 1)$$

$$m4_i = 1/4\,(a_i\,(a_i^2 + 3))$$

$$\mathbf{z}_i = (\mathbf{x}_i', m3_i, m4_i)'$$

Then,

$$LM = \left(\sum\nolimits_{i=1}^{N} d_i\mathbf{z}_i\right)'\left(\sum\nolimits_{i=1}^{N} c_i\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\sum\nolimits_{i=1}^{N} d_i\mathbf{z}_i\right)$$

The commands below will carry out the test. The chi squared reported by the last line has two degrees of freedom.

```
NAMELIST    ; x = one,... $
CREATE      ; y = the dependent variable $
PROBIT      ; Lhs = y ; Rhs = x  $
CREATE      ; ai = b'x ; fi = Phi(ai) ; dfi = N01(ai)
            ; di = (y-fi) * dfi  /(fi*(1-fi)) ; ci = dfi^2 /(fi*(1-fi))
            ; m3i = -1/2*(ai^2-1) ; m4i =  1/4*(ai*(ai^2+3)) $
NAMELIST    ; z  = x,m3i,m4i $
MATRIX      ; List ; LM = di’z * <z'[ci]z> * z'di $
```

We executed the routine for our probit model estimated earlier, with

```
NAMELIST    ; x = one,age,educ,married,hhninc,hhkids,self $
CREATE      ; y = doctor $
```

The result of 93.12115 would lead to rejection of the hypothesis of normality; the 5% critical value for the chi squared variable with two degrees of freedom is 5.99.

```
      LM|             1
--------+--------------
       1|       93.1211
```

E28.4 The WALD Command

The **WALD** command may be used for linear and nonlinear restrictions. The model commands produce a set of names that can be used in **WALD** commands after estimation. For the binary choice commands, these are *b_variable*. The **WALD** command can be used with these names in specified restrictions, with no other information needed. For example:

```
PROBIT      ; Lhs = doctor
            ; Rhs = one,age,educ,married,hhninc,hhkids  $
WALD        ; Fn1 = b_age + b_educ - 0
            ; Fn2 = b_married - 0
            ; Fn3 = b_hhninc + b_hhkids + .3 $
```

(The latter restriction doesn’t make much sense, but we can test it anyway.) The results of this pair of commands are shown below. (The **PROBIT** command was shown earlier.)

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic             =     24.95162
Prob. from Chi-squared[ 3] =       .00002
Functions are computed at means of variables
-----------------------------------------------------------------------------
```

```
--------+------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
WaldFcns|  Coefficient      Error          z    |z|>Z*         Interval
--------+------------------------------------------------------------
 Fncn(1)|    -.01528***     .00369     -4.14  .0000     -.02252   -.00805
 Fncn(2)|     .05226**      .02046      2.55  .0106      .01216    .09237
 Fncn(3)|     .04239        .05065       .84  .4027     -.05689    .14166
--------+------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

You may follow a model command with as many **WALD** commands as you wish.

You can use **WALD** to obtain standard errors for linear or nonlinear functions of parameters. Just ignore the test statistics. Also, **WALD** produces some useful output in addition to the displayed results. The new matrix *varwald* will contain the estimated asymptotic covariance matrix for the set of functions. The new vector *waldfns* will contain the values of the specified functions. A third matrix, *jacobian*, will equal the derivative matrix, $\partial \mathbf{c}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}'$. For the computations above, the three matrices are

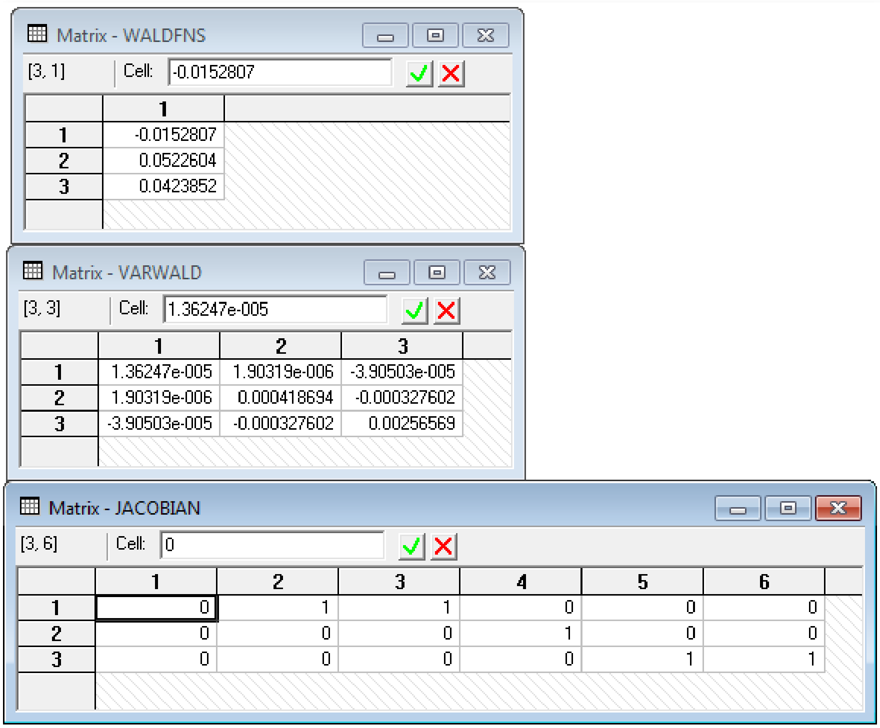


Figure E28.1 Matrix Results for the WALD Command

Thus, the command

MATRIX ; w = waldfns’ <varwald> waldfns $

would recompute the Wald statistic.

```
Matrix W         has  1 rows and  1 columns.
               1
        +--------------
       1|   24.95162
```

E28.5 Imposing Linear Restrictions

Fixed Value and Equality Restrictions

Fixed value and equality restrictions are imposed with

; Rst = the list of settings symbols for free parameters, values for specific values

For example,

```
NAMELIST    ; x = one,age,educ,married,hhninc,hhkids $
LOGIT       ; Lhs  =  doctor ; Rhs = x
            ; Rst  =  b0, b1, b1, 0, b2, b3 $
```

will force the second and third coefficients to be equal and the fourth to equal zero.

Linear Restrictions

These are imposed with

; CML: the set of linear restrictions

(See Section R13.6.2.) This is a bit more general than the Rst function, but similar. For example, to force the restriction that the coefficient on *age* plus that on *educ* equal twice that on *hhninc*, use

```
; CML:  age + educ - 2*hhninc = 0
```

E29: Extended Binary Choice Models

E29.1 Introduction

LIMDEP supports a large variety of models and extensions for the analysis of binary choice. This chapter documents sample selection models, models with endogenous right hand side variables and two step estimation of models that build on probit and logit models.

E29.2 Sample Selection in Probit and Logit Models

The model of sample selection can be extended to the probit and logit binary choice models. In both cases, we depart from

$$\text{Prob}[y_i = 1\,|\mathbf{x}_i] = \text{F}(\boldsymbol{\beta}'\mathbf{x}_i)$$

where $F(t) = \Phi(t)$ for the probit model and $\Lambda(t)$ for the logit model,

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\; u_i \sim N[0,1],\; z_i = 1(z_i^* > 0)$$

y_i, $\mathbf{x}_i$ observed only when $z_i = 1$.

In both cases, as stated, there is no obvious way that the selection mechanism impacts the binary choice model of interest. We modify the models as follows:

For the probit model,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\; \varepsilon_i \sim N[0,1],\; y_i = 1(y_i^* > 0)$$

which is the structure underlying the probit model in any event, and

$$u_i, \varepsilon_i \sim \text{BVN}[(0,0),(1,\rho,1)].$$

This is precisely the structure underlying the bivariate probit model. Thus, the probit model with selection is treated as a bivariate probit model. Some modification of the model is required to accommodate the selection mechanism. The full set of results is presented in Section E33.4. The command is simply

```
BIVARIATE   ; Lhs  = y,z
            ; Rh1 = variables in x
            ; Rh2 = variables in w
            ; Selection $
```

For the logit model, a similar approach does not produce a convenient bivariate model. The probability is changed to

$$\text{Prob}(y_i = 1 \mid \mathbf{x}_i,\varepsilon_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}.$$

With the selection model for z_i as stated above, the bivariate probability for y_i and z_i is a mixture of a logit and a probit model. The log likelihood can be obtained, but it is not in closed form, and must be computed by approximation. We do so with simulation. The model and the background results are presented in Section E54.5. The commands for the model are

PROBIT ; Lhs = z ; Rhs = variables in w ; Hold $
LOGIT ; Lhs = y ; Rhs = variables in x ; Selection $

The motivation for a probit selection mechanism into a logit model does seem ambiguous.

E29.3 Endogenous Variable in a Probit Model

This estimator is for what is essentially a simultaneous equations model. The model equations are

$$y_1{*} = \boldsymbol{\beta}'\mathbf{x} + \alpha y_2 + \varepsilon,\ y_1 = 1[y_1{*} > 0],$$
$$y_2 = \boldsymbol{\gamma}'\mathbf{z} + u,$$
$$(\varepsilon, u) \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\sigma\\ \rho\sigma & \sigma^2\end{pmatrix}\right].$$

Probit estimation based on y_1 and $(\mathbf{x}_1,y_2)$ will not consistently estimate $(\boldsymbol{\beta},\alpha)$ because of the correlation between y_2 and ε induced by the correlation between u and ε. Several methods have been proposed for estimation. One possibility is to use the partial reduced form obtained by inserting the second equation in the first. This will produce consistent estimates of $\boldsymbol{\beta}/(1+\alpha^2\sigma^2+2\alpha\sigma\rho)^{1/2}$ and $\alpha\boldsymbol{\gamma}/(1+\alpha^2\sigma^2+2\alpha\sigma\rho)^{1/2}$. Linear regression of y_2 on $\mathbf{z}$ produces estimates of γ and σ^2, but there is no method of moments estimator of ρ produced by this procedure, so this estimator is incomplete. Newey (1987) suggested a 'minimum chi squared' estimator that does estimate all parameters. A more direct, and actually simpler approach is full information maximum likelihood. Details on the estimation procedure appear below after the application.

To estimate this model, use the command

PROBIT ; Lhs = y1, y2
; Rh1 = independent variables in probit equation
; Rh2 = independent variables in regression equation $

(Note, the probit must be the first equation.) Other optional features relating to fitted values, marginal effects, etc. are the same as for the univariate probit command. We note, marginal effects are computed using the univariate probit probabilities,

$$\text{Prob}[y_1 = 1] \sim \Phi[\beta'x + \alpha y_2]$$

These will approximate the marginal effects obtained from the conditional model (which contain u). When averaged over the sample values, the effect of u will become asymptotically negligible. Predictions, etc. are kept with **; Keep = name**, and so on. Likewise, options for the optimization, such as maximum iterations, etc. are also the same as for the univariate probit model.

Retained Results

The results saved by this binary choice estimator are:

Matrices: *b* = estimate of ($\boldsymbol{\beta}$,α,$\boldsymbol{\gamma}$). Using **; Par** adds σ and ρ to *b*.
varb = asymptotic covariance matrix.

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variable* (includes α) and, *c_variables.*

Last Function: $\Phi(\mathbf{b}'\mathbf{x} + ay_2) = \text{Prob}(y_1 = 1 \mid \mathbf{x},y_2)$.

The *Last Model* names are used with **WALD** to simplify hypothesis tests. The last function is the conditional mean function. The extra complication of the estimator has been used to obtain a consistent estimator of β,α. With that in hand, the interesting function is $E[y_1| \mathbf{x},y_2]$.

```
NAMELIST    ; xdoctor = one,age,hsat,public,hhninc $
NAMELIST    ; xincome = one,age,age*age,educ,female,hhkids $
PROBIT      ; Lhs = doctor,hhninc
            ; Rh1 = xdoctor
            ; Rh2 = xincome $
```

```
-----------------------------------------------------------------------------
Probit   Regression Start Values for DOCTOR
Dependent variable               DOCTOR
Log likelihood function    -16634.88715
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =33279.774 AIC/N =    1.218
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.05627***      .05508    19.18  .0000      .94831   1.16423
     AGE|     .00895***      .00073    12.24  .0000      .00752    .01038
    HSAT|    -.17520***      .00395   -44.31  .0000     -.18295   -.16745
  PUBLIC|     .12985***      .02515     5.16  .0000      .08056    .17914
  HHNINC|    -.01332         .04581     -.29  .7712     -.10310    .07645
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=HHNINC  Mean                 =          .35208
            Standard deviation   =          .17691
            No. of observations  =           27326  Degrees of freedom
Regression  Sum of Squares       =         88.9621           5
Residual    Sum of Squares       =         766.216       27320
Total       Sum of Squares       =         855.178       27325
            Standard error of e  =          .16747
```

```
Fit          R-squared            =         .10403  R-bar squared =   .10386
Model test   F[  5, 27320]        =      634.40260  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =    10059.42844  Akaike I.C.   = -3.57369
             Restricted (b=0)     =     8558.60603  Bayes  I.C.   = -3.57189
             Chi squared [  5]    =     3001.64483  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.40365***      .01704   -23.68  .0000     -.43705   -.37024
     AGE|     .02555***      .00079    32.43  .0000      .02400    .02709
 AGE*AGE|    -.00029***   .9008D-05   -31.68  .0000     -.00030   -.00027
    EDUC|     .01989***      .00045    44.22  .0000      .01901    .02077
  FEMALE|     .00122         .00207      .59  .5538     -.00283    .00527
  HHKIDS|    -.01146***      .00231    -4.96  .0000     -.01599   -.00693
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Initial iterations cannot improve function.Status=3
  Error   805: Initial iterations cannot improve function.Status=3
Function=  .61428384629D+04, at entry,  .61358027527D+04 at exit

-----------------------------------------------------------------------------
Probit with Endogenous RHS Variable
Dependent variable                DOCTOR
Log likelihood function      -6135.80156
Restricted log likelihood   -16599.60800
Chi squared [  11 d.f.]      20927.61288
Significance level                .00000
McFadden Pseudo R-squared       .6303647
Estimation based on N =  27326, K =  13
Inf.Cr.AIC  =12297.603 AIC/N =      .450
--------+--------------------------------------------------------------------
  DOCTOR|                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Coefficients in Probit Equation for DOCTOR
Constant|    1.05627***      .07626    13.85  .0000      .90681   1.20574
     AGE|     .00895***      .00074    12.03  .0000      .00749    .01041
    HSAT|    -.17520***      .00392   -44.72  .0000     -.18288   -.16752
  PUBLIC|     .12985***      .02626     4.94  .0000      .07838    .18131
  HHNINC|    -.01332         .14728     -.09  .9279     -.30200    .27535
        |Coefficients in Linear Regression for HHNINC
Constant|    -.40301***      .01712   -23.55  .0000     -.43656   -.36946
     AGE|     .02551***      .00081    31.37  .0000      .02391    .02710
 AGE*AGE|    -.00028***   .9377D-05   -30.39  .0000     -.00030   -.00027
    EDUC|     .01986***      .00040    50.26  .0000      .01908    .02063
  FEMALE|     .00122         .00207      .59  .5552     -.00284    .00528
  HHKIDS|    -.01144***      .00226    -5.06  .0000     -.01587   -.00701
        |Standard Deviation of Regression Disturbances
Sigma(w)|     .16720***      .00026   639.64  .0000      .16669    .16772
        |Correlation Between Probit and Regression Disturbances
Rho(e,w)|     .02412         .02550      .95  .3442     -.02586    .07409
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Technical Details

The log likelihood is built up from the joint density of y_1 and y_2, which we write as the product of the conditional and the marginal densities,

$$f(y_1,y_2) \;=\; f(y_1|y_2)\, f(y_2).$$

To derive the conditional distribution, we use results for the bivariate normal, and write

$$\varepsilon|u \;=\; [(\rho\sigma)/\sigma^2]\, u \;+\; w,$$

where w is normally distributed with $\text{Var}[w] = (1 - \rho^2)$. Inserting this in the first equation, we have

$$y_1^*|y_2 = \boldsymbol{\beta}'\mathbf{x} + \alpha y_2 \;+\; (\rho/\sigma)u \;+\; w.$$

Therefore,

$$\text{Prob}[y_1 = 1|y_2] \;=\; \Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x} + \alpha y_2 + (\rho/\sigma)u}{\sqrt{1-\rho^2}}\right].$$

Inserting the expression for u, and using the normal density for the marginal distribution of y_2, we obtain the log likelihood function for the sample,

$$\log L = \sum\nolimits_{i=1}^{N} \;\log\Phi\left[(2y_{i1}-1)\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i + \alpha y_{i2} + (\rho/\sigma)(y_{i2}-\boldsymbol{\gamma}'\mathbf{z}_i)}{\sqrt{1-\rho^2}}\right)\right] + \log\left[\frac{1}{\sigma}\phi\left(\frac{y_{i2}-\boldsymbol{\gamma}'\mathbf{z}_i}{\sigma}\right)\right]$$

We use several devices to make estimation easier. First, we use the Olsen transformation to reparameterize the log likelihood in

$$\theta \;=\; 1/\sigma$$

$$\boldsymbol{\delta} \;=\; (1/\sigma)\boldsymbol{\gamma}.$$

We ensure that θ is positive during estimation by estimating

$$\eta \;=\; \log\theta, \text{ so } \theta \;=\; \exp(\eta).$$

To force the correlation to remain in the (-1,+1) interval, we maximize the log likelihood with respect to

$$\tau \;=\; \log\left(\frac{1+\rho}{1-\rho}\right), \text{ so } \;\rho = \frac{\exp(\tau)-1}{\exp(\tau)+1}.$$

The log likelihood is, then,

$$\log L = \sum_{i=1}^{N} \log\Phi\left[(2y_{i1}-1)\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i + \alpha y_{i2} + \rho(\theta y_{i2} - \boldsymbol{\delta}'\mathbf{z}_i)}{\sqrt{1-\rho^2}}\right)\right] + \log\left[\theta\phi\left(\theta y_{i2} - \boldsymbol{\delta}'\mathbf{z}_i\right)\right].$$

(The likelihood can actually be further simplified. Since $\boldsymbol{\beta}$ is a free parameter vector, we can let $\boldsymbol{\beta}_r$ equal $\boldsymbol{\beta}/(1 - \rho^2)^{1/2}$ and α_r equal $\alpha/(1 - \rho^2)^{1/2}$. Then, define $\omega = \rho/(1 - \rho^2)^{1/2}$. The resulting log likelihood is

$$\log L = \sum_{i=1}^{N} \log\Phi\left[(2y_{i1}-1)\left((\boldsymbol{\beta}_r'\mathbf{x}_i + \alpha_r y_{i2}) + \omega(\theta y_{i2} - \boldsymbol{\delta}'\mathbf{z}_i)\right)\right] + \log\left[\theta\phi\left(\theta y_{i2} - \boldsymbol{\delta}'\mathbf{z}_i\right)\right].$$

This simplifies the programming a bit, but does not actually improve the process of optimization.) The delta method is used after estimation to recover the estimates of the original parameters and estimators of their asymptotic variances.

E29.4 Using MAXIMIZE to Estimate Other Parametric Models

The general formulation used earlier suggests a means of extending the binary choice model to distributions other than the ones listed in Section E27.2. In particular, if the model is formulated as a single index regression:

y_i = a binary outcome taking values 0 or 1 with

Prob[$y_i = 1$] = $F(\boldsymbol{\beta}'\mathbf{x}_i)$, such that $F'(\boldsymbol{\beta}'\mathbf{x}_i) \geq 0$ and $0 < F(\boldsymbol{\beta}'\mathbf{x}_i) < 1$,

then any proper probability distribution function will suffice. This is simply a model, with no more or less justification than the logistic or normal distributions.

There are many possibilities that one might consider. The binary probability model is a particularly simple one to formulate, and *LIMDEP*'s **MAXIMIZE** routine is well suited to this type of problem. A template that one might use for this approach would be as follows:

```
NAMELIST    ; x  = the set of Rhs variables $
CREATE      ; y  = ... define the dependent variable $
CALC        ; k  = Col(x) $
MATRIX      ; b0 = Init(k,1,0.0) $
MAXIMIZE    ; Start = b0
            ; Labels = k_beta
            ; Fcn = bx = beta1'x |
                    p = ... the definition of F(bx) |
                    y * Log(p) + (1-y)*Log(1-p) $
```

E29.5 Two Step Estimation Using Binary Choice Models

The essential parts of a two procedure are

Step 1. A model is estimated by least squares or maximum likelihood. Denote the parameters estimated at this step as $\boldsymbol{\theta}_1$. For present purposes, this is the probit or logit model.

Step 2. A second model is estimated in which a predicted value from the model in Step 1 appears on the right hand side of the equation. Denote the full set of parameters estimated at this step as $\boldsymbol{\theta}_2$.

We assume that estimation at both steps is consistent – the modeler will have to verify this on a case by case basis. The remaining computation is the correction of the estimated asymptotic covariance matrix for the estimator at Step 2 for the randomness of the estimator carried forward from Step 1. We base our results for this computation on the Murphy and Topel (1985) paper which presents a general method of doing the calculations. (See Greene (2011) for additional discussion.) There are like results for GMM estimation – see Newey (1984) – however, we restrict our attention to maximum likelihood estimation in *LIMDEP*.

The underlying result is as follows (again, from Greene (2011)): Let $\mathbf{V}_2$ be the uncorrected covariance matrix computed at Step 2, using the parameter estimates obtained at Step 1 as if they were known, and $\mathbf{V}_1$ be the estimator of the asymptotic covariance matrix for the parameter estimates obtained at Step 1. Both of these estimators are based on the respective log likelihood functions. In addition, define

$$\mathbf{C} = \sum_{i=1}^{n}\left(\frac{\partial \log f_{i2}}{\partial \boldsymbol{\theta}_2}\right)\left(\frac{\partial \log f_{i2}}{\partial \boldsymbol{\theta}_1'}\right)$$

and

$$\mathbf{R} = \sum_{i=1}^{n}\left(\frac{\partial \log f_{i2}}{\partial \boldsymbol{\theta}_2}\right)\left(\frac{\partial \log f_{i1}}{\partial \boldsymbol{\theta}_1'}\right)$$

(Note the derivatives shown are the derivatives of individual terms in the two log likelihoods. These appear at various points above for the probit and logit models.) With these in hand, the corrected covariance matrix for the second step estimator is

$$\mathbf{V}_2^* = \mathbf{V}_2 + \mathbf{V}_2[\mathbf{C}\mathbf{V}_1\mathbf{C}' - \mathbf{R}\mathbf{V}_1\mathbf{C}' - \mathbf{C}\mathbf{V}_1\mathbf{R}']\mathbf{V}_2$$

A general case that has been automated in *LIMDEP* is a model of the form:

y_1 = a binary variable specified by a probit or logit formulation,

y_2 = a dependent variable whose conditional mean function is a function of $E[y_1]$.

Models of this sort could in principle be estimated by full information maximum likelihood. We consider two step estimation instead, which is often simpler. Models for which the second step shown above is automated are the following:

- Probit and probit with heteroscedasticity,
- Truncated regression,
- Tobit and tobit with heteroscedasticity,
- Poisson and negative binomial regression,
- Linear regression.

For these models, the estimation procedure is the following two steps:

```
PROBIT        ; Lhs = y1 ; Rhs = as usual
or LOGIT      ; Prob = py
              ; Hold $
Model name    ; Lhs = y2
              ; Rhs = as usual,py   Note, py, not y1.
              ; 2Step = py $
```

In the example shown below, a probit model is estimated and the results are held for the second step. At the second step, a Poisson regression model is estimated. (Results for the probit model are omitted.) The second set of estimates shown omit the Murphy and Topel correction.

```
NAMELIST   ; x = one,age,educ,married,hhninc,hhkids $
PROBIT     ; Lhs = doctor  ; Rhs = x ; Prob = pdoc ; Hold $
POISSON    ; Lhs = hospvis ; Rhs = one,age,educ,married,pdoc
           ; 2step = pdoc $
POISSON    ; Lhs = hospvis ; Rhs = one,age,educ,married,pdoc  $
```

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable               HOSPVIS
Log likelihood function     -13352.51694
Restricted log likelihood   -13433.21441
Chi squared [    4 d.f.]       161.39493
Significance level                .00000
McFadden Pseudo R-squared       .0060073
Estimation based on N =   27326, K =   5
Inf.Cr.AIC  =26715.034 AIC/N =      .978
Murphy/Topel 2Step VC matrix:P=     PDOC
Model estimated: Jun 13, 2011, 21:22:28
Chi- squared =148819.56673  RsqP= .0372
G  - squared = 21457.91309  RsqD= .0075
Overdispersion tests: g=mu(i)  :   4.164
Overdispersion tests: g=mu(i)^2:   4.268
-----------------------------------------------------------------------------
```

HOSPVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	-1.29249***	.41061	-3.15	.0016	-2.09728	-.48770
AGE	.01115**	.00464	2.40	.0163	.00205	.02025
EDUC	-.08171***	.01211	-6.74	.0000	-.10546	-.05797
MARRIED	-.05946	.04328	-1.37	.1696	-.14429	.02538
PDOC	-.35623	.76816	-.46	.6428	-1.86179	1.14933

(Uncorrected)

HOSPVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	-1.29249***	.38799	-3.33	.0009	-2.05293	-.53205
AGE	.01115**	.00438	2.55	.0109	.00257	.01973
EDUC	-.08171***	.01162	-7.03	.0000	-.10449	-.05893
MARRIED	-.05946	.03874	-1.53	.1249	-.13539	.01648
PDOC	-.35623	.72080	-.49	.6212	-1.76897	1.05651

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

The preceding includes a fairly large number of possible specifications, given all the different combinations. (Any of the first step models may be used with any of the second step models.) But, an essentially infinite number of possible different specifications remain. If you wish to use this procedure, you may have to program the second step correction yourself to do so. *LIMDEP*'s various programming features should make this fairly easy. To illustrate, we will present a moderately complicated case in detail. For the example, we consider a multinomial logit model for a y_2 which has three outcomes and a y_1 determined by a probit model. The model is

$$y_1^* = \boldsymbol{\theta}'\mathbf{z} + \varepsilon_1,$$

$$y_1 = \mathbf{1}(y_1^* > 0),$$

$$E[y_1] = \Phi(\boldsymbol{\theta}'\mathbf{z}),\ \varepsilon \sim N[0,\sigma^2],\ \Phi(.) = \text{standard normal CDF},$$

$$\text{Prob}[y_2 = j] = e_j / (e_0 + e_1 + e_2),\ j = 0, 1, 2,$$

$$e_0 = 1$$

$$e_1 = \exp[\ \boldsymbol{\beta}_1'\mathbf{x} + \gamma_1\Phi(\boldsymbol{\theta}'\mathbf{z})\ ]$$

$$e_2 = \exp[\ \boldsymbol{\beta}_2'\mathbf{x} + \gamma_2\Phi(\boldsymbol{\theta}'\mathbf{z})\ ]$$

At Step 1, $\boldsymbol{\theta}$ is estimated by maximizing the log likelihood

$$\log L_1 = \sum\nolimits_{i=1}^{n} \log f_{1i}(y_{1i}, \mathbf{z}_i \mid \boldsymbol{\theta}) = \sum\nolimits_{i=1}^{n} \log \Phi\left(q_i \boldsymbol{\theta}' \mathbf{z}_i\right), \text{ where } q_i = 2y_{1i} - 1.$$

After the first step is complete, the predictions, $\Phi(\boldsymbol{\theta}'\mathbf{z})$, are computed using the maximum likelihood estimates, then the log likelihood for the second model is maximized with respect to $\boldsymbol{\beta}_1,\gamma_1,\boldsymbol{\beta}_2,\gamma_2$ while treating the predictions as if they were observed data. The second step log likelihood function is

$$\log L_2 = \sum_{i=1}^{n} \log f_{2i}(y_{2i},\mathbf{x}_i,\Phi(\boldsymbol{\theta}'\mathbf{z}_i)\,|\,\boldsymbol{\beta}_1,\gamma_1,\boldsymbol{\beta}_2,\gamma_2)$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{2} d_{ij} \log \text{Prob}[y_{2i} = j]-\text{, where } d_{ij} = 1 \text{ if } y_{2i} = j,\ j = 0,1,2$$

Each step produces its own estimated parameter vector and asymptotic covariance matrix. The matrices needed for the correction are:

$$\mathbf{C} = \sum_{i=1}^{n} \begin{bmatrix} (d_{i1}-P_{i1})\begin{pmatrix}\mathbf{x}_i \\ \Phi(\boldsymbol{\theta}'\mathbf{z}_i)\end{pmatrix} \\ (d_{i2}-P_{i2})\begin{pmatrix}\mathbf{x}_i \\ \Phi(\boldsymbol{\theta}'\mathbf{z}_i)\end{pmatrix}\end{bmatrix} \times \left[(d_{i1}-P_{i1})\gamma_1 + (d_{i2}-P_{i2})\gamma_2\right]\phi(\boldsymbol{\theta}'\mathbf{z}_i)\mathbf{z}_i'$$

$$\mathbf{R} = \sum_{i=1}^{n} \begin{bmatrix} (d_{i1}-P_{i1})\begin{pmatrix}\mathbf{x}_i \\ \Phi(\boldsymbol{\theta}'\mathbf{z}_i)\end{pmatrix} \\ (d_{i2}-P_{i2})\begin{pmatrix}\mathbf{x}_i \\ \Phi(\boldsymbol{\theta}'\mathbf{z}_i)\end{pmatrix}\end{bmatrix} \times \left[\frac{q_i\phi(\boldsymbol{\theta}'\mathbf{z}_i)}{\Phi(q_i\boldsymbol{\theta}'\mathbf{z}_i)}\mathbf{z}_i'\right]$$

(Derivatives for the multinomial logit log likelihood above appear later in this manual.)

This part of the routine is set up for the particular application. The remainder is general, and need not be changed.

```
NAMELIST    ; x   = … define the Rhs for the multinomial logit model
            ; z   = … define the Rhs for the probit model
CREATE      ; y1 = … dependent variable in the probit model
            ; y2 = … dependent variable in logit model
```

Estimate the probit model. The **IMR = lambda** is just for convenience. It computes the q*N01/Phi in the first log likelihood. Pick up other terms now.

```
PROBIT      ; Lhs = y1 ; Rhs = z ; Prob = prob ; Hold(IMR = lambda) $
CREATE      ; den1 = N01(b'z) $
MATRIX      ; v1 = varb $
```

Augment the Rhs of the logit model with the fitted probability from the probit model, then fit the logit model.

```
NAMELIST    ; xp = x,prob ; xbrep = xp,xp$
LOGIT       ; Lhs = y2 ; Rhs = xp $
```

Get the subvectors of the logit parameter vector and the coefficients on the fitted probability.

```
CALC        ; k = Col(xp) ; j21 = k+1 ; j22 = 2*k
            ; gamma1 = b(k) ; gamma2 = b(j22) $
MATRIX      ; b1 = b(1:k) ; b2 = b(j21:j22) $
```

Compute the scalars that appear in the summations in the construction of the **C** and **R** matrices.

```
CREATE      ; d0  = (y2 = 0) ; d1 = (y2 = 1) ; d2 = (y2 = 2)
            ; e0  = 1 ; e1 = Exp(b1'xp) ; e2 = Exp(b2'xp)
            ; p0  = e0 / (e0 + e1 + e2 ) ; p1 = e1 * p0 ; p2 = e2 * p0
            ; u1  = (d1 – p1 ) ; u2 = (d2 – p2 )
            ; dc1= u1 * (u1*gamma1 + u2*gamma2)*den1
            ; dc2= u2 * (u1*gamma1 + u2*gamma2)*den1
            ; dr1= u1 * lambda ; dr2 = u2 * lambda $
```

Note the matrix constructions. The namelist[variable]namelist format is specifically for computing matrices of the form of **C** and **R** in the expressions above. We compute both matrices in two parts, then stack the parts.

```
MATRIX      ; cm1 = xp' [dc1] z ; cm2 = xp' [dc2 ] z
            ; rm1 = xp' [dr1] z ; rm2 = xp' [dr2 ] z
            ; c = [cm1 / cm2] ; r = [rm1 / rm2] $
```

The last computation computes the corrected covariance matrix, then shows the results.

```
MATRIX      ; t = c * v1 * c' – c * v1 * r' – r * v1 * c'
            ; v2 = varb + varb * t * varb
            ; Stat(b,v2,xbrep) $
```

To illustrate, we used the procedure with the following variable definitions: The data set contains *newhsat*, which is a self assessment of health satisfaction ranked from zero to 10. (The raw data contain the variable *hsat*. *Newhsat* corrects some obvious coding errors.) We created the variable *lhsat* which is zero for *newhsat* less than six, one for *newhsat* from six to eight and two otherwise.

```
CREATE      ; lhsat = 0 + ((newhsat=6)+(newhsat=7)+(newhsat=8))+2*(newhsat > 8) - 1 $
NAMELIST    ; x = one,age,educ,married,working,bluec,whitec,self
            ; z = one,age,educ,married,hhninc,hhkids $
CREATE      ; y1 = doctor ; y2 = lhsat $
```

Estimates produced by the procedure are shown below. The intermediate output is omitted. The first set of results is the uncorrected results of the **LOGIT** command. The second use the Murphy and Topel correction to the asymptotic covariance matrix.

```
-----------------------------------------------------------------------------
Multinomial Logit Model
Dependent variable                Y2
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      Y2|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    2.12189***      .34637     6.13  .0000     1.44302   2.80076
     AGE|    -.00755*        .00394    -1.91  .0555     -.01528    .00018
    EDUC|     .03653***      .01038     3.52  .0004      .01619    .05688
 MARRIED|     .13272***      .03473     3.82  .0001      .06466    .20078
 WORKING|     .34639***      .05318     6.51  .0000      .24216    .45062
   BLUEC|    -.12961**       .05675    -2.28  .0224     -.24084   -.01838
  WHITEC|     .06000         .05407     1.11  .2672     -.04598    .16598
    SELF|    -.04926         .07452     -.66  .5086     -.19532    .09681
    PROB|   -3.17753***      .64402    -4.93  .0000    -4.43978  -1.91527
        |Characteristics in numerator of Prob[Y = 2]
Constant|    2.59418***      .40842     6.35  .0000     1.79369   3.39467
     AGE|    -.03235***      .00461    -7.02  .0000     -.04138   -.02332
    EDUC|     .05585***      .01199     4.66  .0000      .03235    .07936
 MARRIED|     .12444***      .04153     3.00  .0027      .04305    .20583
 WORKING|     .38404***      .06193     6.20  .0000      .26266    .50542
   BLUEC|    -.27785***      .06647    -4.18  .0000     -.40813   -.14757
  WHITEC|    -.07213         .06209    -1.16  .2453     -.19383    .04956
    SELF|     .07701         .08528      .90  .3665     -.09014    .24417
    PROB|   -3.73917***      .75481    -4.95  .0000    -5.21857  -2.25977
--------+--------------------------------------------------------------------
(Corrected)
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  Matrix|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2.12189***      .38264     5.55  .0000     1.37193   2.87185
     AGE|    -.00755*        .00429    -1.76  .0785     -.01596    .00086
    EDUC|     .03653***      .01143     3.20  .0014      .01413    .05894
 MARRIED|     .13272***      .03815     3.48  .0005      .05795    .20749
 WORKING|     .34639***      .05329     6.50  .0000      .24195    .45084
   BLUEC|    -.12961**       .05678    -2.28  .0224     -.24089   -.01833
  WHITEC|     .06000         .05408     1.11  .2672     -.04600    .16599
    SELF|    -.04926         .07453     -.66  .5087     -.19533    .09682
    PROB|   -3.17753***      .70554    -4.50  .0000    -4.56035  -1.79470
Constant|    2.59418***      .43099     6.02  .0000     1.74947   3.43890
     AGE|    -.03235***      .00480    -6.73  .0000     -.04177   -.02293
    EDUC|     .05585***      .01266     4.41  .0000      .03104    .08067
 MARRIED|     .12444***      .04308     2.89  .0039      .04000    .20888
 WORKING|     .38404***      .06198     6.20  .0000      .26256    .50551
   BLUEC|    -.27785***      .06649    -4.18  .0000     -.40816   -.14754
  WHITEC|    -.07213         .06208    -1.16  .2453     -.19382    .04955
    SELF|     .07701         .08529      .90  .3666     -.09016    .24419
    PROB|   -3.73917***      .79080    -4.73  .0000    -5.28911  -2.18923
--------+--------------------------------------------------------------------
```

E29.6 Other Models that Build on the Binary Choice Models

A variety of the other models that are estimated with *LIMDEP* are built upon the binary choice framework, particularly the probit model. Some of the model extensions that may be of interest are:

- Bivariate and multivariate extensions of the probit model. These extend the latent regression model to a multivariate regression framework:

$$y_1^* = \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1,\ y_1 = 1 \text{ if } y_1^* > 0, 0 \text{ otherwise,}$$

$$y_2^* = \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2,\ y_2 = 1 \text{ if } y_2^* > 0, 0 \text{ otherwise,}$$

$$\ldots$$

$$y_M^* = \boldsymbol{\beta}_M'\mathbf{x}_M + \varepsilon_M,\ y_M = 1 \text{ if } y_M^* > 0, 0 \text{ otherwise,}$$

  The disturbances in the equations are allowed to be freely correlated. The bivariate probit model restricts this to two equations, but includes some useful extensions of the model.

  - Bivariate probit with exponential heteroscedasticity
  - Bivariate probit with sample selection: (y_1 ,$\mathbf{x}_1$) only observed when $y_2 = 1$.
  - Partial observability models: Only y_1y_2 is observed. (There are three variants.)

  The unrestricted multivariate probit (MVP) model stated above allows up to 20 equations.

- Multinomial probit (MNP). The MNP model is part of *NLOGIT* Version 5 and is not available in *LIMDEP* Version 10. The MNP model modifies the MVP model above by changing the observation mechanism. The observed outcome is an indicator, j, which denotes which of the M y_j*s is the maximum. The interpretation is that the right hand sides of the regressions are the random utilities associated with M choices, and the individual chooses the one which gives greatest utility.

- Sample selection models. This is a group of models that build on a regression type of model,

$$f(Y^*) = g(\boldsymbol{\gamma}'\mathbf{z} + u) \text{ where } f(.) \text{ is the probability distribution of some observed variable, which depends on a latent or observed regression model}$$

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ y = 1 \text{ if } y^* > 0, 0 \text{ otherwise,}$$

  Y^* and $\mathbf{z}$ are observed only when $y = 1$, ε and u are correlated.

  Examples include the bivariate probit model mentioned earlier, a regression model (Heckman’s model of sample selection), an ordered probability model, and a Poisson regression model for counts. The models are presented in Chapter E54.

- Ordered probability models are forms of the probit and logit models in which there are more than two outcomes, coded 0,1,2,... The observed outcome occurs with probability drawn from the normal or logistic distribution, with the range of the latent variable divided into more than two parts. The base form of the ordered probability model is

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \ \ \varepsilon_i \sim \text{standard normal or standard logistic,}$$

$$y_i = 0 \text{ if } y_i^* \le \mu_0,$$

$$1 \text{ if } \mu_0 < y_i^* \le \mu_1,$$

$$2 \text{ if } \mu_1 < y_i^* \le \mu_2,$$

$$\ldots$$

$$J \text{ if } y_i^* > \mu_{J-1}.$$

The observed counterpart to y_i^* is y_i. Note that the probit model that has been discussed in this chapter is the special case when $J = 1$.

- Zero inflation models for count data. The models for count data specify that

$$\text{Prob}[y_i = j] = g(\boldsymbol{\gamma}'\mathbf{z}_i), j = 0,1,\ldots$$

The zero inflation models extend this framework to include the possibility that observations of zero may arise from two regimes. In regime 1, y always equals zero. In regime 2, y follows the distribution above. The determination of which regime applies is modeled as the outcome of a binary choice,

$$\text{Prob[regime 1]} = F(\boldsymbol{\beta}'\mathbf{x}), \ \ \text{Prob[regime 2]} = 1 - F(\boldsymbol{\beta}'\mathbf{x}).$$

The zero inflation models for count data are discussed in Section E43.6.

- Split population models. Parametric models of duration are based on data which are 'complete,' – a transition takes place and 'censored' – the transition has not taken place at the time of observation, and it is assumed that it will take place eventually. The split population models relax this assumption by extending the duration model with a binary choice equation that models the censoring process. The implication is that some observations which are observed as censored might be reasonably treated as if they would never actually experience the transition. See Section E61.5 for this development.

E30: Fixed and Random Effects Models for Binary Choice

E30.1 Introduction

The parametric models discussed in Chapters E27-E29 are extended to panel data formats. Four specific parametric model formulations are provided as internal procedures in *LIMDEP* for these binary choice models. These are the same ones described earlier, less the Burr distribution which is not included in this set. Four classes of models are supported:

- **Fixed effects:** $\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$,
  α_i may be correlated with $\mathbf{x}_{it}$,

- **Random effects:** $\text{Prob}[y_{it} = 1] = \text{Prob}[\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i > 0]$,
  u_i is uncorrelated with $\mathbf{x}_{it}$,

- **Random parameters:** $\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}_i'\mathbf{x}_{it})$,
  $\boldsymbol{\beta}_i \mid i \sim h(\boldsymbol{\beta}|i)$ with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$

- **Latent class:** $\text{Prob}[y_{it} = 1|\text{class j}] = F(\boldsymbol{\beta}_j'\mathbf{x}_{it})$,
  $\text{Prob}[\text{class = j}] = F_j(\boldsymbol{\theta})$

The last two models provide various extensions of the basic form shown above.

NOTE: None of these panel data models require balanced panels. The group sizes may always vary.

NOTE: None of these panel data models are provided for the Burr (scobit) model.

All formulations are treated the same for the five models, probit, logit, extreme value, Gompertz and arctangent.

NOTE: The random effects estimator requires individual data. The fixed effects estimator allows grouped data.

The third and fourth arise naturally in a panel data setting, but in fact, can be used in cross section frameworks as well. The fixed and random effects estimators require panel data. The fixed and random effects models are described in this chapter. Random parameters and latent class models are documented in Chapter E31.

The probabilities and density functions supported here are as follows:

Probit

$$F = \int_{-\infty}^{\beta'\mathbf{x}_i} \frac{\exp(-t^2/2)}{\sqrt{2\pi}} dt = \Phi(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \phi(\boldsymbol{\beta}'\mathbf{x}_i)$$

Logit

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

Complementary log log

$$F = 1 - \exp(-\exp(\boldsymbol{\beta}'\mathbf{x}_i)) = C(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \exp(\boldsymbol{\beta}'\mathbf{x}_i)[1 - C(\boldsymbol{\beta}'\mathbf{x}_i)]$$

Gompertz, or type 1 extreme value

$$F = \exp(-\exp(-\boldsymbol{\beta}'\mathbf{x}_i)) = G(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \exp(-\boldsymbol{\beta}'\mathbf{x}_i)G(\boldsymbol{\beta}'\mathbf{x}_i)$$

Arctangent

$$F = 2/\pi \arctan(\exp(\boldsymbol{\beta}'\mathbf{x}_i)), \qquad f = 2/\pi\, [1/(1 + \exp^2(\boldsymbol{\beta}'\mathbf{x}_i))\,]$$

The applications in this chapter are based on the German health care data used throughout the documentation (See Section E2.4). The data are an unbalanced panel of observations on health care utilization by 7,293 individuals. The group sizes in the panel number as follows: T_i: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987. There are altogether 27,326 observations. The variables in the file that are used here are

doctor = 1 if number of doctor visits > 0, 0 otherwise,
hhninc = household nominal monthly net income in German marks / 10000,
hhkids = 1 if children under age 16 in the household, 0 otherwise,
educ = years of schooling,
married = marital status,
female = 1 for female, 0 for male,
docvis = number of visits to the doctor,
hospvis = number of visits to the hospital,
newhsat = self assessed health satisfaction, coded 0,1,...,10.

(The data on health satisfaction in the raw data file, in variable *hsat*, contained some obvious coding errors. Our corrected data are in *newhsat*.)

E30.2 Commands

The essential model command for the models described in this chapter are

PROBIT **LOGIT** **COMPLOG** **GOMPERTZ** **ARCTANGENT**	**; Lhs = dependent variable** **; Rhs = independent variables - not including one** **; Panel** **; ... specification of the panel data model $**

As always, panels may be balanced or unbalanced. The panel is indicated with

SETPANEL ; Group = group identifier
; Pds = count variable to be created $

Thereafter,

; Panel

in the model command is sufficient to specify the panel setting. In circumstances where you have set up the count variable yourself, you may also use the explicit declaration in the command:

; Pds = the fixed number of periods if the panel is balanced
; Pds = a variable which, within a group, repeats the number of observations in the group

One or the other of these two specifications is required for the fixed and random effects estimators.

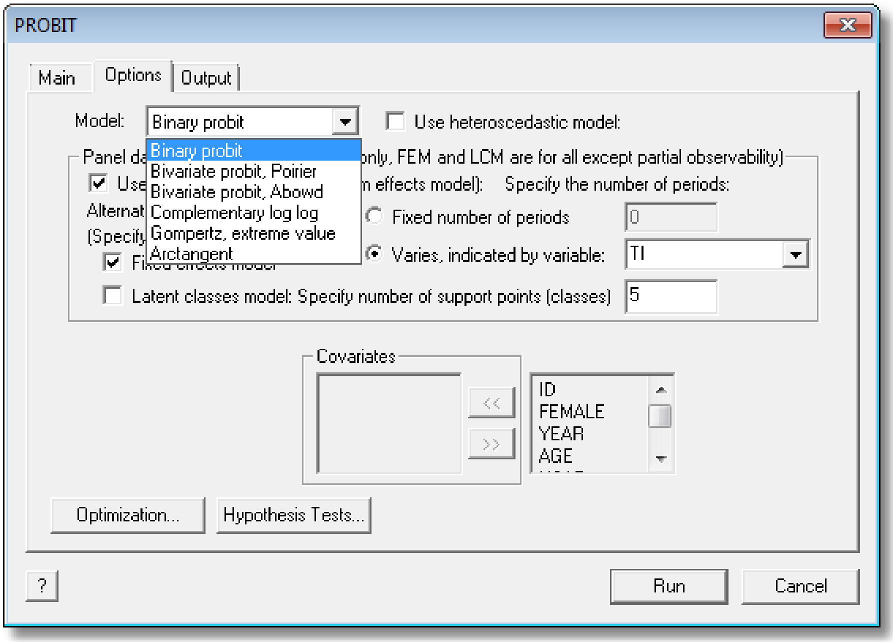
NOTE: For these estimators, you should not attempt to manage missing data. Just leave observations with missing values in the sample. *LIMDEP* will automatically bypass the missing values. Do not use **SKIP**, as it will undermine the setting of **; Pds = specification**.

The estimator produces and saves the coefficient estimator, *b* and covariance matrix, *varb*, as usual. Unless requested, the estimated fixed effects coefficients are not retained. (They are not reported regardless.) To save the vector of fixed effects estimates, α in a matrix named *alphafe*, add

; Parameters

to the command. The fixed effects estimators allow up to 100,000 groups. However, only up to 50,000 estimated constant terms may be saved in *alphafe*.

The Options pages of the Model:Binary Choice/Probit and Model:Binary Choice/ Logit provide command builders for the panel data models. The probit, complementary log log, arctangent and extreme value models are all in the **PROBIT** command builder.



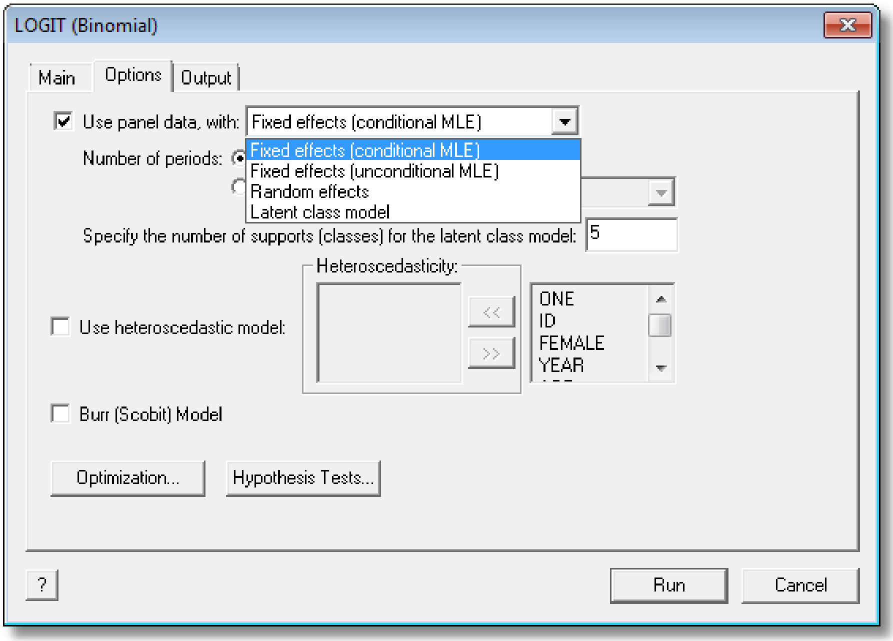


Figure E30.1 Command Builder for Panel Data Binary Logit Models

E30.3 Clustering, Stratification and Robust Covariance Matrices

The robust estimator based on sample clustering and stratification is available for the parametric binary choice models. Full details appear in Chapter R10 for the general case and Section E27.5.2 for the parametric binary choice models of interest here. The option for clustering is offered in the command builders for most of the nonlinear model and binary choice routines in the Model Estimates submenu. This will differ a bit from model to model. The one for the probit model is shown below in Figure E30.2. The Model Estimates dialog box is selected at the bottom of the Output page, then the clustering is specified in the next dialog box.

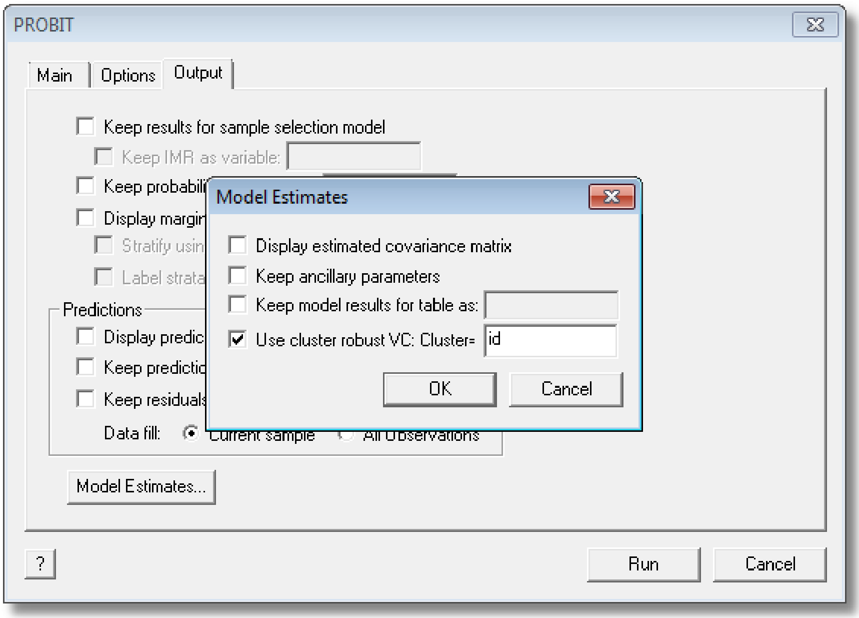


Figure E30.2 Command Builder for a Probit Model

This sampling setup may be used with any of the binary choice estimators. Do note, however, you should not use it with panel data models. The so called ‘clustering’ corrections are already built into the panel data estimators. (This is unlike the linear regression case, in which some authors argue that the correction should be used even when fixed or random effects models are estimated.)

To illustrate, the following shows the setup for the panel data set described in the preceding section. We have also artificially reduced the sample to 1,015 observations, 29 groups of 35 individuals, all of whom were observed seven times. The information below would appear with a model command that used this configuration of the data to construct a robust covariance matrix.

```
SAMPLE      ; 1-5000 $
REJECT      ; _groupti < 7 $
NAMELIST    ; x = age,educ,hhninc,hhkids,married $
PROBIT      ; Lhs = doctor ; Rhs = one,x
            ; Cluster = 7
            ; Stratum = 35
            ; Describe $
```

These results appear before any results of the probit command. They are produced by the **; Describe** specification in the command.

```
========================================================================
 Summary of Sample Configuration for Two Level Stratified Data
========================================================================
 Stratum #   Stratum     Number Groups          Group Sizes
            Size (obs)  Sample   FPC.       1       2       3 ...    Mean
==========  ==========  =============  =================================
         1          35        5 1.0000       7       7       7 ...     7.0
         2          35        5 1.0000       7       7       7 ...     7.0
        (Rows 3 – 28 omitted)
        29          35        5 1.0000       7       7       7 ...     7.0

+-----------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.      |
| Sample of   1015 observations contained    145 clusters defined by    |
|      7 observations (fixed number) in each cluster.                   |
| Sample of   1015 observations contained     29 strata defined by      |
|     35 observations (fixed number) in each stratum.                   |
+-----------------------------------------------------------------------+
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                DOCTOR
Log likelihood function      -621.15030
Restricted log likelihood    -634.14416
Chi squared [   5 d.f.]        25.98772
Significance level               .00009
McFadden Pseudo R-squared      .0204904
Estimation based on N =   1015, K =   6
Inf.Cr.AIC  = 1254.301 AIC/N =    1.236
Hosmer-Lemeshow chi-squared =  18.58245
P-value=  .01726 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|      .71039       2.41718      .29  .7688    -4.02720   5.44797
     AGE|      .00659        .03221      .20  .8378     -.05655    .06973
    EDUC|     -.05898        .14043     -.42  .6745     -.33421    .21625
  HHNINC|     -.13753       1.25599     -.11  .9128    -2.59921   2.32416
  HHKIDS|     -.11452        .56015     -.20  .8380    -1.21240    .98336
 MARRIED|      .29025        .82535      .35  .7251    -1.32741   1.90791
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E30.4 One and Two Way Fixed Effects Models

The fixed effects models are estimated by unconditional maximum likelihood. The command for requesting the model is

PROBIT
LOGIT
COMPLOG
GOMPERTZ
ARCTANGENT

; Lhs = dependent variable
; Rhs = independent variables - not including one
; Panel
; Fixed Effects or ; FEM $

NOTE: Your Rhs list should not include a constant term, as the fixed effects model fits a complete set of constants for the set of groups. If you do include one in your Rhs list, it is automatically removed prior to beginning estimation.

The fixed effects model assumes a group specific effect:

$$\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$$

where α_i is the parameter to be estimated. You may also fit a two way fixed effects model

$$\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \gamma_t)$$

where γ_t is an additional, time (period) specific effect. The time specific effect is requested by adding

; Time

to the command if the panel is balanced, and

; Time = variable name

if the panel is unbalanced. For the unbalanced panel, we assume that overall, the sample observation period is

$$t = 1,2,..., T$$

and that the 'Time' variable gives for the specific group, the particular values of t that apply to the observations. Thus, suppose your overall sample is five periods. The first group is three observations, periods 1, 2, 4, while the second group is four observations, 2, 3, 4, 5. Then, your panel specification would be

; Pds = Ti, for example, where *Ti* = (3, 3, 3), (4, 4, 4, 4)
and **; Time = Pd,** for example, where *Pd* = (1, 2, 4), (2, 3, 4, 5).

NOTE: See the discussion in Section E30.4.2 that describes how this model is estimated. It places an important restriction on the two way fixed effects model.

Standard Model Specifications for the Fixed Effects Binary Choice Models

This is the full list of general specifications supported for this model.

Controlling Output from Model Commands

; Par keeps ancillary parameter matrix *alphafe* containing fixed effects.
; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Prob = name saves probabilities as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Starting values for the iterations are obtained by fitting the basic model without fixed effects. Thus, the initial results in the output for these models will be the binary choice models discussed in the preceding chapters. You will see a constant term in these results even though you have not included one in your commands. This is used to get starting value for the fixed effects. Iterations begin with the restricted model that forces all the fixed effects to equal the constant term in the restricted model. You may provide your own starting values for the slope parameters with

; Start = ... the list of values for β.

Do not include a set of constants. You may also provide a starting value which will be used identically for all the fixed effects by including one extra value at the end of your list of starting values.

Results that are kept for this model are

Matrices:	*b*	= estimate of $\boldsymbol{\beta}$
	varb	= asymptotic covariance matrix for estimate of $\boldsymbol{\beta}$.
	alphafe	= estimated fixed effects if the command contains **; Parameters**
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function
Last Model:	*b_variables*	
Last Function:	None	

The upper limit on the number of groups is 100,000. Technical details on the method of estimation appear in Section E30.4.2. Partial effects are computed locally with **; Partial Effects** in the command. The post estimation **PARTIAL EFFECTS** command does not have the set of constant terms, some of which are infinite, so the probabilities cannot be computed.

E30.4.1 Application

The gender and kids present dummy variables are time invariant and are omitted from the model. Nonlinear models are like linear models in that time invariant variables will prevent estimation. This is not due to the ‘within’ transformation producing columns of zeros. The within transformation of the data is not used for nonlinear models. A similar effect does arise in the derivatives of the log likelihood, however, which will halt estimation because of a singular Hessian.

The results of fitting models with no fixed effects, with the person specific effects and with both person and time effects are listed below. The results are partially reordered to enable comparison of the results, and some of the results from the pooled estimator are omitted.

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = age,educ,hhninc,newhsat $
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects $
PROBIT      ; Lhs = doctor ; Rhs = x
            ; FEM
            ; Panel
            ; Parameters
            ; Partial Effects $
PROBIT      ; Lhs = doctor ; Rhs = x
            ; FEM
            ; Panel
            ; Time Effects
            ; Parameters
            ; Partial Effects $
```

These are the results for the pooled data without fixed effects.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                 DOCTOR
Log likelihood function    -16639.23971
Restricted log likelihood  -18019.55173
Chi squared [   4 d.f.]      2760.62404
Significance level               .00000
McFadden Pseudo R-squared      .0766008
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =33288.479 AIC/N =    1.218
Hosmer-Lemeshow chi-squared =  20.51061
P-value=  .00857 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|     .00856***      .00074    11.57  .0000      .00711    .01001
    EDUC|    -.01540***      .00358    -4.30  .0000     -.02241   -.00838
  HHNINC|    -.00668         .04657     -.14  .8859     -.09795    .08458
 NEWHSAT|    -.17499***      .00396   -44.21  .0000     -.18275   -.16723
Constant|    1.35879***      .06243    21.77  .0000     1.23644   1.48114
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the estimates for the one way fixed effects model.

```
-----------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable                 DOCTOR
Log likelihood function     -9187.45120
Estimation based on N =  27326, K =4251
Inf.Cr.AIC  =26876.902 AIC/N =     .984
Unbalanced panel has   7293 individuals
Skipped 3046 groups with inestimable ai
PROBIT (normal)  probability model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|     .04701***      .00438    10.74  .0000      .03844    .05559
    EDUC|    -.07187*        .04111    -1.75  .0804     -.15244    .00870
  HHNINC|     .04883         .10782      .45  .6506     -.16249    .26015
 NEWHSAT|    -.18143***      .00805   -22.53  .0000     -.19721   -.16564
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

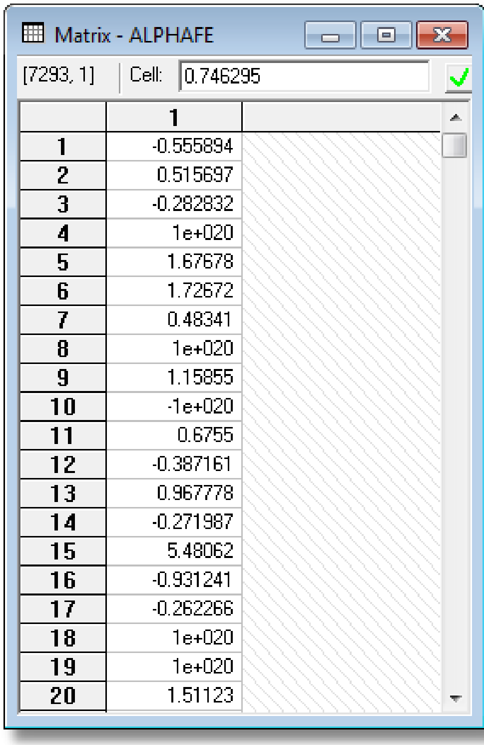


Figure E30.3 Estimated Fixed Effects

Note that the results report that 3046 groups had inestimable fixed effects. These are individuals for which the Lhs variable, *doctor*, was the same in every period, including 1525 groups with $T_i = 1$. If there is no within group variation in the dependent variable for a group, then the fixed effect for that group cannot be estimated, and the group must be dropped from the sample. The **; Parameters** specification requests that the estimates of α_i be kept in a matrix, *alphafe*. Groups for which α_i is not estimated are filled with the value -1.E20 if y_{it} is always zero and +1.E20 if y_{it} is always one, as shown above.

The log likelihood function has increased from -16,639.24 to -9187.45 in computing the fixed effects model. The chi squared statistic is twice the difference, or 14,903.57. This would far exceed the critical value for 95% significance, so at least at first take, it would seem that the hypothesis of no fixed effects should be rejected. There are two reasons why this test would be invalid. First, because of the incidental parameters issue, the fixed effects estimator is inconsistent. As such, the statistic just computed does not have precisely a chi squared distribution, even in large samples. Second, the fixed effects estimator is based on a reduced sample. If the test were valid otherwise, it would have to be based on the same data set. This can be accomplished by using the commands

```
CREATE   ; meandr = Group Mean(doctor, Str = id) $
REJECT   ; meandr < .1 | meandr > .9 $
PROBIT   ; Lhs = doctor ; Rhs = one,x $
```

(The mean value must be greater than zero and less than one. For groups of seven, it can be as high as 6/7 = .86.) Using the reduced sample, the log likelihood for the pooled sample would be -10,852.71. The chi squared is 11,573.31 which is still extremely large. But, again, the statistic does not have the large sample chi squared distribution that allows a formal test. It is a rough guide to the results, but not precise as a formal rule for building the model.

In order to compute marginal effects, it is necessary to compute the index function, which does require an α_i. The mean of the estimated values is used for the computation. The results for the pooled data are shown for comparison below the fixed effects results.

These are the partial effects for the fixed effects model.

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]   with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Estimated E[y|means,mean alphai]=    .625
Estimated scale factor for dE/dx=    .379
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01783***     1.22903     6.39  .0000      .01237    .02330
    EDUC|    -.02726        -.49559    -1.40  .1628     -.06554    .01102
  HHNINC|     .01852         .01048      .45  .6542     -.06253    .09957
 NEWHSAT|    -.06882***     -.77347    -5.96  .0000     -.09144   -.04619
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the partial effects for the pooled model.

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
Average partial effects for sample obs.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00297***      .20554    11.66  .0000      .00247    .00347
    EDUC|    -.00534***     -.09618    -4.30  .0000     -.00778   -.00291
  HHNINC|    -.00232        -.00130     -.14  .8859     -.03401    .02937
 NEWHSAT|    -.06075***     -.65528   -49.40  .0000     -.06316   -.05834
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the two way fixed effects estimates. The time effects, which are usually few in number, are shown in the model results, unlike the group effects.

```
-----------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable               DOCTOR
Log likelihood function     -9175.69958
Estimation based on N =  27326, K =4257
Inf.Cr.AIC  =26865.399 AIC/N =     .983
Model estimated: Jun 15, 2011, 11:00:11
Unbalanced panel has   7293 individuals
Skipped 3046 groups with inestimable ai
No. of period specific effects= 6
PROBIT (normal)  probability model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|     .03869***      .01310     2.95  .0031      .01301    .06437
    EDUC|    -.07985*        .04130    -1.93  .0532     -.16080    .00109
  HHNINC|     .05329         .10807      .49  .6219     -.15852    .26510
 NEWHSAT|    -.18090***      .00806   -22.44  .0000     -.19670   -.16510
 Period1|    -.08649         .15610     -.55  .5795     -.39244    .21946
 Period2|    -.00782         .13926     -.06  .9552     -.28076    .26513
 Period3|     .08766         .12423      .71  .4804     -.15583    .33116
 Period4|     .03048         .10907      .28  .7799     -.18330    .24425
 Period5|    -.02437         .09372     -.26  .7948     -.20807    .15932
 Period6|     .05075         .07761      .65  .5131     -.10136    .20287
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]   with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Estimated E[y|means,mean alphai]=    .625
Estimated scale factor for dE/dx=    .379
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01467***     1.01123     4.35  .0000      .00806    .02129
    EDUC|    -.03029        -.55056    -1.49  .1370     -.07021    .00964
  HHNINC|     .02021         .01144      .48  .6289     -.06176    .10218
 NEWHSAT|    -.06861***     -.77109    -4.34  .0000     -.09962   -.03761
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E30.4.2 Technical Details

The fixed effects model is fit essentially by ‘brute force.’ *LIMDEP* actually estimates the full $K + N$ up to 100,150 coefficients by Newton’s method. It is possible to fit the huge number of coefficients because we take advantage of the properties of the sparse second derivatives matrix of the log likelihood. One of the implications, however, is that there is no covariance matrix computed for the fixed effects. As such, it is not possible to do any kind of inference for individual fixed effects.

The two way fixed effects estimator is computed by actually creating the time specific dummy variables and adding them to the model – see the results above. This means that the usual 150 parameter limit on model size applies to the number of variables in the model plus the number of periods (minus one).

Marginal effects in the fixed effects model are computed at the means of the data and with the sample average of the fixed effects estimates as the constant term.

The unconditional log likelihood is maximized by using Newton’s method. A full discussion of the method is given in Chapter R23. A short sketch of the result is given here, for the logit model. (The results for the other binary choice models are similar. For the models that have asymmetric probability functions, complementary log log, Gompertz and arctangent, the expressions below become more complicated as the zeros and ones are treated separately as required.) The log likelihood is

$$\log L = \sum_{i=1}^{n} \log\left[\prod_{t=1}^{T_i} \Lambda[(2y_{it}-1)(\alpha_i + \beta'\mathbf{x}_{it})]\right]$$

Let p_{it}, y_{it}, $\mathbf{x}_{it}$ and $q_{it} = 2y_{it} - 1$ denote the obvious components of this function. Then,

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it}-p_{it})\mathbf{x}_{it} = \mathbf{g}_{\beta}$$

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{t=1}^{T_i}(y_{it}-p_{it}) = g_i$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\sum_{i=1}^{N}\sum_{t=1}^{T_i} p_{it}(1-p_{it})\mathbf{x}_{it}\mathbf{x}_{it}' = \mathbf{H}_{\beta\beta'}$$

$$\frac{\partial^2 \log L}{\partial \alpha_i^2} = -\sum_{t=1}^{T_i} p_{it}(1\text{-}p_{it}) = h_{ii}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \alpha_i} = -\sum_{t=1}^{T_i} p_{it}(1\text{-}p_{it})\mathbf{x}_{it} = \mathbf{h}_{\beta i}$$

Assemble the full set of first derivatives in a $(K+N)\times 1$ vector, $\mathbf{g}$ and the full set of second derivatives in a $(K+N)\times(K+N)$ matrix, $\mathbf{H}$. The iteration for Newton’s method is

$$\begin{aligned}\boldsymbol{\gamma}_{s+1} &= \boldsymbol{\gamma}_s - \mathbf{H}_s^{-1}\mathbf{g}_s \\ &= \boldsymbol{\gamma}_s + \mathbf{d}_s,\end{aligned}$$

where $\boldsymbol{\gamma}$ denotes the full $(K+N)\times 1$ parameter vector, $(\boldsymbol{\beta}',\alpha_1,\alpha_2,...,\alpha_N)'$ and s indexes iterations. This iteration then, computes a change vector, $\mathbf{d}_s$ as the product of the matrix and vector of derivatives. In principle, the matrix $\mathbf{H}$ is huge, which makes this computation unwieldy. However, the lower right $N\times N$ submatrix of $\mathbf{H}$ (the very large part) is a diagonal matrix – see above. Therefore, it is not necessary actually to compute the entire matrix. The change vector can be computed as a sum of $K\times 1$ vectors which are themselves functions only of the scalar diagonal parts of the submatrix and the $K\times K$ submatrix at the upper left, all of which is very easily done and requires no more computer memory than a conventional estimator, say least squares for a regression.

There is an important qualification to be made in the preceding. Consider the first order condition for the ith group specific constant term:

$$\frac{\partial \log L}{\partial \alpha_\iota} = \sum_{t=1}^{T_i} (y_{it} - p_{it}) = g_i$$

Now, suppose for group i, y_{it} is always one. Then, inserting the probability and expanding, this first order condition becomes

$$\frac{\partial \log L}{\partial \alpha_\iota} \left(\sum_{t=1}^{T_i} y_{it} = T_i\right) = T_i - \sum_{t=1}^{T_i} F(\beta'\mathbf{x}_{it} + \alpha_i) = 0$$

In order for this condition to be met, each probability in the last term must equal 1.0, which means, if the data and parameters are finite, α_i must go to $+\infty$. Thus, if every outcome in a group is one, the constant term is not estimable in any binary choice model. The same result occurs if every y_{it} in a group equals zero. Thus, such groups must be dropped from the sample. (See the example above.)

We use Newton's method for the computations, so the actual Hessian is available for estimation of the asymptotic covariance matrix of the estimators. Let $\mathbf{H}_{\beta\alpha'}$ denote the $K\times N$ submatrix of $\mathbf{H}$ obtained as $[\mathbf{h}_{\beta 1}, \mathbf{h}_{\beta 2}, ..., \mathbf{h}_{\beta N}]$ and let $\mathbf{H}_{\alpha\alpha'}$ denote the $N\times N$ diagonal lower right submatrix of $\mathbf{H}$ obtained as diag$[h_{ii}]$. Then, the estimator of the asymptotic covariance matrix for the MLE of $\boldsymbol{\beta}$ is the upper left submatrix of $-\mathbf{H}^{-1}$. Using the partitioned inverse formula, this is

$$\text{Asy.Var}[\mathbf{b}] = [-(\mathbf{H}_{\beta\beta'} - \mathbf{H}_{\beta\alpha'}(\mathbf{H}_{\alpha\alpha'})^{-1}\mathbf{H}_{\alpha\beta'})]^{-1}$$

The first matrix is given above. By inserting the formulas given above, and exploiting the fact that $\mathbf{H}_{\alpha\alpha'}$ is a diagonal matrix, we obtain the result

$$\mathbf{H}_{\beta\alpha'}(\mathbf{H}_{\alpha\alpha'})^{-1}\mathbf{H}_{\alpha\beta'} = \sum_{i=1}^{N} \frac{1}{h_{ii}}(\mathbf{h}_{\beta i})(\mathbf{h}_{\beta i})'.$$

This produces a sum of $K\times K$ matrices which is of the form of a moment matrix and which is easily computed. Thus, the asymptotic covariance matrix for the estimated coefficient vector is easily obtained in spite of the size of the problem. (In fact, for these binary choice models, the Hessian is actually in the form of a 'within groups' moment matrix for a panel. This result is derived in Greene (2011).

Two considerations remain. First, it is not possible to compute the asymptotic covariance matrix for the fixed effects estimator (unless there are relatively few of them). Using the partitioned inverse formula once again, we can show that the elements of Asy.Var[**a**] are contained in

$$\text{Asy.Var}[\mathbf{a}] = [-(\mathbf{H}_{\alpha\alpha'} - \mathbf{H}_{\alpha\beta'}(\mathbf{H}_{\beta\beta'})^{-1}\mathbf{H}_{\beta\alpha'})]^{-1}.$$

The *ij*th element of the matrix to be inverted is

$$(\mathbf{H}_{\alpha\alpha'} - \mathbf{H}_{\alpha\beta'}(\mathbf{H}_{\beta\beta'})^{-1}\mathbf{H}_{\beta\alpha'})_{ij} = \mathbf{1}(i=j)h_{ii} - \mathbf{h}_{\beta i}'(\mathbf{H}_{\beta\beta'})^{-1}\mathbf{h}_{\beta j}$$

This is a full $N \times N$ matrix, and so the model size problem will apply – it is not feasible to manipulate this matrix.

Finally, note that in Asy.Var[**b**], the terms are of order NT minus a sum of N order T outer products. Therefore, the end result is the inverse of an order NT matrix, which will converge to zero. What this establishes is that **b** does converge to a parameter in the sense that its asymptotic covariance matrix converges to zero. However, it converges to a function that deviates from $\boldsymbol{\beta}$ to the extent that plim a_i deviates from α_i. The asymptotic covariance matrix of the fixed effects estimators above is an $N \times N$ matrix that is the inverse of an order T matrix. Since T is fixed and may be very small, the fixed effects estimators are not consistent.

NOTE: Full estimation of the fixed effects model in this fashion encounters the *incidental parameters* problem. Some of the implications of this problem are discussed in Chapter R24. Also, a particular group specific effect, α_i cannot be estimated if y_{it} takes the same value (1 or 0) in every period. If the number of periods is small, this is likely to happen fairly often. You will see an indication in the results of how many such groups had to be dropped from the estimation. See the application above.

Little is known about the impact of the incidental parameters problem on ML estimators of binary choice models beyond the long established 100% bias of the logit estimator in the case of $T = 2$. The following table, extracted from Greene (2004a, pp. 98-119) is as extensive a study of the issue as is currently available. It is based on Monte Carlo analysis of probit and logit models with a continuous variable coefficient, β, and a dummy variable coefficient, δ. While Monte Carlo studies are never definitive, this should provide a moderately good guide to the extent of the problem for binary choice estimators. The table entry estimates the ratio of the expected value of the estimator to the parameter it is estimating for several sample sizes.

Means of empirical sampling distributions, N = 1,000 individuals based on 200 replications.

	$T=2$		$T=3$		$T=5$		$T=8$		$T=10$		$T=20$	
	β	δ	β	δ	β	δ	β	δ	β	δ	β	δ
Logit Coef	2.020	2.027	1.698	1.668	1.379	1.323	1.217	1.156	1.161	1.135	1.069	1.062
Logit ME[a]	1.676	1.660	1.523	1.477	1.319	1.254	1.191	1.128	1.140	1.111	1.034	1.052
Probit Coef	2.083	1.938	1.821	1.777	1.589	1.407	1.328	1.243	1.247	1.169	1.108	1.068
Probit ME[a]	1.474	1.388	1.392	1.354	1.406	1.231	1.241	1.152	1.190	1.110	1.088	1.047
Ord Probit	2.328	2.605	1.592	1.806	1.305	1.415	1.166	1.220	1.131	1.158	1.058	1.068

[a] Average ratio of estimated marginal effect to true marginal effect

Table E30.1 Monte Carlo Simulations of Incidental Parameters Problem

E30.5 Conditional MLE of the Fixed Effects Logit Model

Two nonlinear models, the binomial logit and Poisson regression can be estimated by conditional maximum likelihood. (The MLE of the linear model is the within estimator. In principle, the exponential loglinear regression model also provides the needed sufficient statistics, but we have not seen this model employed in practice.) This is a specialized approach that was devised to deal with the problem of large numbers of incidental parameters discussed in the preceding section. We consider the logit case here and the count models in Section E44.4.1. (This model was studied, among others, by Chamberlain (1980).) The log likelihood for the binomial logit model with fixed effects is

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log\Lambda\left[\left(2y_{it}-1\right)(\boldsymbol{\beta}'\mathbf{x}_{it}+\alpha_i)\right]$$

The first term, $2y_{it}$ - 1, makes the sign negative for y_{it} = 0 and positive for y_{it} = 1, and $\Lambda(.)$ is the logistic probability, $\Lambda(z) = 1/[1 + \exp(-z)]$. Direct maximization of this log likelihood involves estimation of N+K parameters, where N is the number of groups. As N may be extremely large, this is a potentially difficult estimation problem. As we saw in the preceding section, direct estimation with up to 100,000 coefficients is feasible. But, the method discussed here is not restricted – the number of groups is unlimited because the fixed effects coefficients are not estimated. Rather, the fixed effects are conditioned out of the log likelihood. The main appeal of this approach, however, is that whereas the brute force estimator of the preceding section is subject to the incidental parameters bias, the conditional estimator is not; it is consistent even for small T (even for T = 2).

The contribution to the likelihood function of the T_i observations for group i can be conditioned on the sum of the observed outcomes to produce the conditional log likelihood,

$$L_c = \frac{\prod_{t=1}^{T_i} \exp[y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}]}{\sum_{\textit{all arrangements of } T_i \textit{ outcomes with the same sum}} \prod_{s=1}^{T_i} \exp[y_{is}\boldsymbol{\beta}'\mathbf{x}_{is}]}$$

$$= \frac{\exp\left[\sum_{t=1}^{T_i} y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}\right]}{\sum_{\textit{all arrangements of } T_i \textit{ outcomes with the same sum}} \exp\left[\sum_{s=1}^{T_i} d_{is}\boldsymbol{\beta}'\mathbf{x}_{is}\right]}.$$

This function can be maximized with respect to the slope parameters, $\boldsymbol{\beta}$, with no need to estimate the fixed effects parameters. The number of terms in the denominator of the probability may be exceedingly large, as it is the sum of T* terms where T* is equal to the binomial coefficient $\binom{T_i}{S_i}$ and S_i is the sum of the binary outcomes for the ith group. This can be extremely large. The computation of the denominator is accomplished by means of a recursion presented in Krailo and Pike (1984). Let the denominator be denoted $A(T_i,S_i)$. The authors show that for any T and S the function obeys the recursion

$$A(T,S) = A(T\text{-}1,S) + \exp(\mathbf{x}_{iT}'\boldsymbol{\beta})A(T\text{-}1,S\text{-}1)$$

with initial conditions $A(T,s) = 0$ if $T < s$ and $A(T,0) = 1$.

This enables rapid computation of the denominator for T_i up to 200 which is the internal limit. (If your model is this large, expect this computation to be quite time consuming. Although 200 periods (or more) is technically feasible, the number of terms rises geometrically in T_i, and more than 20 or 30 or so is likely to test the limits of the program (as well as your patience). Note, as well that when the sum the observations is zero or T_i, the conditional probability is one, since there is only a single way that each of these can occur. Thus, groups with sums of zero or T_i fall out of the computation. There is one exception. If you are fitting a discrete choice model (see the discussion of **CLOGIT** in Chapter E38) with more than 100 choices, you can use this estimator for models with up to 200 choices. Note in this case, although T_i may be very large, S_i will equal one, so the problem is simple.

Estimation of this model is done with Newton's method. When the data set is rich enough both in terms of variation in $\mathbf{x}_{it}$ and in S_i, convergence will be quick and simple.

E30.5.1 Command

The command for estimation of the model by this method is

LOGIT ; Lhs = dependent variable
; Rhs = dependent variables (do not include one)
; Pds = fixed number of periods or variable for group sizes $

NOTE: You must omit the **; FEM** from the logit command. This is the default panel data estimator for the binary logit model. Use **; Fixed Effects** or **; FEM** to request the unconditional estimator discussed in the previous section.

You may use weights with this estimator. Presumably, these would reflect replications of the observations. Be sure that the weighting variable takes the same value for all observations within a group. The specification would be

; Wts = variable, Noscale

The **Noscaling** option should be used here if the weights are replication factors. If not, then do be aware that the scaling will make the weights sum to the sample size, not the number of groups.

Results that are retained with this estimator are the usual ones from estimation:

Matrices: *b* = estimate of β
varb = asymptotic covariance matrix for estimate of β

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

E30.5.2 Application

The following will fit the binary logit model using the two methods noted. Bear in mind that with $T_i < \underline{7}$, the unconditional estimator is inconsistent and in fact likely to be substantially biased. The conditional estimator is consistent. Based on the simulation results cited earlier, the second results should exceed the first by roughly 40%. Marginal effects are shown as well. Computation is discussed below.

```
NAMELIST    ; x = age,educ,hhninc,newhsat $
LOGIT       ; Lhs = doctor ; Rhs = x,one $
LOGIT       ; Lhs = doctor ; Rhs = x
            ; Panel $          (Chamberlain conditional estimator)
LOGIT       ; Lhs = doctor ; Rhs = x
            ; Panel ; FEM $  (unconditional estimator)
```

These are the pooled estimates.

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function    -16639.86860
Restricted log likelihood  -18019.55173
Chi squared [    4 d.f.]     2759.36627
Significance level               .00000
McFadden Pseudo R-squared      .0765659
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =33289.737 AIC/N =    1.218
Hosmer-Lemeshow chi-squared =  23.04975
P-value=  .00330 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
     AGE|     .01366***      .00121    11.26  .0000      .01128    .01604
    EDUC|    -.02604***      .00585    -4.45  .0000     -.03750   -.01458
  HHNINC|    -.01231         .07670     -.16  .8725     -.16264    .13801
 NEWHSAT|    -.29181***      .00681   -42.86  .0000     -.30515   -.27846
Constant|    2.28922***      .10379    22.06  .0000     2.08580   2.49265
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the conditional maximum likelihood estimates followed by the unconditional fixed effects estimates. For these data, the unconditional estimates are closer to the conditional ones than might have been expected, but still noticeably higher as the received results would predict. The suggested proportionality result also seems to be operating, but with an unbalanced panel, this would not necessarily occur, and should not be used as any kind of firm rule (save, perhaps for the case of $T_i = 2$).

```
+----------------------------------------------------+
| Panel Data Binomial Logit Model                    |
| Number of individuals          =    7293           |
| Number of periods              =TI                 |
| Conditioning event is the sum of DOCTOR            |
+----------------------------------------------------+
```

```
----------------------------------------------------------------------------
Logit Model for Panel Data
Dependent variable                DOCTOR
Log likelihood function      -6092.58175
Estimation based on N =  27326, K =   4
Inf.Cr.AIC  =12193.164 AIC/N =     .446
Hosmer-Lemeshow chi-squared = *********
P-value=  .00000 with deg.fr. =       8
Fixed Effect Logit Model for Panel Data
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------------
     AGE|     .06391***      .00659     9.70  .0000      .05100    .07683
    EDUC|    -.09127         .05752    -1.59  .1126     -.20401    .02147
  HHNINC|     .06121         .16058      .38  .7031     -.25352    .37594
 NEWHSAT|    -.23717***      .01208   -19.63  .0000     -.26086   -.21349
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------+-------------------------------------------------------------------
FIXED EFFECTS Logit  Model
Dependent variable                DOCTOR
Log likelihood function      -9279.06752
Estimation based on N =  27326, K =4251
Inf.Cr.AIC  =27060.135 AIC/N =     .990
Unbalanced panel has   7293 individuals
Skipped 3046 groups with inestimable ai
LOGIT (Logistic) probability model
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Index function for probability
     AGE|     .07925***      .00738    10.74  .0000      .06479    .09372
    EDUC|    -.11803*        .06779    -1.74  .0817     -.25090    .01484
  HHNINC|     .07814         .18102      .43  .6660     -.27665    .43294
 NEWHSAT|    -.30367***      .01376   -22.07  .0000     -.33064   -.27670
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

When the panel is balanced, the estimator also produces a frequency count for the conditioning sums. For example, if we restrict our sample to the individuals who are in the sample for all seven periods, the following table will also appear with the results.

```
+---------------------------------------------------+
| Panel Data Binomial Logit Model                   |
| Number of individuals          =     887          |
| Number of periods              =       7          |
| Conditioning event is the sum of DOCTOR           |
| Distribution of sums over the  7 periods:         |
| Sum         0      1      2      3      4      5      6 |
| Number     48     73     82    100    115    116    151 |
| Pct.     5.41   8.23   9.24  11.27  12.97  13.08  17.02 |
| Sum         7      8      9     10     11     12     13 |
| Number    202      0      0      0      0      0      0 |
| Pct.    22.77    .00    .00    .00    .00    .00    .00 |
+---------------------------------------------------+
```

This count would be meaningless in an unbalanced panel, so it is omitted.

How should you choose which estimator to use? We should note that the two approaches will generally give different numerical answers. The conditional and unconditional log likelihoods are different. In general, you should use the conditional estimator if T is not relatively large. The conditional estimator is less efficient by construction, but consistency trumps efficiency at this level. In addition, if you have more than 50,000 groups, you must use the conditional estimator. If, on the other hand, T is larger than, say, 10, and N is less than 50,000, then the unconditional estimator might be preferred. The additional consideration discussed in the next section might also weigh in favor of the unconditional estimator.

E30.5.3 Estimating the Individual Constant Terms

The conditional fixed effects estimator for the logit model specifically eliminates the fixed effects, so they are not directly estimated. Without them, however, the parameter estimates are of relatively little use. Fitted probabilities and marginal effects will both require some estimate of a constant term. You can request post estimation computation of the fixed effects by using the specification

; Parameters

This saves a matrix named *alphafe* in your matrix work area. This will be a vector with number of elements equal to the number of groups, containing an ad hoc estimate of α_i for the groups for which there is within group variation in $y_{it.}$ We note how this is done. The logit model is

$$\text{Prob}[y_{it} = 1|\mathbf{x}_{it}] = \Lambda(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i) \text{ where } \Lambda(z) = \exp(z)/[1+\exp(z)]$$

After estimation of $\boldsymbol{\beta}$, we treat the $\boldsymbol{\beta}'\mathbf{x}_{it}$ part of this as known, and let $z_{it} = \boldsymbol{\beta}'\mathbf{x}_{it}$. These are now just data. As such, the log likelihood for group i would be

$$\log L_i = \Sigma_t \log \Lambda[(2y_{it} - 1)(z_{it} + \alpha_i)]$$

The likelihood equation for α_i would be

$$\Sigma_t (y_{it} - P_{it}) = 0 \text{ where } P_{it} = \Lambda(z_{it} + \alpha_i)$$

The implicit solution for α_i is given by

$$\Sigma_t y_{it} = \Sigma_t w_{it} / (a_i + w_{it}) \text{ where } w_{it} = \exp(z_{it}) \text{ and } a_i = \exp(-\alpha_i).$$

If y_{it} is always zero or always one in every period, t, then there is no solution to maximizing this function. The corresponding element of *alphafe* will be set equal to -1.d20 or +1.d20 But, if the y_{it}s differ, then the α_i that equates the left and right hand sides can be found by a straightforward search. The remaining rows of *alphafe* will contain the individual specific solutions to these equations. (This is the method that Heckman and MaCurdy (1980) suggested for estimation of the fixed effects probit model.)

We emphasize, this is not the maximum likelihood estimator of α_i because the conditional estimator of $\boldsymbol{\beta}$ is not the unconditional MLE. Nor, in fact, is it consistent in N. It is consistent in T_i, but that is not helpful here since T_i is fixed, and presumably small. This estimator is a means to an end. The estimated marginal effects can be based on this estimator – it will give a reasonable estimator of an overall average of the constant terms, which is all that is needed for the marginal effects. Individual predicted probabilities remain ambiguous.

E30.5.4 A Hausman Test for Fixed Effects in the Logit Model

The fixed effects estimator is illustrated with the data used in the preceding examples: Note that the first estimator is the pooled estimator. Under the alternative hypothesis of fixed effects, it is inconsistent. Under the null, it is consistent and efficient. The second estimator is the conditional MLE and the third one is the unconditional fixed effects estimator. The unconditional fixed estimator cannot be used for formal testing because of the incidental parameters problem – it is inconsistent. The pooled estimator and the conditional fixed effects estimator use different samples, so the likelihoods are not comparable. Therefore, testing for the joint significance of the effects is problematic for the conditional estimator. What one can do is use a Hausman test. The test is constructed as follows:

H_0: There are no fixed effects; unconditional ML estimators are $\mathbf{b}_0$ and $\mathbf{V}_0$

H_1: There are fixed effects: conditional ML estimators are $\mathbf{b}_1$ and $\mathbf{V}_1$

Under H_0, $\mathbf{b}_0$ is consistent and efficient, while $\mathbf{b}_1$ is consistent but inefficient. Under H_1, $\mathbf{b}_0$ is inconsistent while $\mathbf{b}_1$ is consistent and efficient. The Hausman statistic would therefore be

$$H = (\mathbf{b}_1 - \mathbf{b}_0)' [\mathbf{V}_1 - \mathbf{V}_0]^{-1}(\mathbf{b}_1 - \mathbf{b}_0)$$

The statistic can be constructed as follows:

```
NAMELIST    ; x = the independent variables, not including one $
LOGIT       ; Lhs =  ... ; Rhs = x, one $
CALC        ; k = Col(x) $
MATRIX      ; b0 = b(1:k) ; v0 = varb(1:k,1:k) $
LOGIT       ; Lhs = ... ; Rhs = x ; Pds = ... ; FEM $
MATRIX      ; b1 = b ; v1 = varb $
MATRIX      ; d = b1 - b0 ; List ; h = d’ * Nvsm(v1, -v0) * d $
```

We apply this to our innovation data by defining *x = imprtshr,fdishare,logsales,relsize,prod* and the dependent variable is *innov*. The remaining commands are generic.

The three sets of parameter estimates were given earlier. The Hausman statistic using the procedure suggested above is

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = age,educ,hhninc,newhsat $
LOGIT       ; Lhs = doctor ; Rhs = x, one $
CALC        ; k = Col(x) $
MATRIX      ; b0 = b(1:k) ; v0 = Varb(1:k,1:k) $
LOGIT       ; Lhs = doctor ; Rhs = x ; Panel $
MATRIX      ; b1 = b ; v1 = varb $
MATRIX      ; d = b1 - b0 ; List ; h = d' * Nvsm(v1, -v0) * d $
```

The final result of the **MATRIX** command is

```
       H|             1
--------+--------------
       1|       98.1550
```

This statistic has four degrees of freedom. The critical value from the chi squared table is 9.49, so based on this test, we would reject the null hypothesis of no fixed effects.

E30.6 Random Effects Models for Binary Choice

The five models we have developed here can also be fit with random effects instead of fixed effects. The structure of the random effects model is

$$z_{it} \mid u_i = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$$

where u_i is the unobserved heterogeneity for the ith individual,

$$u_i \sim N[0,\sigma_u^2],$$

and ε_{it} is the stochastic term in the model that provides the *conditional* distribution.

$$\text{Prob}[y_{it} = 1 \mid \mathbf{x}_{it}, u_i] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + u_i),\ i = 1,...,N,\ t = 1,...,T_i.$$

where $F(.)$ is the distribution discussed earlier (normal, logistic, extreme value, Gompertz). Note that the unobserved heterogeneity, u_i is the same in every period. The parameters of the model are fit by maximum likelihood. As usual in binary choice models, the underlying variance,

$$\sigma^2 = \sigma_u^2 + \sigma_\varepsilon^2$$

is not identified. The reduced form parameter,

$$\rho = \frac{\sigma_u^2}{\sigma_\varepsilon^2 + \sigma_u^2},$$

is estimated directly. With the normalization that we used earlier, $\sigma_\varepsilon^2 = 1$, we can determine

$$\sigma_u = \sqrt{\frac{\rho}{1-\rho}}.$$

Further discussion of the estimation of the structural parameters appears at the end of this section.

The model command for this form of the model is

PROBIT
LOGIT
COMPLOG
GOMPERTZ
ARCTANGENT

; Lhs = dependent variable
; Rhs = independent variables
; Panel
; Random Effects $

NOTE: For this model, your Rhs list should include a constant term, *one*.

Standard Model Specifications for the Random Effects Binary Choice Models

This is the full list of general specifications applicable to this model estimator. See Chapter E1 and references noted there for further details on these specifications.

Controlling Output from Model Commands

; Par keeps ancillary parameter ρ with main parameter β vector in *b*.
; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Hpt = n sets the number of points to use for Hermite quadrature
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Prob = name saves probabilities as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Marginal effects are computed by setting the heterogeneity term, u_i to its expected value of zero. Restrictions may be tested and imposed exactly as in the model with no heterogeneity. Since restrictions can be imposed on all parameters, including ρ, you can fix the value of ρ at any desired value. Do note that forcing the ancillary parameter, in this case, ρ, to equal a slope parameter will almost surely produce unsatisfactory results, and may impede or even prevent convergence of the iterations.

Starting values for the iterations are obtained by fitting the basic model without random effects. Thus, the initial results in the output for these models will be the binary choice models discussed in the preceding sections. You may provide your own starting values for the parameters with

; Start = ... the list of values for β, value for ρ

There is no natural moment based estimator for ρ, so a relatively low guess is used as the starting value instead. The starting value for ρ is approximately .2 ($\theta = [2\rho/(1-\rho)]^{1/2} \approx .29$ – see the technical details below. Maximum likelihood estimates are then computed and reported, along with the usual diagnostic statistics. (An example appears below.) This model is fit by approximating the necessary integrals in the log likelihood function by Hermite quadrature. An alternative approach to estimating the same model is by Monte Carlo simulation. You can do exactly this by fitting the model as a random parameters model with only a random constant term.

Your data might not be consistent with the random effects model. That is, there might be no discernible evidence of random effects in your data. In this case, the estimate of ρ will turn out to be negligible. If so, the estimation program issues a diagnostic and reverts back to the original, uncorrelated formulation and reports (again) the results for the basic model.

Results that are kept for this model are

Matrices:	*b*	= estimate of **β**
	varb	= asymptotic covariance matrix for estimate of **β**
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function
	rho	= estimated value of ρ
	varrho	= estimated asymptotic variance of estimator of ρ
Last Model:	*b_variables, ru*	

Last Function: Prob($y = 1|\mathbf{x},u=0$) (Note: None if you use **; RPM** to fit the RE model.)

The additional specification **; Par** in the command requests that ρ be included in *b* and the additional row and column corresponding to ρ be included in *varb*. If you have included **; Par**, *rho* and *varrho* will also appear at the appropriate places in *b* and *varb*.

NOTE: The hypothesis of no group effects can be tested with a Wald test (simple t test) or with a likelihood ratio test. The LM approach, using **; Maxit = 0** with a zero starting value for ρ does not work in this setting because with ρ = 0, the last row of the covariance matrix turns out to contain zeros.

E30.6.1 Application

The following study fits the probit model under four sets of assumptions. The first uses the pooled estimator, then corrects the standard errors for the clustering in the data. The second is the unconditional fixed effects estimator. The third and fourth compute the random effects estimator, first by quadrature, using the Butler and Moffitt method and the second using maximum simulated likelihood with Halton draws. The output is trimmed in each model to compare only the estimates and the marginal effects.

```
NAMELIST    ; x = age,educ,hhninc,newhsat $
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
PROBIT      ; Lhs  = doctor ; Rhs = x,one ; Partial Effects
            ; Cluster = id $
PROBIT      ; Lhs  = doctor ; Rhs = x ; Partial Effects
            ; Panel ; FEM $
PROBIT      ; Lhs  = doctor ; Rhs = x,one ; Partial Effects
            ; Panel ; Random Effects $
```

The random parameters model described in Chapter E31 provides an alternative estimator for the random effects model based on maximum simulated likelihood rather than with Hermite quadrature. The general syntax is used below for a probit model to illustrate the method.

```
PROBIT      ; Lhs  = doctor ; Rhs = x,one  ; Partial Effects
            ; Panel ; RPM  ; Fcn = one(n) ; Pts = 25 ; Halton $
CALC        ; List ; b(6)^2/(1+b(6)^2) $
```

These are the pooled estimates with corrected standard errors.

```
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of  27326 observations contained   7293 clusters defined by  |
| variable ID       which identifies by a value a cluster ID.         |
+---------------------------------------------------------------------+
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable               DOCTOR
Log likelihood function    -16639.23971
Restricted log likelihood  -18019.55173
Chi squared [   4 d.f.]      2760.62404
Significance level               .00000
McFadden Pseudo R-squared      .0766008
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =33288.479 AIC/N =    1.218
Hosmer-Lemeshow chi-squared =  20.51061
P-value=  .00857 with deg.fr. =       8
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|      .00856***      .00098     8.76  .0000      .00664    .01047
    EDUC|     -.01540***      .00499    -3.09  .0020     -.02517   -.00562
  HHNINC|     -.00668         .05646     -.12  .9058     -.11735    .10398
 NEWHSAT|     -.17499***      .00490   -35.72  .0000     -.18460   -.16539
Constant|     1.35879***      .08475    16.03  .0000     1.19268   1.52491
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The unconditional fixed effects estimates appear next. They differ greatly from the pooled estimates. It is worth noting that under the random effects assumption, neither the pooled nor these fixed effects estimates are consistent.

```
-----------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable               DOCTOR
Log likelihood function     -9187.45120
Estimation based on N =  27326, K =4251
Inf.Cr.AIC  =26876.902 AIC/N =     .984
Model estimated: Jun 15, 2011, 14:02:10
Unbalanced panel has   7293 individuals
Skipped 3046 groups with inestimable ai
PROBIT (normal)  probability model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|      .04701***      .00438    10.74  .0000      .03844    .05559
    EDUC|     -.07187*        .04111    -1.75  .0804     -.15244    .00870
  HHNINC|      .04883         .10782      .45  .6506     -.16249    .26015
 NEWHSAT|     -.18143***      .00805   -22.53  .0000     -.19721   -.16564
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the random effects estimates. The variance of u and correlation parameter ρ are given explicitly in the results. In the MSL random effects estimates that appear next, only the standard deviation of u is given. Squaring the 1.37554428 gives 1.892122, which is nearly the same as the 1.888060 given in the first results. In order to compare the first estimates to the MSL estimates, it is necessary to divide the first by the estimate of $1+\rho$. Thus, the scaled coefficient on *age* in the first set of estimates would be 0.019322; that on *educ* would be -.027611, and so on. Thus, the two sets of estimates are quite similar.

```
-----------------------------------------------------------------------------
Random Effects Binary Probit Model
Dependent variable               DOCTOR
Log likelihood function    -15614.50229
Restricted log likelihood  -16639.23971
Chi squared [   1 d.f.]      2049.47485
Significance level               .00000
McFadden Pseudo R-squared      .0615856
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =31241.005 AIC/N =    1.143
Unbalanced panel has   7293 individuals
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01305***      .00119    10.97  .0000      .01072    .01538
    EDUC|    -.01840***      .00594    -3.10  .0020     -.03005   -.00675
  HHNINC|     .06299         .06387      .99  .3240     -.06218    .18817
 NEWHSAT|    -.19418***      .00520   -37.32  .0000     -.20437   -.18398
Constant|    1.42666***      .09644    14.79  .0000     1.23765   1.61567
     Rho|     .39553***      .01045    37.84  .0000      .37504    .41601
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Coefficients  Probit   Model
Dependent variable               DOCTOR
Log likelihood function    -15619.14356
Restricted log likelihood  -16639.23971
Chi squared [   1 d.f.]      2040.19230
Significance level               .00000
McFadden Pseudo R-squared      .0613067
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =31250.287 AIC/N =    1.144
Model estimated: Jun 15, 2011, 14:04:01
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01288***      .00083    15.58  .0000      .01126    .01450
    EDUC|    -.01823***      .00395    -4.61  .0000     -.02598   -.01048
  HHNINC|     .06741         .05108     1.32  .1870     -.03271    .16752
 NEWHSAT|    -.19383***      .00435   -44.58  .0000     -.20235   -.18531
        |Means for random parameters
Constant|    1.42554***      .06828    20.88  .0000     1.29172   1.55936
        |Scale parameters for dists. of random parameters
Constant|     .80930***      .01088    74.38  .0000      .78797    .83062
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The random parameters approach provides an alternative way to estimate a random effects model. A comparison of the two sets of results illustrates the general result that both are consistent estimators of the same parameters. We note, however, the Hermite quadrature approach produces an estimator of $\rho = \sigma_u^2/(1 + \sigma_u^2)$ while the RP approach produces an estimator of σ_u. To check the consistency of the two approaches, we compute an estimate of ρ based on the RP results. The result below demonstrates the near equivalence of the two approaches.

```
CALC ; List ; b(6)^2/(1+b(6)^2)$
[CALC] *Result*=       .3957574
```

These are the four sets of estimated partial effects.

```
Pooled
-----------------------------------------------------------------------------
Average partial effects for sample obs.
--------+--------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z        |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00297***       .20554     8.83   .0000       .00231    .00363
    EDUC|    -.00534***      -.09618    -3.09   .0020     -.00874   -.00195
  HHNINC|    -.00232         -.00130     -.12   .9058     -.04074    .03610
 NEWHSAT|    -.06075***      -.65528   -39.87   .0000     -.06374   -.05777
--------+--------------------------------------------------------------------
Unconditional Fixed Effects
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]
Estimated E[y|means,mean alphai]=    .625
Estimated scale factor for dE/dx=    .379
--------+--------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z        |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01783***      1.22903     6.39   .0000       .01237    .02330
    EDUC|    -.02726         -.49559    -1.40   .1628     -.06554    .01102
  HHNINC|     .01852          .01048      .45   .6542     -.06253    .09957
 NEWHSAT|    -.06882***      -.77347    -5.96   .0000     -.09144   -.04619
--------+--------------------------------------------------------------------
Random Effects
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]
Observations used for means are All Obs.
--------+--------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z        |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00376***       .25254    11.06   .0000       .00310    .00443
    EDUC|    -.00531***      -.09261    -3.10   .0020     -.00866   -.00195
  HHNINC|     .01817          .00986      .99   .3239     -.01793    .05426
 NEWHSAT|    -.05600***      -.58577   -37.33   .0000     -.05894   -.05306
--------+--------------------------------------------------------------------
```

```
Random Constant Term
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Scale Factor for Marginal Effects   .3541
--------+--------------------------------------------------------------------
        |     Partial                           Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00456***       .28882    11.14  .0000       .00376     .00536
    EDUC|    -.00646***      -.10635    -5.06  .0000     -.00896    -.00396
  HHNINC|     .02387          .01223     1.32  .1882     -.01168     .05942
 NEWHSAT|    -.06864***      -.67771   -33.24  .0000     -.07269    -.06459
--------+--------------------------------------------------------------------
```

E30.6.2 Technical Details for the Random Effects Models

The structure of the random effects model is

$$z_{it} \mid u_i = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + \sigma_u u_i$$

where u_i ~ N[0,1], and ε_{it} is the stochastic term in the model that provides the *conditional* distribution.

$$\text{Prob}[y_{it} = 1 \mid \mathbf{x}_{it}, u_i] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i),\ i = 1,...,N,\ t = 1,...,T_i.$$

where $F(.)$ is the distribution discussed earlier (normal, logistic, extreme value, Gompertz, arctangent). The parameter vector for the random effects model is

$$\boldsymbol{\theta} = [\beta_1,...,\beta_K, \rho]'.$$

With the usual normalization, $\sigma_\varepsilon = 1$ and $\sigma_u = \sqrt{\rho/(1-\rho)}$. The log likelihood function is

$$\log L = \Sigma_i \log L_i$$

where $\log L_i$ is the contribution of the ith individual (group) to the total. Conditioned on u_i, the joint probability for the ith group is

$$\text{Prob}[Y_{i1} = y_{i1},...,Y_{iTi} = y_{iTi} \mid \mathbf{x}_{i1,...,}u_i] = \prod_{t=1}^{T_i} F\left[\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i\right]^{y_{it}} \left(1 - F\left[\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i\right]\right)^{1-y_{it}}$$

where now, u_i is normalized to unit variance. Since u_i is unobserved, it is necessary to obtain the unconditional log likelihood by taking the expectation of this over the distribution of u_i. For convenience, write the tth term in the probability above as $G(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)$, where $\gamma = \sigma_u$, so that

$$L_i \mid u_i = \prod_{t=1}^{T_i} G(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i).$$

Then,

$$L_i = E_{ui}[L_i \mid u_i] = \int_{-\infty}^{\infty} \frac{\exp(-u_i^2/2)}{\sqrt{2\pi}} \prod_{t=1}^{T_i} P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)\, du_i$$

NOTE: It can be seen in the likelihood function that it is necessary actually to compute the product of the probabilities for the group, not the sum of the logs. For this reason, the number of observations in a group cannot be extremely large. Since the probability is likely to be on the order of .25 or so, the product of 100 probabilities is on the order of 10^{-100}. This means that the end result is more rounding error than result. In worse cases, the computation will 'overflow' – that is, exceed the computer's capacity to compute the value. For example, the correct result for the product of 100 probabilities on the order of .01 cannot be computed in the accuracy of the computer, which is about $10^{+/-380}$. The diagnostic that this estimator produces mentions a 'Bad counter...' When the counter for group size exceeds 100, the estimator assumes that you have made some kind of error.

Then, finally,

$$\log L \quad = \quad \sum_{i=1}^{N} \log L_i$$

The function is maximized by solving the likelihood equations:

$$\frac{\partial \log L}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} \;=\; \sum_{i=1}^{N} \frac{\partial \log L_i}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} \;=\; \mathbf{0}.$$

For convenience below, let $\boldsymbol{\theta}$ denote the full parameter vector, $[\boldsymbol{\beta},\gamma]'$.

The integration is done with Hermite quadrature. Make the change of variable to $v_i = u_i/\sqrt{2}$. Then,

$$\log L_i \;=\; \log \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v_i^2) \prod_{t=1}^{T_i} P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta v_i)\; dv_i$$

where $\delta = \gamma\times\sqrt{2}$ [so $\rho = \delta^2/(2 + \delta^2)$] and $\sigma_u = [\rho/(1-\rho)]^{1/2}$]. The integral of the form $\int_{-\infty}^{\infty} \exp(-v^2) g(v) dv$ is approximated by the Hermite quadrature,

$$\int_{-\infty}^{\infty} \exp(-v^2) g(v) dv \;\approx\; \sum_{h=1}^{H} w_h\, g(z_h)$$

where w_h are the weights and z_h are the abscissas for the approximation. (See Section R23.3.1 Butler and Moffitt (1982) and Abramovitz and Stegun (1972) for further details.) Collecting terms, then, the log likelihood is computed with

$$\text{logL} \;\approx\; \sum_{i=1}^{N} \;\log \left\{ \frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h) \right] \right\}$$

The derivatives of the log likelihood function are approximated as well,

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} \approx \sum_{i=1}^{N} \frac{\left\{\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h)\right]\left[\sum_{t=1}^{T_i}\frac{\partial \log P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h)}{\partial \boldsymbol{\theta}}\right]\right\}}{\left\{\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h)\right]\right\}}$$

Note that L_i and its derivatives are approximated separately. The summation involves two separate integrals. We use a 20 point quadrature by default, but you can change the number of quadrature points by including **; Hpt = p** in the command, where 'p' is the desired number of points, from 4 to 96 (even). In some cases, the accuracy of the computations will improve with the number of quadrature points. However, the amount of computation will as well (linearly).

The variance, δ, appears linearly in the function along with **β**, so no complication is added by this additional parameter as the summation is done over the abscissas. In each case, the term

$$P(y_{it}, \boldsymbol{\beta}'\mathrm{x}_{it} + \gamma z_h) \quad = \quad F\left[\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma z_h\right]^{y_{it}}\left(1 - F\left[\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma z_h\right]\right)^{1-y_{it}}$$

so

$$\log P(y_{it}, \boldsymbol{\beta}'\mathrm{x}_{it} + \gamma z_h) \quad = \quad y_{it} \log F_{it} + (1 - y_{it})\log(1 - F_{it}).$$

Thus,

$$\frac{\partial \log P(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h)}{\partial \boldsymbol{\theta}} = \left(\frac{y_{it}}{F_{it}} - \frac{1 - y_{it}}{1 - F_{it}}\right) g_{it}(.)\begin{bmatrix}\mathbf{x}_{it} \\ z_h\end{bmatrix}$$

The forms of the particular density functions, $g_{it}(\bullet)$, differs among the five models. The functional forms appear in Section E27.2.1. Using the functions defined there, the log derivatives, $g(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)$ are as follows:

Probit: $$\frac{(2y_{it} - 1)\phi(\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)}{\Phi[(2y_{it} - 1)(\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)]}$$

Logit: $(2y_{it} - 1)\{1 - \Lambda[(2y_{it} - 1)(\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)]\}$

Comp. log log: $\exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)\times\{ y_{it}[1-C(.)]/C(.) - (1 - y_{it})\}$

Gompertz: $\exp[-(\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)]\times\{y_{it} - (1 - y_{it})G(.)/[1-G(.)]\}$

Arctangent: $$\frac{(2y_i - 1)2}{\pi}\frac{1}{F_i}\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1 + [\exp(\boldsymbol{\beta}'\mathbf{x}_i)]^2}$$

The asymptotic covariance matrix is estimated by the BHHH estimator,

$$\mathbf{H} = \left[\sum_{i=1}^{N}\left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)\left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)'\right]^{-1}$$

E31: Random Parameter Models for Binary Choice

E31.1 Introduction

The parametric binary choice models discussed in Chapter E27 are extended to panel data formats as internal procedures. Four classes of models are supported:

- **Fixed effects:** $\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$,
  α_i correlated with $\mathbf{x}_{it}$,
- **Random effects:** $\text{Prob}[y_{it} = 1] = \text{Prob}[\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i > 0]$,
  u_i uncorrelated with $\mathbf{x}_{it}$,
- **Random parameters:** $\text{Prob}[y_{it} = 1] = F(\boldsymbol{\beta}_i'\mathbf{x}_{it})$,
  $\boldsymbol{\beta}_i \mid i \sim h(\boldsymbol{\beta}|i)$ with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$
- **Latent class:** $\text{Prob}[y_{it} = 1|\text{class j}] = F(\boldsymbol{\beta}_j'\mathbf{x}_{it})$,
  $\text{Prob}[\text{class} = \text{j}] = F_j(\boldsymbol{\theta})$

The first two were developed in Chapter E30. This chapter documents the use of random parameters (mixed) and latent class models for binary choice. Technical details on estimation of random parameters are given in Chapter R24. Technical details for estimation of latent class models are given in Chapter R25.

NOTE: None of these panel data models require balanced panels. The group sizes may always vary.

The random parameters and latent class models do not require panel data. You may fit them with a cross section. If you omit **; Pds** and **; Panel** in these cases, the cross section case, $T_i = 1$, is assumed. (You can also specify **; Pds = 1**.) Note that this group of models (and all of the panel data models described in the rest of this manual) does not use the **; Str** = variable specification for indicating the panel – that is only for **REGRESS**.

The probabilities and density functions supported here are as follows:

Probit

$$F = \int_{-\infty}^{\beta'\mathbf{x}_i} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt = \Phi(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \phi(\boldsymbol{\beta}'\mathbf{x}_i)$$

Logit

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

Complementary log log

$$F = 1 - \exp(-\exp(\boldsymbol{\beta}'\mathbf{x}_i)) = \mathrm{C}(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \exp(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \mathrm{C}(\boldsymbol{\beta}'\mathbf{x}_i)]$$

Gompertz, or type 1 extreme value

$$F = \exp(-\exp(-\boldsymbol{\beta}'\mathbf{x}_i)) = \mathrm{G}(\boldsymbol{\beta}'\mathbf{x}_i) \qquad f = \exp(-\boldsymbol{\beta}'\mathbf{x}_i)\mathrm{G}(\boldsymbol{\beta}'\mathbf{x}_i)$$

Arctangent

$$F = 2/\pi \arctan(\exp(\boldsymbol{\beta}'\mathbf{x}_i)), \qquad f = 2/\pi\ [1/(1 + \exp^2(\boldsymbol{\beta}'\mathbf{x}_i))$$

E31.2 Binary Choice Models with Random Parameters

There is a growing literature on the random parameters modeling approach in transportation studies associated primarily with the discrete choice models described in the *NLOGIT 5 Reference Guide*. We have extended the random parameters model to the binary choice models as well as many other models including the tobit and exponential regression models. Some of the relevant background literature includes Revelt and Train (1998), Train (1998), Brownstone and Train (1999), and Greene (2001a). (In that literature, the models are described under the heading ‘mixed logit’ models. We will require a broader rubric for our purposes.) The structure of the random parameters model is based on the conditional probability

$$\mathrm{Prob}[y_{it} = 1|\, \mathbf{x}_{it}, \boldsymbol{\beta}_i] = F(\boldsymbol{\beta}_i'\mathbf{x}_{it}),\ i = 1,...,N,\ t = 1,...,T_i.$$

where $F(.)$ is the distribution discussed earlier (normal, logistic, extreme value, Gompertz). The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals)

$$\mathrm{E}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

(the second term is optional – the mean may be constant),

$$\mathrm{Var}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] = \boldsymbol{\Sigma}.$$

The model is operationalized by writing

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i \text{ where } \mathbf{v}_i \sim \mathrm{N}[\mathbf{0},\mathbf{I}].$$

As noted earlier, the heterogeneity term is optional. In addition, it may be assumed that some of the parameters are nonrandom. It is convenient to analyze the model in this fully general form here. One can easily accommodate nonrandom parameters just by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$. The command structure for these models makes this simple to do.

NOTE: If there is no heterogeneity in the mean, and only the constant term is considered random – the model may specify that some parameters are nonrandom – then this model is equivalent to the random effects model of the preceding section.

E31.2.1 Command for the Random Parameters Models

The basic model command for this form of the model is

PROBIT
LOGIT
COMPLOG
GOMPERTZ
ARCTANGENT

; Lhs = dependent variable
; Rhs = independent variables
; Panel or Pds = fixed periods or count variable
; RPM
; Fcn = random parameters specification $

NOTE: For this model, your Rhs list should include a constant term.

NOTE: The **; Pds** specification is optional. You may fit these models with cross section data.

Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one, x1, x2, x3, x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be constant. For those that you wish to specify as random, use

; Fcn = variable name (distribution),
variable name (distribution), ...

Three distributions may be specified. All random variables have mean 0.

n = standard normal distribution, variance = 1,
t = triangular (tent shaped) distribution in [-1,+1], variance = 1/6,
u = standard uniform distribution [-1,1], variance = 1/3,
l = lognormal distribution, variance = exp(.5),
o = tent shaped distribution with one anchor at zero
g = log gamma
or *c* = variance = 0. (The parameter is not random.)

Each of these is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. The normal distribution is used most often, but there are several other possibilities. Numerous other formats for random parameters are described in Section R24.3. Those results all apply to the binary choice models. To specify that the constant term and the coefficient on *x*1 are each normally distributed with given mean and variance, use

; Fcn = one(n), x1(n).

This specifies that the first and second coefficients are random while the remainder are not. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

The results include estimates of the means and standard deviations of the distributions of the random parameters and the estimates of the nonrandom parameters. The log likelihood shown in the results is conditioned on the random draws, so one might be cautious about using it to test hypotheses, for example, that the parameters are random at all by comparing it to the log likelihood from the basic model with all nonrandom coefficients. The test becomes valid as R increases, but the 50 used in our application is probably too few. With several hundred draws, one could reliably use the simulated log likelihood for testing purposes.

Correlated Random Parameters

The preceding defines an estimator for a model in which the covariance matrix of the random parameters is diagonal. To extend it to a model in which the parameters are freely correlated, add

; Correlation (or just ; Cor)

to the command. An example appears below.

Heterogeneity in the Means

The preceding examples have specified that the mean of the random variable is fixed over individuals. If there is measured heterogeneity in the means, in the form of

$$E[\beta_{ki}] = \beta_k + \Sigma_m \delta_{km} z_{mi}$$

where z_m is a variable that is measured for each individual, then the command may be modified to

; RPM = list of variables in z

In the data set, these variables must be repeated for each observation in the group. In the application below, we have specified that the random parameters have different means for individuals depending on gender and marital status.

Autocorrelation

You may change the character of the heterogeneity from a time invariant effect to an AR(1) process,

$$v_{kit} = \rho_k v_{ki,t-1} + w_{kit}.$$

(See Section R24.3 for details.)

E31.2.2 Results from the Estimator and Applications

The results produced by this estimator begin with the familiar diagnostic statistics, likelihood function, information criteria, etc. The coefficient estimates are possibly rearranged so that the nonrandom parameters appear first. In the base case of a diagonal covariance matrix, the means of the random parameters appear next, followed in the same order by the estimated scale parameters. The example below illustrates. For normally distributed parameters, these are the standard deviations. For other distributions, these scale factors are multiplied by the relevant standard deviation to obtain the standard deviation of the parameter. For example, if we had specified

; Fcn = educ(u)

in the model command, then the parameter on *educ* would be defined to have mean 1.697 and standard deviation .08084 times 1/sqr(6). (The uniform draw is transformed to be *U*[-1,+1].)

```
SAMPLE     ; All $
SETPANEL   ; Group = id ; Pds = ti $
NAMELIST   ; x = age,educ,hhninc,hsat $
LOGIT      ; Lhs = doctor ; Rhs = x,one
           ; Partial Effects
           ; Panel
           ; RPM
           ; Fcn = one(n),hhninc(n),hsat(n)
           ; Pts = 25
           ; Halton $
```

```
-----------------------------------------------------------------------------
Logit    Regression Start Values for DOCTOR
Dependent variable               DOCTOR
Log likelihood function    -16639.59764
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =33289.195 AIC/N =    1.218
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01366***      .00121    11.25  .0000      .01128    .01603
    EDUC|    -.02603***      .00585    -4.45  .0000    -.03749   -.01457
Constant|    2.28946***      .10379    22.06  .0000    2.08604   2.49288
  HHNINC|    -.01221         .07670     -.16  .8735    -.16254    .13812
    HSAT|    -.29185***      .00681   -42.87  .0000    -.30519   -.27850
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Coefficients  Logit    Model
Dependent variable               DOCTOR
Log likelihood function    -15617.53717
Restricted log likelihood  -16639.59764
Chi squared [   3 d.f.]      2044.12094
Significance level               .00000
McFadden Pseudo R-squared      .0614234
Estimation based on N =  27326, K =   8
Inf.Cr.AIC  =31251.074 AIC/N =    1.144
Unbalanced panel has   7293 individuals
LOGIT (Logistic) probability model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01541***      .00100    15.39  .0000      .01344    .01737
    EDUC|    -.02538***      .00475    -5.34  .0000     -.03469   -.01607
        |Means for random parameters
Constant|    1.77433***      .08285    21.42  .0000     1.61195   1.93671
  HHNINC|     .08517         .06181     1.38  .1682     -.03598    .20632
    HSAT|    -.23532***      .00541   -43.50  .0000     -.24592   -.22471
        |Scale parameters for dists. of random parameters
Constant|    1.37499***      .01982    69.36  .0000     1.33614   1.41384
  HHNINC|     .18336***      .03792     4.84  .0000      .10904    .25768
    HSAT|     .00080         .00204      .39  .6960     -.00319    .00479
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point    .6436
Scale Factor for Marginal Effects   .2294
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00353***      .23902    15.53  .0000      .00309    .00398
    EDUC|    -.00582***     -.10241    -5.36  .0000     -.00795   -.00369
  HHNINC|     .01954         .01069     1.38  .1686     -.00827    .04735
    HSAT|    -.05398***     -.56914   -29.82  .0000     -.05753   -.05043
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

When the random parameters are specified to be correlated, the output is changed. The parameter vector in this case is written

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Gamma}\,\mathbf{v}_i$$

where $\boldsymbol{\Gamma}$ is a lower triangular Cholesky matrix. In this case, the nonrandom parameters and the means of the random parameters are reported as before. The table then reports $\boldsymbol{\Gamma}$ in two parts. The diagonal elements are reported first. These would correspond to the case above. The nonzero elements of $\boldsymbol{\Gamma}$ below the diagonal are reported next, rowwise. In the example below, there are three random parameters, so there are 1 + 2 elements below the main diagonal of $\boldsymbol{\Gamma}$ in the reported results. The covariance matrix for the random parameters in this specification is

$$\text{Var}[\boldsymbol{\beta}_i] = \boldsymbol{\Omega} = \boldsymbol{\Gamma}\mathbf{A}\boldsymbol{\Gamma}'$$

where $\mathbf{A}$ is the known diagonal covariance matrix of $\mathbf{v}_i$. For normally distributed parameters, $\mathbf{A} = \mathbf{I}$. This matrix is reported separately after the tabled coefficient estimates. Finally, the square roots of the diagonal elements of the estimate of $\boldsymbol{\Omega}$ are reported, followed by the correlation matrix derived from $\boldsymbol{\Omega}$. The example below illustrates.

```
LOGIT          ; Lhs = doctor ; Rhs = x,one
               ; Partial Effects
               ; Pds = _groupti
               ; RPM
               ; Fcn = one(n),hhninc(n),newhsat(n)
               ; Correlated
               ; Pts = 25
               ; Halton $
```

```
-----------------------------------------------------------------------------
Random Coefficients  Logit    Model
Dependent variable               DOCTOR
Log likelihood function    -15606.79747
Restricted log likelihood  -16639.59764
Chi squared [   6 d.f.]      2065.60035
Significance level               .00000
McFadden Pseudo R-squared      .0620688
Estimation based on N =  27326, K =  11
Inf.Cr.AIC  =31235.595 AIC/N =    1.143
Unbalanced panel has   7293 individuals
LOGIT (Logistic) probability model
Simulation based on  25 Halton draws
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01471***      .00101    14.61  .0000      .01274    .01668
    EDUC|    -.02740***      .00475    -5.77  .0000     -.03670   -.01810
        |Means for random parameters
Constant|    1.98083***      .08660    22.87  .0000     1.81111   2.15056
  HHNINC|     .09438         .06586     1.43  .1518     -.03470    .22346
    HSAT|    -.25657***      .00615   -41.74  .0000     -.26861   -.24452
        |Diagonal elements of Cholesky matrix
Constant|    1.90753***      .07911    24.11  .0000     1.75248   2.06257
  HHNINC|     .91257***      .08028    11.37  .0000      .75522   1.06991
    HSAT|     .01770***      .00203     8.74  .0000      .01373    .02167
        |Below diagonal elements of Cholesky matrix
lHHN_ONE|    -.00234         .10500     -.02  .9822     -.20813    .20344
lHSA_ONE|    -.08124***      .00932    -8.71  .0000     -.09951   -.06297
lHSA_HHN|     .09466***      .00433    21.88  .0000      .08617    .10314
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Implied covariance matrix of random parameters
Var_Beta|             1             2             3
--------+------------------------------------------
       1|       3.63867    -.00447279      -.154960
       2|    -.00447279       .832783      .0865698
       3|      -.154960      .0865698      .0158724

Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       1.90753
       2|       .912570
       3|       .125986

Implied correlation matrix of random parameters
Cor_Beta|             1             2             3
--------+------------------------------------------
       1|       1.00000    -.00256946      -.644803
       2|    -.00256946       1.00000       .752973
       3|      -.644803       .752973       1.00000

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point    .6464
Scale Factor for Marginal Effects   .2286
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00336***      .22640    14.71  .0000      .00291    .00381
    EDUC|    -.00626***     -.10967    -5.78  .0000     -.00838   -.00414
  HHNINC|     .02157         .01175     1.43  .1522     -.00796    .05110
    HSAT|    -.05864***     -.61557   -27.65  .0000     -.06280   -.05448
--------+--------------------------------------------------------------------
```

Finally, if you specify that there is observable heterogeneity in the means of the parameters with

; RPM = list of variables

then the model changes to

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\,\mathbf{v}_i.$$

The elements of $\boldsymbol{\Delta}$, rowwise, are reported after the decomposition of $\boldsymbol{\Gamma}$. The example below, which contains gender and marital status, illustrates. Note that a compound name is created for the elements of $\boldsymbol{\Delta}$.

```
LOGIT     ; Lhs = doctor ; Rhs = x,one
          ; Partial Effects
          ; Panel
          ; RPM = female,married
          ; Fcn = one(n),hhninc(n),hsat(n)
          ; Correlated
          ; Pts = 25
          ; Halton $
```

```
-----------------------------------------------------------------------------
Random Coefficients  Logit    Model
Dependent variable                DOCTOR
Log likelihood function    -15470.04441
Restricted log likelihood  -16639.59764
Chi squared [  12 d.f.]      2339.10646
Significance level               .00000
McFadden Pseudo R-squared      .0702874
Estimation based on N =  27326, K =  17
Inf.Cr.AIC  =30974.089 AIC/N =    1.134
Model estimated: Jun 15, 2011, 18:43:49
Unbalanced panel has   7293 individuals
LOGIT (Logistic) probability model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01375***      .00104    13.24  .0000      .01171    .01578
    EDUC|    -.00913*        .00488    -1.87  .0613     -.01870    .00043
        |Means for random parameters
Constant|    1.58591***      .12092    13.11  .0000     1.34890   1.82291
  HHNINC|     .10102         .12817      .79  .4306     -.15018    .35223
    HSAT|    -.25929***      .01173   -22.11  .0000     -.28228   -.23630
        |Diagonal elements of Cholesky matrix
Constant|    1.85093***      .07867    23.53  .0000     1.69674   2.00512
  HHNINC|    1.17355***      .08054    14.57  .0000     1.01570   1.33140
    HSAT|     .00147         .00202      .73  .4682     -.00250    .00543
```

```
        |Below diagonal elements of Cholesky matrix
lHHN_ONE|     .15728         .10367      1.52  .1293     -.04592    .36047
lHSA_ONE|    -.06741***      .00926     -7.28  .0000     -.08555   -.04926
lHSA_HHN|     .07996***      .00426     18.78  .0000      .07161    .08831
        |Heterogeneity in the means of random parameters
cONE_FEM|     .26949***      .09017      2.99  .0028      .09276    .44622
cONE_MAR|     .11320         .10064      1.12  .2607     -.08404    .31044
cHHN_FEM|     .10364         .12514       .83  .4075     -.14162    .34891
cHHN_MAR|    -.08432         .13820      -.61  .5418     -.35520    .18655
cHSA_FEM|     .03242***      .01081      3.00  .0027      .01124    .05360
cHSA_MAR|    -.01361         .01218     -1.12  .2638     -.03748    .01026
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Implied covariance matrix of random parameters
Var_Beta|          1             2             3
--------+------------------------------------------
       1|      3.42595       .291109      -.124767
       2|      .291109       1.40195      .0832340
       3|     -.124767      .0832340      .0109393

Implied standard deviations of random parameters
S.D_Beta|          1
--------+--------------
       1|      1.85093
       2|      1.18404
       3|      .104591

Implied correlation matrix of random parameters
Cor_Beta|          1             2             3
--------+------------------------------------------
       1|      1.00000       .132831      -.644484
       2|      .132831       1.00000       .672107
       3|     -.644484       .672107       1.00000

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point    .6687
Scale Factor for Marginal Effects   .2215
--------+--------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00305*        .19821      1.89  .0591     -.00012    .00621
    EDUC|    -.00202       -.03425     -1.28  .1994     -.00511    .00107
  HHNINC|     .02238         .01178       .38  .7014     -.09203    .13679
    HSAT|    -.05744       -.58287      -.70  .4825     -.21776    .10288
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Results saved by this estimator are:

Matrices: *b* = estimate of θ
varb = asymptotic covariance matrix for estimate of θ.
gammaprm = the estimate of Γ
beta_i = individual specific parameters, if **; Par** is requested
sdbeta_i = individual specific parameter standard deviations if **; Par** is requested

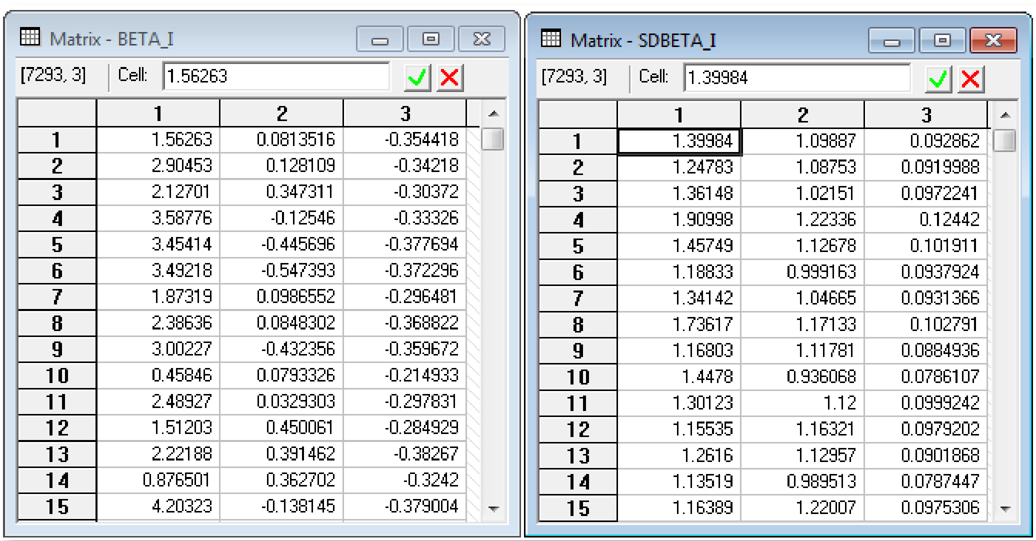
Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

Simulation based estimation is time consuming. The sample size here is fairly large (27,326 observations). We limited the simulation to 25 Halton draws. The amount of computation rises linearly with the number of draws. A typical application of the sort pursued here would use perhaps 300 draws, or 12 times what we used. Estimation of the last model required two minutes and 30 seconds, so in full production, estimation of this model might take 30 minutes. In general, you can get an idea about estimation times by starting with a small model and a small number of draws. The amount of computation rises linearly with the number of draws – that is the main consumer. It also rises linearly with the number of random parameters. The time spent fitting the model will rise only slightly with the number of nonrandom numbers. Finally, it will rise linearly with the number of observations. Thus, a model with a doubled sample and twice as many draws will take four times as long to estimate as one with the original sample and number of draws.

When you include **; Par** in the model command, two additional matrices are created, *beta_i* and *sdbeta_i*. Extensive detail on the computation of these matrices is provided in Section R24.5. For the final specification described above, the results would be as shown in Figure E31.1.



Matrix - BETA_I [7293, 3] Cell: 1.56263

	1	2	3
1	1.56263	0.0813516	-0.354418
2	2.90453	0.128109	-0.34218
3	2.12701	0.347311	-0.30372
4	3.58776	-0.12546	-0.33326
5	3.45414	-0.445696	-0.377694
6	3.49218	-0.547393	-0.372296
7	1.87319	0.0986552	-0.296481
8	2.38636	0.0848302	-0.368822
9	3.00227	-0.432356	-0.359672
10	0.45846	0.0793326	-0.214933
11	2.48927	0.0329303	-0.297831
12	1.51203	0.450061	-0.284929
13	2.22188	0.391462	-0.38267
14	0.876501	0.362702	-0.3242
15	4.20323	-0.138145	-0.379004

Matrix - SDBETA_I [7293, 3] Cell: 1.39984

	1	2	3
1	1.39984	1.09887	0.092862
2	1.24783	1.08753	0.0919988
3	1.36148	1.02151	0.0972241
4	1.90998	1.22336	0.12442
5	1.45749	1.12678	0.101911
6	1.18833	0.999163	0.0937924
7	1.34142	1.04665	0.0931366
8	1.73617	1.17133	0.102791
9	1.16803	1.11781	0.0884936
10	1.4478	0.936068	0.0786107
11	1.30123	1.12	0.0999242
12	1.15535	1.16321	0.0979202
13	1.2616	1.12957	0.0901868
14	1.13519	0.989513	0.0787447
15	1.16389	1.22007	0.0975306

Figure E31.1 Estimated Conditional Parameter Means

E31.2.3 Controlling the Simulation

R is the number of points in the simulation. Authors differ in the appropriate value. Train recommends several hundred. Bhat suggests 1,000 is an appropriate value. The program default is 100. You can choose the value with

; Pts = number of draws, R

The value of 50 that we set in our experiments above was chosen purely to produce an example that you could replicate without spending an inordinate amount of waiting for the results.

The standard approach to simulation estimation is to use random draws from the specified distribution. As suggested immediately above, good performance in this connection requires very large numbers of draws. The drawback to this approach is that with large samples and large models, this entails a huge amount of computation and can be very time consuming. Some authors have documented dramatic speed gains with no degradation in simulation performance through the use of a small number of Halton draws instead of a large number of random draws. Halton sequences are discussed in Section R24.7. Authors (e.g., Bhat (2001)) have found that a Halton sequence of draws with only one tenth the number of draws as a random sequence is equally effective. To use this approach, add

; Halton

to your model command.

In order to replicate an estimation, you must use the same random draws. One implication of this is that if you give the identical model command twice in sequence, you will not get the identical set of results because the random draws in the sequences will be different. To obtain the same results, you must reset the seed of the random number generator with a command such as

CALC ; Ran(seed value) $

(Note that we have used **Ran(12345)** before some of our earlier examples, precisely for this reason. The specific value you use for the seed is not of consequence; any odd number will do.

The random sequence used for the model estimation must be the same in order to obtain replicability. In addition, during estimation of a particular model, the same set of random draws must be used for each person every time. That is, the sequence $\mathbf{v}_{i1}$, $\mathbf{v}_{i2}$, ..., $\mathbf{v}_{iR}$ used for each individual must be same every time it is used to calculate a probability, derivative, or likelihood function. (If this is not the case, the likelihood function will be discontinuous in the parameters, and successful estimation becomes unlikely.) One way to achieve this which has been suggested in the literature is to store the random numbers in advance, and simply draw from this reservoir of values as needed. Because *LIMDEP* is able to use very large samples, this is not a practical solution, especially if the number of draws is large as well. We achieve the same result by assigning to each individual, *i*, in the sample, their own random generator seed which is a unique function of the global random number seed, *S*, and their group number, *i*;

$$\text{Seed}(S,i) = S + 123.0 \times i, \text{ then minus 1.0 if the result is even.}$$

Since the global seed, *S*, is a positive odd number, this seed value is unique, at least within the several million observation range of *LIMDEP*.

E31.2.4 Other Options

Standard Model Specifications for the Random Parameters Binary Choice Models

This is the full list of general specifications applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps individual specific parameter estimates.
; Margin	displays marginal effects.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Prob = name	saves probabilities as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test
; Rst = list imposes equality and fixed value restrictions
; CML: spec imposes linear restrictions on parameters during estimation.

Marginal effects are computed by setting the heterogeneity terms to their expected value of zero.

E31.2.5 The Parameter Vector and Starting Values

Starting values for the iterations are obtained by fitting the basic model without random parameters. Other parameters are set to zero. Thus, the initial results in the output for these models will be the binary choice models discussed in the preceding sections. You may provide your own starting values for the parameters with

; Start = ... the list of values for θ.

The parameter vector is laid out as follows, in this order:

$\alpha_1, ..., \alpha_K$ are the K nonrandom parameters,

$\beta_1,...,\beta_M$ are the M means of the distributions of the random parameters,

$\sigma_1,\sigma_2,...,\sigma_M$ are the M scale parameters for the distributions of the random parameters.

These are the essential parameters. If you have specified that parameters are to be correlated, then the σs are followed by the below diagonal elements of $\mathbf{\Gamma}$. (The σs are the diagonal elements.) If you have specified heterogeneity variables, **z**, then the preceding are followed by the rows of $\mathbf{\Delta}$. Consider an example: The model specifies:

; RPM = z1,z2
; Rhs = one,x1,x2,x3,x4 ? base parameters β_1, β_2, β_3, β_4, β_5
; Fcn = one(n),x2(n),x4(n)
; Cor

Then, after rearranging, the model becomes

Variable	Parameter
$x1$	α_1
$x3$	α_2
one	$\beta_1 + \sigma_1 v_{i1} \qquad\qquad\qquad\qquad + \delta_{11}z_{i1} + \delta_{12}z_{i2}$
$x2$	$\beta_2 + \sigma_2 v_{i2} + \gamma_{21}v_{i1} \qquad\qquad + \delta_{11}z_{i1} + \delta_{12}z_{i2}$
$x4$	$\beta_3 + \sigma_3 v_{i3} + \gamma_{31}v_{i1} + \gamma_{32}v_{i2} + \delta_{11}z_{i1} + \delta_{12}z_{i2}$

and the parameter vector would be

$$\boldsymbol{\theta} = \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2, \sigma_3, \gamma_{21}, \gamma_{31}, \gamma_{32}, \delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}, \delta_{31}, \delta_{32}.$$

You may use **; Rst** and **; CML** to impose restrictions on the parameters. Use the preceding as a guide to the arrangement of the parameter vector. We do note, using **; Rst** to impose fixed value, such as zero restrictions, will generally work well. Other kinds of restrictions, particularly across the parts of the parameter vector, will generally produce unfavorable results.

The variances of the underlying random variables are given earlier, 1 for the normal distribution, 1/3 for the uniform, and 1/6 for the tent distribution. The σ parameters are only the standard deviations for the normal distribution. For the other two distributions, σ_k is a scale parameter. The standard deviation is obtained as $\sigma_k/\sqrt{3}$ for the uniform distribution and $\sigma_k/\sqrt{6}$ for the triangular distribution. When the parameters are correlated, the implied covariance matrix is adjusted accordingly. The correlation matrix is unchanged by this.

E31.2.6 A Dynamic Probit Model

We consider estimation of the dynamic (habit persistence) probit model

$$y_{it}^* = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma y_{i,t-1} + \varepsilon_{it} + \sigma u_i, \ \mathrm{t} = 0,...,T_i, \ i = 1,...,N$$
$$y_{it} = 1(y_{it}^* > 0).$$

Simple estimation of the model by maximum likelihood is clearly inappropriate owing to the random effect. ML random effects is likewise inconsistent because $y_{i,t-1}$ will be correlated with the random effect. Following Heckman (1981), a suggested formulation and procedure for estimation are as follows: Treat the initial condition as an equilibrium, in which

$$y_{i0}^* = \phi + \boldsymbol{\delta}'\mathbf{x}_{i0} + \varepsilon_{i0} + \tau u_i$$
$$y_{i0} = 1(y_{i0*} > 0)$$

and retain the preceding model for periods $1,...,T_i$. Note that the same random effect, u_i appears throughout, but the scaling parameter and the slope vector are different in the initial period. The lagged value of y_{it} does not appear in period 0. This model can be estimated in this form with the random parameters estimator in *LIMDEP*. Use the following procedure. Set up the variables:

d_{it} = 1 in period 1, 0 in all other periods,

f_{it} = $1 - d_{it}$ = 1 in all periods except period 1,

$\mathbf{x}_{it}$ = the set of regressors in the model, 0 in the first period,

$\mathbf{x}_{i0}$ = the set of regressors in the model in period 0, 0 in all other periods,

$y_{i,-1}$ = $y_{i,t-1}$ in periods $1,...,T_i$, 0 in the first period.

Then, the encompassing model is

$$y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\mathbf{x}_{i0} + \phi d_{it} + \alpha f_{it} + \gamma y_{i,-1} + \varepsilon_{it} + \sigma f_{it} u_i + \tau d_{it} u_i,$$
$$y_{it} = 1(y_{it}^* > 0), \ t = 0,1,...,T_i.$$

The commands you might use to set up the data would follow these steps. First, use **CREATE** to set up your group size count variable, *_groupti*.

```
CREATE    ; yit = the dependent variable
          ; yit1 = yit[-1] ? Make sure that yit1 = 0 in the first period.
          ; t = Trn(-ti,1)  or whatever means to set up 1,2,...Ti + 1
          ; dit = (t=1) ; fit = (t > 1) $
CREATE    ; set up the xit and xi0 sets of variables $
```

The estimation command is a random parameters probit model. We make use of a special feature of the RPM that allows the random component of the random parameters to be shared by more than one parameter. This is precisely what is needed to have both τu_i and σu_i appear in the equation without forcing $\tau = \sigma$.

```
PROBIT   ; Lhs = yit
         ; Rhs = xit, xi0, yit1, dit, fit
         ; Panel
         ; RPM
         ; Fcn = dit(n), fit(n)
         ; Common
         ; ... any other desired specifications for the estimation $
```

A refinement of this model assumes that $u_i = \boldsymbol{\lambda}'\mathbf{z}_i + w_i$ for a set of time invariant variables. (See Hyslop (1999) and Greene (2011). One possibility is the vector of group means of the variables $\mathbf{x}_{it}$. (Only the time varying variables would be included in these means.) These can be created and included as additional Rhs variables.

E31.3 Latent Class Models for Binary Choice

The binary choice model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is

$$\text{Prob}[Y_{it} = y_{it} \mid \mathbf{x}_{it}] = F(y_{it},\boldsymbol{\beta}'\mathbf{x}_{it}) = P(i,t),\ y_{it} = 0 \text{ or } 1.$$

Henceforth, we use the term 'group' to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. Unobserved heterogeneity in the distribution of y_{it} is assumed to impact the density in the form of a random effect. The continuous distribution of the heterogeneity is approximated by using a finite number of 'points of support.' The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J 'classes' with a model which allows for heterogeneity as follows: The probability of observing y_{it} given that regime j applies is

$$P(i,t|j) = \text{Prob}[Y_{it} = y_{it}| \mathbf{x}_{it}, j]$$

where the density is now specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$P(i,t|j) = F[y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta_j], \text{Prob}[\text{class} = j] = F_j$$

We formulate this approximation more generally as,

$$P(i,t|j) = F[y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}_j'\mathbf{x}_{it}], F_j = \exp(\theta_j) / \Sigma_j \exp(\theta_j), \text{ with } \theta_J = 0.$$

In this formulation, each group has its own parameter vector, $\boldsymbol{\beta}_j' = \boldsymbol{\beta} + \boldsymbol{\delta}_j$, though the variables that enter the mean are assumed to be the same. (This can be changed by imposing restrictions on the full parameter vector, as described below.) This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters. You may also specify that the latent class probabilities depend on person specific characteristics, so that

$$\theta_{ij} = \boldsymbol{\theta}_j'\mathbf{z}_i, \boldsymbol{\theta}_J = \mathbf{0}.$$

The estimation command for this model is

PROBIT
LOGIT
COMPLOG
GOMPERTZ
ARCTANGENT

; Lhs = ...
; Rhs = independent variables
; LCM (for latent class model)
; Panel $

The default number of support points is five. You may set J from two to nine classes with

; Pts = the value

Use

; LCM = list of variables in $\mathbf{z}_i$

to specify the multinomial logit form of the latent class probabilities.

Standard Model Specifications for the Latent Class Binary Choice Models

This is the full list of general specifications from Chapter E1. Those marked by '*' are not available or not applicable to this model estimator. See Chapter E1 and references noted there for further details on these specifications.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Margin displays marginal effects.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Prob = name saves probabilities as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Some particular values computed for the latent class model are

; Group = the index of the most likely latent class
; Cprob = estimated posterior probability for the most likely latent class

You can obtain a listing of these two results by using

; List

The posterior probabilities for each individual are saved by the following steps:

1. Create a set of variables, pr1=0, pr2=0,… (using any names you wish) so that there is one variable for each class

2. Create a namelist for these variables:

   NAMELIST ; prgroup = pr1,pr2,… $

   Again, use any name you wish.

3. In the model command, include

   ; Classp = the namelist name.

You can use the **; Rst = list** option to structure the latent class model so that different variables appear in different classes. Alternatively, you can use this to force the Heckman and Singer form of the model as follows, where we use a three class model as an example:

```
NAMELIST    ; x = ... one, list of variables $
CALC        ; k1 =  Col(x) - 1 $
LOGIT       ; Lhs = ... ; Rhs = x ; LCM ; Pts = 3
            ; Rst = d1, k1_b, d2, k1_0, d3, k1_0, t1, t2, t3 $
```

Estimates retained by this model include

Matrices: *b* = full parameter vector, $[\beta_1', \beta_2', \ldots F_1, \ldots, F_J]$
varb = full covariance matrix
Note that *b* and *varb* involve $J\times(K+1)$ estimates.

Two additional matrices are created:

b_class = a $J\times K$ matrix with each row equal to the corresponding β_j
class_pr = a $J\times 1$ vector containing the estimated class probabilities

If the command specifies **; Parameters**, then the additional matrix created is:

beta_i = individual specific parameters

Scalars: *kreg* = number of variables in Rhs list
nreg = total number of observations used for estimation
logl = maximized value of the log likelihood function
exitcode = exit status of the estimation procedure

E31.3.1 Application

To illustrate the model, we will fit probit models with three latent classes as alternatives to the continuously varying random parameters models in the preceding section. This model requires a fairly rich data set – it will routinely fail to find a maximum if the number of observations in a group is small. In addition, it will break down if you attempt to fit too many classes. (This point is addressed in Heckman and Singer.)

The model estimates include the estimates of the prior probabilities of group membership. As shown in Section R25.7.1, it is also possible to compute the posterior probabilities for the groups, conditioned on the data. The **; List** specification will request a listing of these. The final illustration below shows this feature for a small subset of the data used above. The models use the following commands: The first is the pooled probit estimator. The second is a basic, three class LCM. The third models the latent class probabilities as functions of the gender and marital status dummy variables. The final model command fits a comparable random parameters model. We will compare the two estimated models.

Fit the pooled probit model first, basic latent class, then latent class with the gender and marital status dummy variables in the class probabilities.

```
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects
            ; Cluster = id $
MATRIX      ; betapool = b' $
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects
            ; Pds = _groupti
            ; LCM
            ; Pts = 3 $
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects
            ; Pds = _groupti
            ; LCM = female,married
            ; Pts = 3
            ; Parameters $
```

Fit the random parameters probit model with heterogeneity in means.

```
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects
            ; Pds = _groupti
            ; RPM = female,married
            ; Fcn = one(n),hhninc(n),newhsat(n)
            ; Correlated
            ; Pts = 25
            ; Halton
            ; Parameters $
```

These are the estimated parameters of the pooled probit model. The cluster correction is shown with the pooled results.

```
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of  27326 observations contained   7293 clusters defined by  |
| variable ID       which identifies by a value a cluster ID.         |
+---------------------------------------------------------------------+
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                DOCTOR
Log likelihood function     -16638.96591
Restricted log likelihood   -18019.55173
Chi squared [    4 d.f.]      2761.17165
Significance level                .00000
McFadden Pseudo R-squared       .0766160
Estimation based on N =  27326, K =    5
Inf.Cr.AIC  =33287.932 AIC/N =     1.218
Hosmer-Lemeshow chi-squared =  20.59314
P-value=  .00831 with deg.fr. =       8
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------
        |Index function for probability
     AGE|      .00855***      .00098     8.75  .0000      .00664    .01047
    EDUC|     -.01539***      .00499    -3.08  .0020     -.02517   -.00561
  HHNINC|     -.00663         .05646     -.12  .9066     -.11729    .10404
    HSAT|     -.17502***      .00490   -35.72  .0000     -.18462   -.16542
Constant|     1.35894***      .08475    16.03  .0000     1.19282   1.52505
--------+--------------------------------------------------------------
```

These are the estimates of the basic three class latent class model.

```
-----------------------------------------------------------------------
Latent Class / Panel Probit   Model
Dependent variable                DOCTOR
Log likelihood function    -15609.05992
Restricted log likelihood  -16638.96591
Chi squared [  13 d.f.]      2059.81198
Significance level               .00000
McFadden Pseudo R-squared      .0618972
Estimation based on N =  27326, K =  17
Inf.Cr.AIC  =31252.120 AIC/N =    1.144
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
Model fit with  3 latent classes.
--------+--------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------
        |Model parameters for latent class 1
     AGE|      .01388***      .00228     6.10  .0000      .00942    .01835
    EDUC|     -.00381         .01146     -.33  .7399     -.02627    .01866
  HHNINC|     -.07299         .15239     -.48  .6320     -.37166    .22569
    HSAT|     -.20115***      .01709   -11.77  .0000     -.23466   -.16765
Constant|     2.08411***      .23986     8.69  .0000     1.61399   2.55424
        |Model parameters for latent class 2
     AGE|      .01336***      .00183     7.29  .0000      .00977    .01696
    EDUC|     -.01886**       .00815    -2.31  .0206     -.03483   -.00289
  HHNINC|      .06824         .10660      .64  .5221     -.14069    .27717
    HSAT|     -.20129***      .00994   -20.26  .0000     -.22076   -.18181
Constant|     1.15407***      .17393     6.64  .0000      .81317   1.49498
        |Model parameters for latent class 3
     AGE|      .00547         .00464     1.18  .2390     -.00363    .01456
    EDUC|     -.04318**       .01911    -2.26  .0239     -.08063   -.00572
  HHNINC|      .30044         .21747     1.38  .1671     -.12579    .72668
    HSAT|     -.14638***      .01965    -7.45  .0000     -.18489   -.10786
Constant|      .24354         .31547      .77  .4401     -.37478    .86186
        |Estimated prior probabilities for class membership
Class1Pr|      .40689***      .04775     8.52  .0000      .31331    .50048
Class2Pr|      .45729***      .03335    13.71  .0000      .39192    .52266
Class3Pr|      .13581***      .02815     4.82  .0000      .08063    .19100
--------+--------------------------------------------------------------
```

The three class latent class model is extended to allow the prior class probabilities to differ by sex and marital status.

```
-----------------------------------------------------------------------------
Latent Class / Panel Probit   Model
Dependent variable               DOCTOR
Log likelihood function    -15471.73843
Restricted log likelihood  -16638.96591
Chi squared [  19 d.f.]      2334.45496
Significance level               .00000
McFadden Pseudo R-squared      .0701502
Estimation based on N =  27326, K =  21
Inf.Cr.AIC  =30985.477 AIC/N =    1.134
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
     AGE|     .01225***      .00240     5.11  .0000      .00755    .01695
    EDUC|     .01438         .01311     1.10  .2725     -.01130    .04007
  HHNINC|    -.02303         .16581     -.14  .8895     -.34801    .30194
    HSAT|    -.17738***      .01802    -9.84  .0000     -.21271   -.14205
Constant|    1.76773***      .25126     7.04  .0000     1.27528   2.26018
        |Model parameters for latent class 2
     AGE|     .00185         .00409      .45  .6508     -.00616    .00986
    EDUC|    -.03067**       .01439    -2.13  .0331     -.05888   -.00245
  HHNINC|     .23788         .18111     1.31  .1890     -.11709    .59285
    HSAT|    -.15169***      .01623    -9.35  .0000     -.18349   -.11989
Constant|     .44044*        .26021     1.69  .0905     -.06957    .95045
        |Model parameters for latent class 3
     AGE|     .01401***      .00199     7.02  .0000      .01010    .01791
    EDUC|    -.00399         .00847     -.47  .6372     -.02060    .01261
  HHNINC|     .03018         .11424      .26  .7916     -.19372    .25408
    HSAT|    -.21215***      .01178   -18.01  .0000     -.23524   -.18906
Constant|    1.13165***      .18329     6.17  .0000      .77241   1.49088
        |Estimated prior probabilities for class membership
   ONE_1|    -.53375**       .21925    -2.43  .0149     -.96347   -.10403
FEMALE_1|    1.18549***      .13400     8.85  .0000      .92284   1.44813
MARRIE_1|    -.33518**       .16234    -2.06  .0390     -.65336   -.01700
   ONE_2|    -.51961*        .26512    -1.96  .0500    -1.03924    .00002
FEMALE_2|    -.31028*        .18197    -1.71  .0882     -.66694    .04638
MARRIE_2|    -.42489**       .18253    -2.33  .0199     -.78265   -.06713
   ONE_3|       0.0    .....(Fixed Parameter).....
FEMALE_3|       0.0    .....(Fixed Parameter).....
MARRIE_3|       0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------

+--------------------------------------------------------------+
|  Prior class probabilities at data means for LCM variables |
|   Class 1     Class 2     Class 3     Class 4     Class 5   |
|    .36905      .17087      .46008      .00000      .00000   |
+--------------------------------------------------------------+
```

Since the class probabilities now differ by observation, the program reports an average using the data means. The earlier fixed prior class probabilities are shown below the averages for this model. The extension brings only marginal changes in the averages, but this does not show the variances across the different demographic segments (female/male, married/single) which may be substantial.

These are the estimated 'individual' parameter vectors.

Matrix - BETA_I

[7293, 5] Cell: 0.0102607

	1	2	3	4	5
1	0.0102607	-0.0116558	0.0920886	-0.192872	0.937109
2	0.0128258	0.00673472	0.000489164	-0.190681	1.50544
3	0.0128752	0.00215468	0.0171934	-0.196419	1.35305
4	0.0123536	0.0108518	-0.0107307	-0.182296	1.64918
5	0.0126052	0.00565629	0.00648296	-0.190281	1.47322
6	0.0125599	0.0100674	-0.00962577	-0.184772	1.61996
7	0.0119442	-0.0033133	0.0459222	-0.195801	1.18672
8	0.0111877	-0.00342029	0.0530367	-0.190218	1.19511
9	0.0131413	4.52367e-005	0.022728	-0.200897	1.27902
10	0.00646755	-0.0197847	0.156178	-0.173777	0.72784
11	0.0125112	0.00632947	0.0047974	-0.188783	1.49699
12	0.0117798	-0.00466994	0.0524595	-0.196149	1.14441
13	0.012093	-0.00625951	0.055629	-0.200375	1.08684
14	0.00631425	-0.020861	0.161567	-0.173881	0.694631
15	0.0123544	0.0131466	-0.0193273	-0.179614	1.72513
16	0.0037759	-0.026427	0.204927	-0.161251	0.550411
17	0.0134314	-0.00181918	0.0271311	-0.205271	1.21273
18	0.0127887	0.00690157	0.000193792	-0.190206	1.51155
19	0.0124609	0.0121641	-0.016595	-0.181569	1.69093
20	0.0126153	0.0104261	-0.0114597	-0.184769	1.63096

Figure E31.2 Latent Class Parameter Estimates

The random parameters model in which parameter means differ by sex and marital status and are correlated with each other is comparable to the full latent class model shown above.

```
-----------------------------------------------------------------------------
Random Coefficients  Probit   Model
Dependent variable              DOCTOR
Log likelihood function    -15469.87914
Restricted log likelihood  -16638.96591
Chi squared [  12 d.f.]      2338.17354
Significance level               .00000
McFadden Pseudo R-squared      .0702620
Estimation based on N =  27326, K =  17
Inf.Cr.AIC  =30973.758 AIC/N =    1.133
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
Simulation based on  25 Halton draws
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01161***      .00086    13.51  .0000      .00993    .01330
    EDUC|    -.00704*        .00407    -1.73  .0833     -.01501    .00093
        |Means for random parameters
Constant|    1.29395***      .09898    13.07  .0000     1.09995   1.48795
  HHNINC|     .08845         .10690      .83  .4080     -.12108    .29798
    HSAT|    -.21458***      .00954   -22.50  .0000     -.23327   -.19589
        |Diagonal elements of Cholesky matrix
Constant|    1.04680***      .04364    23.98  .0000      .96126   1.13234
  HHNINC|     .69686***      .04676    14.90  .0000      .60521    .78851
    HSAT|     .00014         .00120      .12  .9049     -.00220    .00248
        |Below diagonal elements of Cholesky matrix
lHHN_ONE|     .10493*        .05843     1.80  .0725     -.00960    .21946
lHSA_ONE|    -.03295***      .00517    -6.37  .0000     -.04309   -.02282
lHSA_HHN|     .04592***      .00248    18.54  .0000      .04107    .05078
        |Heterogeneity in the means of random parameters
cONE_FEM|     .20456***      .07264     2.82  .0049      .06218    .34694
cONE_MAR|     .07909         .08153      .97  .3320     -.08070    .23888
cHHN_FEM|     .08596         .10341      .83  .4059     -.11672    .28863
cHHN_MAR|    -.07299         .11495     -.63  .5254     -.29828    .15230
cHSA_FEM|     .02966***      .00873     3.40  .0007      .01256    .04677
cHSA_MAR|    -.00931         .00991     -.94  .3474     -.02873    .01011
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Implied covariance matrix of random parameters
Var_Beta|             1              2              3
--------+------------------------------------------
       1|       1.09579        .109842      -.0344941
       2|       .109842        .496629       .0285454
       3|     -.0344941       .0285454      .00319490

Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       1.04680
       2|       .704719
       3|      .0565235

Implied correlation matrix of random parameters
Cor_Beta|             1              2              3
--------+------------------------------------------
       1|       1.00000        .148897       -.582977
       2|       .148897        1.00000        .716624
       3|      -.582977        .716624        1.00000
```

These are the estimated marginal effects from the three models estimated, the pooled probit model, the three class latent class model and a comparable random parameters model, respectively.

```
Pooled
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
Average partial effects for sample obs.
--------+--------------------------------------------------------------------
        |     Partial                           Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00297***      .20548     8.83  .0000       .00231    .00363
    EDUC|    -.00534***    -.09614    -3.09  .0020     -.00873   -.00195
  HHNINC|    -.00230        -.00129     -.12  .9066     -.04072    .03612
    HSAT|    -.06076***    -.65534   -39.87  .0000     -.06375   -.05777
--------+--------------------------------------------------------------------
3 Class Latent Class
--------+--------------------------------------------------------------------
     AGE|     .00446***      .28510     7.28  .0000       .00326    .00566
    EDUC|    -.00572***    -.09511    -2.64  .0082     -.00997   -.00148
  HHNINC|     .01510         .00780      .61  .5433     -.03360    .06381
    HSAT|    -.06917***    -.68884   -19.60  .0000     -.07609   -.06225
--------+--------------------------------------------------------------------
3 Class Heterogeneous Priors
-----------------------------------------------------------------------------
     AGE|     .00406***      .26197     7.00  .0000       .00292    .00520
    EDUC|    -.00064        -.01069     -.27  .7838     -.00519    .00391
  HHNINC|     .01657         .00865      .68  .4953     -.03106    .06420
    HSAT|    -.06804***    -.68420   -20.83  .0000     -.07444   -.06164
--------+--------------------------------------------------------------------
Random Parameters
-----------------------------------------------------------------------------
     AGE|     .00424***      .27768     3.18  .0015       .00162    .00685
    EDUC|    -.00257        -.04379    -1.48  .1385     -.00597    .00083
  HHNINC|     .03226         .01711      .55  .5814     -.08242    .14695
    HSAT|    -.07827        -.79992    -1.22  .2216     -.20379    .04724
--------+--------------------------------------------------------------------
```

E32: Semiparametric and Nonparametric Models for Binary Choice

E32.1 Introduction

This chapter will present three non- and semiparametric estimators for binary choice models. Familiar parametric estimators of binary response models, such as the probit and logit are based on the log likelihood criterion,

$$\log L \;=\; \frac{1}{n}\sum_{i=1}^{n} \log F(y_i \,|\, \boldsymbol{\beta}'\mathbf{x}_i)\;.$$

The Cramer-Rao theory justifies this procedure on the basis of efficiency of the parameter estimates. But, it is to be noted that the criterion is not a function of the ability of the model to predict the response. Moreover, in spite of the widely observed similarity of the predictions from the different models, the issue of which parametric family (normal, logistic, etc.) is most appropriate has never been settled, and there exist no formal tests to resolve the question in any given setting. Various estimators have been suggested for the purpose of broadening the parametric family, so as to relax the restrictive nature of the model specification. Two semiparametric estimators are presented in *LIMDEP*, Manski's (1975, 1985) and Manski and Thompson's (1985, 1987) maximum score (MSCORE) estimator and Klein and Spady's (1993) kernel density estimator.

The MSCORE estimator is constructed specifically around the prediction criterion

Choose $\boldsymbol{\beta}$ to maximize $S \;=\; \Sigma_i\, [y_i^* \;\times z_i^*]$,

where y_i^* = sign (-1/1) of the dependent variable

z_i^* = the sign (-1/1) of $\boldsymbol{\beta}'\mathbf{x}_i$.

Thus, the MSCORE estimator seeks to maximize the number of correct predictions by our familiar prediction rule – predict $y_i = 1$ when the estimated Prob[$y_i = 1$] is greater than .5, *assuming that the true, underlying probability function is symmetric*. In those settings, such as probit and logit, in which the density is symmetric, the sign of the argument is sufficient to define whether the probability is greater or less than .5. For the asymmetric distributions, this is not the case, which suggests a limitation of the MSCORE approach. The estimator does allow another degree of freedom in the choice of a quantile other than .5 for the prediction rule – see the definition below – but this is only a partial solution unless one has prior knowledge about the underlying density.

Klein and Spady's semiparametric density estimator is based on the specification

$$\text{Prob}[y_i = 1] \;=\; P(\boldsymbol{\beta}'\mathbf{x}_i)$$

where P is an unknown, continuous function of its argument with range [0,1]. The function P is not specified a priori; it is estimated with the parameters. The probability function provides the location for the index that would otherwise be provided by a constant term. The estimation criterion is

$$\log L \;=\; \frac{1}{n}\sum_{i=1}^{n} [y_i \log P_n(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)\log(1-P_n(\boldsymbol{\beta}'\mathbf{x}_i))]$$

where P_n is the estimator of P and is computed using a kernel density estimator.

The third estimator is a nonparametric treatment of binary choice based on the index function estimated from a parametric model such as a logit model.

E32.2 Maximum Score Estimation - MSCORE

Maximum score is a semiparametric approach to estimation which is based on a prediction rule. The base case (quantile = ½) is

$$S = \Sigma_i [y_i^* \times z_i^*],$$

where y_i^* is the sign (-1/1) of the dependent variable and z_i^* is the counterpart for the fitted model; z_i^* = the sign (-1/1) of $\boldsymbol{\beta}'\mathbf{x}_i$. Thus, this base case is formulated precisely upon the ability of the sign of the estimated index function to predict the sign of the dependent variable (which, in the binary response models, is all that we observe). Formally, MSCORE maximizes the sample score function

$$\text{Max}_{\boldsymbol{\beta}\varepsilon\mathbf{B}}\, S_{n\alpha}(\boldsymbol{\beta}) = (1/n)\Sigma_i[y_i^* - (1\text{-}2\alpha)]\text{Sgn}(\boldsymbol{\beta}'\mathbf{x}_i),$$

where $\mathbf{B} = \{\boldsymbol{\beta}\ \varepsilon\ R^K : \|\boldsymbol{\beta}\| = 1\}.$

The sample data consist of n observations $[y_i^* ,\mathbf{x}_i]$ where y_i^* is the binary response. Input of y_i is the usual binary variable taking values zero and one; y_i^* is obtained internally by converting zeros to minus ones. The quantile, α, is between zero and one and is provided by the user. The vector $\mathbf{x}_i$ is the usual set of K regressors, usually including a constant. An equivalent problem is to maximize the normalized sample score function

$$S_{N\alpha}^*(\boldsymbol{\beta}) = (1/n)[S_{n\alpha}(\boldsymbol{\beta}) / W_n + 1],$$

where $W_n = (1/n)\Sigma_i w_i$

and $w_i = \text{abs}(y_i^* - (1\text{-}2\alpha)).$

This may then be rewritten as

$$S_{n\alpha}^*(\boldsymbol{\beta}) = \Sigma_i\, w_i^* \times \mathbf{1}[y_i^* = \text{Sgn}(\boldsymbol{\beta}'\mathbf{x}_i)],$$

where $w_i^* = w_i / W_n.$

and $\mathbf{1}[\bullet]$ is the indicator function which equals 1 if the condition in the brackets is true and 0 otherwise. Thus, in the preceding, $\mathbf{1}[\bullet]$ equals 1 if the sign of the index function, $\boldsymbol{\beta}'\mathbf{x}_i$, correctly predicts y_i^*. The normalized sample score function is, thus, a weighted average of the prediction indicators. If α = ½, then w_i^* equals $1/n$, and the normalized score is the fraction of the observations for which the response variable is correctly predicted. Maximum score estimation can therefore be interpreted as the problem of finding the parameters that maximize a weighted average number of correct predictions for the binary response.

The following shows how to use the **MSCORE** command and gives technical details about the procedure. An application is given with the development of **NPREG**, which is a companion program, in Section E32.4.

E32.2.1 Command for MSCORE

The mandatory part of the command for invoking the maximum score estimator

MSCORE ; Lhs = y ; Rhs = x list of independent variables $

The first element of *x* should be *one*. The variable *y* is a binary dependent variable, coded 0/1. The following are the optional specifications for this command. The default values given are used by *LIMDEP* if the option is not specified on the command. **MSCORE** is designed for relatively small problems. The internal limits are 15 parameters and 10,000 observations.

E32.2.2 Options Specific to the Maximum Score Estimator

Quantile

The quantile defines the way the score function is computed. The default of .5 dictates that the score is to be calculated as $(1/n)$ times the number of correctly predicted signs of the response variable. You may choose any value between 0 and 1 with

; Qnt = quantile (default = .5; this is α).

Number of Bootstrap Replications

Bootstrap estimates are computed as follows: After computing the point estimate, **MSCORE** generates R bootstrap samples from the data by sampling n observations with replacement. The entire point estimation procedure, including computation of starting values is repeated for each one. Let $\mathbf{b}$ be the maximum score estimate, R be the number of bootstrap replications, and $\mathbf{d}_i$ be the ith bootstrap estimate. The mean squared deviation matrix,

$$\mathbf{MSD} = (1/R)\Sigma_i\,[(\mathbf{d}_i - \mathbf{b})(\mathbf{d}_i - \mathbf{b})'],$$

is computed from the bootstrap estimates. This is reported in the output as if it were the estimated covariance matrix of the estimates. But, it must be noted that there is no theory to suggest that this is correct. In purely practical terms, the deviations are from the point estimate, not the mean of the bootstrap estimates. The results are merely suggestive. The use of **; Test:** should also be done with this in mind. Use

; Nbt = number of bootstraps (default = 20)

to set the number of bootstrap iterations.

Analysis of Ties

The specification for analysis of ties is

; Ties to analyze ties (default = no)

If the **; Ties** option is chosen, **MSCORE** reports information about regions of the parameter space discovered during the endgame searches for which the sample score is tied with the score at the final estimates. If a tie is found in a region, **MSCORE** records the endpoints of the interval, the current search direction, and some information which records each observation's contribution to the sample score in the region. It is possible to determine whether ties found on separate great circle searches represent disjoint regions or intersections of different great circles. Since the region containing the final estimates is partially searched in each iteration, the tie checking procedure records extensive information about this region. For each region, **MSCORE** reports the minimum and maximum angular direction from the final estimates. These are labeled PSI-low and PSI-high. The parameter values associated with these endpoints are also reported.

If tie regions are found that are far from the point estimate, it may be that the global maximum remains to be found. If so, it may be useful to rerun the estimator using a starting value in the tied region. The existence of many tie regions does not necessarily indicate an unreliable estimate. Particularly in large samples, there may be a large number of disjoint regions in a small neighborhood of the global maximum.

Number of Endgame Iterations

The number of endgame iterations is specified with

; End = number endgame iterations (default = 5)

A given set of great circle searches may miss a direction of increase in the score function. Moreover, even if the trial maximum is a true local maximum, it may not be a global maximum. For these reasons, upon finding a trial maximum, **MSCORE** conducts a user specified number of 'endgame iterations.' These are simply additional iterations of the maximization algorithm. The random search method is such that with enough of these, the entire parameter space would ultimately be searched with probability one. If the endgame iterations provide no improvement in the score, the trial maximum is deemed the final estimate. If an improvement is made during an endgame search, the current estimate is updated as usual and the search resumes. The logic of the algorithm depends on the endgame searches to ensure that all regions of the parameter space are investigated with some probability. The density of the coverage is an increasing function of the number of endgame searches.

There are no formal rules for the number of endgame searches. It should probably increase with K and (perhaps a little less certainly) with n. But, because the step function more closely approximates a continuous population score function, it may be that fewer endgame searches will be needed as N increases.

Starting Values

Starting values are specified with

; Start = starting values (default = none).

If starting values are not provided by the user, they are computed as follows: For each of the K parameters, we form a vector equal to the kth column of an identity matrix. The sample score function is evaluated at this vector, and the kth parameter is set equal to this value. At the conclusion, the starting vector is normalized to unit length. If you do provide your own starting values, they will be normalized to unit length before the iterations are begun.

Technical Output

Technical output is specified with

; Output = 4 or 5 for output of trace of bootstraps to output file
(default = neither).

This is used to control the amount of information about the bootstrap iterations that is produced. This can generate hundreds or thousands of lines of output, depending on the number of bootstrap estimates computed and the number of endgame searches requested. This information is displayed on the screen, in order to trace the progress of execution. In general, the output is not especially informative except in the aggregate. That is, individual lines of this trace are likely to be quite similar. The default is not to retain information about individual bootstraps or endgame searches in the file. Use **; Output = 4** to request only the bootstrap iterations (one line of output per). Use **; Output = 5** to include, in addition, the corresponding information about the endgame searches.

E32.2.3 General Options for MSCORE

The following general options used with the nonlinear estimators in *LIMDEP* are available for MSCORE:

; Covariance Matrix:	to display MSE matrix (default = no), same as **; Printvc**
; List	to display predicted values (default = no list)
; Keep = name	to retain predictions in name (default = no)
; Res = name	to retain fitted values in name (default = no)
; Test: spec	to specify restriction (default = none)
; Maxit = n	to set maximum iterations (default = 50)

Note the earlier caution about the MSD matrix when using the **; Test:** option. The **; Rst = ...** and **; CML:** options for imposing restrictions are not available with this estimator.

E32.2.4 Output from MSCORE

Output from **MSCORE** consists of the following, in the order in which it will appear on your screen or your output file:

1. The iteration summary for the primary estimation procedure (this is labeled bootstrap sample 0') and, if you have requested them, the bootstrap sample estimations. With each one, we report the number of iterations, the number of completed 'endgame iterations' (see the discussion above), the maximum normalized score, and the change in the normalized score.

2. Echo of input parameters in your command.

3. The score function and normalized score function evaluated at three different points:

    a. naive, the first element of β is 1 or -1 and all other values are 0,
    b. the starting values,
    c. the final estimates.

4. The deviations of the bootstrap estimates from the point estimates are summarized in the root mean square error and mean absolute angular deviation between them.

5. The point estimates of the parameters.

NOTE: The estimates are presented in *LIMDEP*'s standard format for parameter estimates. If you have computed bootstrap estimates, the mean square deviation matrix (from the point estimate) is reported as if it were an estimate of the covariance matrix of the estimates. This includes 'standard errors,' 't ratios,' and 'prob. values.' These may, in fact, not be appropriate estimates of the asymptotic standard errors of these parameter estimates. Discussion appears in the references below.

If you change the number of bootstrap estimates, you may observe large changes in these standard errors. This is not to be interpreted as reflecting any changes in the precision of the estimates. If anything, it reflects the unreliability of the bootstrap MSD matrix as an estimate of the asymptotic covariance matrix of the estimates. It has been shown that the asymptotic distribution of the maximum score estimator is not normal. (See Kim and Pollard (1990).) Moreover, even under the best of circumstances, there is no guarantee that the bootstrap estimates or functions of them (such as t ratios), converge to anything useful.

6. A cross tabulation of the predictions of the model vs. the actual values of the Lhs variable.

7. If the model has more than two parameters, and you have requested analysis of the ties, the results of the endgame searches are reported last. Records of ties are recorded in your output file if one is opened, but not displayed on your screen.

The predicted values computed by **MSCORE** are the sign of $\mathbf{b}'\mathbf{x}_i$, coded 0 or 1. Residuals are $y_i - \hat{y}_i$, which will be 1, 0, or -1. The **; List** specification also produces a listing of $\mathbf{b}'\mathbf{x}_i$. The last column of the listing, labeled Prob[y = 1] is the probabilities computed using the standard normal distribution. Since the probit model has not been used to fit the model, these may be ignored.

Results which are saved by **MSCORE** are:

b = final estimates of parameters
varb = mean squared deviation matrix for bootstrap estimates
score = scalar, equal to the maximized value of the score function

The *Last Model* labels are *b_variable*. But, note once again, that the underlying theory needed to justify use of the Wald statistic does not apply here.

E32.2.5 Technical Details

The score function maximized by MSCORE is a step function in contrast to the smooth criterion maximized by, e.g., *LIMDEP*'s probit estimator. As such, the method used here is quite unlike the familiar gradient/search algorithms used for differentiable criteria.

Let $\boldsymbol{\beta}^\circ$ be the current best estimate of $\boldsymbol{\beta}$, and let there be K parameters. MSCORE selects a set of K-1 orthogonal vectors, $\mathbf{c}_1...\mathbf{c}_{K-1}$ all orthogonal to $\boldsymbol{\beta}^\circ$. The score function is then maximized on the great circle connecting $\boldsymbol{\beta}^\circ$ and $\mathbf{c}_1$. The maximum occurs on one or more intervals of positive length on the great circle. If the score is increased relative to that for $\boldsymbol{\beta}^\circ$, the new best estimate becomes the midpoint of the interval. In case of a tie (recall, the score function is a step function), the interval with midpoint closest to the current estimate is chosen. If there is no function improvement, the old estimate is retained. MSCORE then repeats the process with the great circle connecting the new best estimate and $\mathbf{c}_2$. The process is repeated until all K-1 directions have been searched. This process constitutes an iteration. Iterations are continued until no improvement of the function is achieved.

The basis vectors, $\mathbf{c}_1...\mathbf{c}_{K-1}$ are chosen as follows: For each vector $\mathbf{c}_k$, K independent draws from a random number generator are used to produce a vector uniformly distributed on a K-dimensional hypercube $[-1,1]^K$. The K-1 vectors so produced with $\boldsymbol{\beta}^\circ$ are then orthogonalized using the Gram-Schmidt procedure. (Excessively short vectors are discarded and replaced to insure numerical stability.) The vectors are then normalized to unit length. This method insures that all search directions from $\boldsymbol{\beta}^\circ$ are generated with strictly positive probability, but the distribution of directions is not uniform because of the nonlinearity of the transformation. There is an exception to this procedure if an iteration occurs following an iteration in which there was an improvement in the score on two or more of the great circle searches. Let $\boldsymbol{\beta}^\circ$ be the initial estimate and let $\boldsymbol{\beta}^1$ be the final estimate on the most recent iteration. Then the great circle connecting these points is the first direction searched in the current iteration. The remaining directions are searched at random.

We note an important aspect of this procedure. Because the search direction is random and because the criterion is a step function, small changes in the random sample can lead to sizable changes in the parameter estimates. In particular, we have found experimentally that consecutive runs of MSCORE with the same data produced noticeably different parameter estimates, though only occasionally any change in the score. (In almost all cases, the score, itself, was unchanged.) *LIMDEP* imposes one degree of control on this. The seed for the random number generator is always set to the same value upon entry to MSCORE. As such, you will always get the same results with a given data set and model specifications. However, be aware that small changes, (e.g., in the number of observations or in the set of regressors) can bring noticeable changes in the parameter estimates. Once again, 'what counts' is the score function, not the parameters.

When there are only two parameters, the parameter space is the unit circle, and there is only one great circle to search. As such, the algorithm is guaranteed to find the global maximum in one iteration. In problems of higher dimension, convergence cannot be assured in a finite number of iterations.

If bootstrap estimates are computed, we compute the angular deviation between them and the point estimates. (I.e., the angle between them, ArcCos[$(\mathbf{d}_i\mathbf{b}_i/(\mathbf{d}_i'\mathbf{d}_i \times \mathbf{b}_i'\mathbf{b}_i)^{1/2}$].) Since all parameter vectors have unit length, the angular deviation is also the great circle distance between the estimates along the unit hypersphere. The root mean square angular deviation is the square root of the average squared deviation. We also report the mean absolute value of the angular deviations. The units for both are radians. Discussion of this and other computational aspects of this estimator may be found in Manski and Thompson (1985).

E32.2.6 Extensions

The MSCORE procedure may be used to compute Han's (1987) Maximum Rank Correlation (MRC) estimator for binary response models. The MRC estimator is defined to be the value of $\boldsymbol{\beta}$ that maximizes

$$R(\boldsymbol{\beta}) \;=\; \Sigma_{i,j}\,(y_i - y_j) \times \text{Sgn}(\boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\beta}'\mathbf{x}_j).$$

This can be computed by using MSCORE with quantile = .5 on a sample constructed from the original $[y_i,\mathbf{x}_i]$ as follows: For each distinct (i,j) pair for which y_i is not equal to y_j, compute an observation m consisting of

$$d_m = 1 \text{ if } y_i=1 \text{ and } y_j = 0, \text{ and 0 otherwise,}$$

$$\mathbf{x}_m = \mathbf{x}_i - \mathbf{x}_j.$$

This may require some processing outside of *LIMDEP* since this may generate a sample far larger than the original. However, once constructed, the estimation is simple.

Han claims that the estimator is consistent and asymptotically normally distributed. It requires somewhat more stringent assumptions than maximum score. If they are met, the estimator may be more efficient than maximum score.

Manski (1987) analyzes the model

$$y_{i,t} \;=\; \boldsymbol{\beta}'\mathbf{x}_{i,t} + c_i + u_{i,t},\;\; t = 0,1,$$

where c_i is a random effect. The observable indicator $z_{i,t}$ is defined in the usual way for binomial response models; $z_{i,t} = 1$ if $y_{i,t} > 0$, and 0 otherwise. Under the assumptions stated in Manski's paper, the model may be estimated by maximum score by using the reduced sample

$$z_i \;=\; z_{i,1} - z_{i,0}$$

and

$$\mathbf{w}_i \;=\; \mathbf{x}_{i,1} - \mathbf{x}_{i,0}.$$

E32.3 Klein and Spady's Semiparametric Binary Choice Model

Klein and Spady's semiparametric density estimator is based on the specification

$$\text{Prob}[y_i = 1] \;=\; P(\boldsymbol{\beta}'\mathbf{x}_i)$$

where P is an unknown, continuous function of its argument with range [0,1]. The function P is not specified a priori; it is estimated with the parameters. The probability function provides the location for the index that would otherwise be provided by a constant term. The estimation criterion is

$$\log L \;=\; \frac{1}{n}\sum\nolimits_{i=1}^{n}\;[y_i \log P_n(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)\log(1-P_n(\boldsymbol{\beta}'\mathbf{x}_i))]$$

where P_n is the estimator of P and is computed using a kernel density estimator. The probability function is estimated with a kernel estimator,

$$P_n(\boldsymbol{\beta}'\mathbf{x}_i) \;=\; \frac{\sum\nolimits_{j=1}^{n}\;\dfrac{y_j}{h}K\left(\dfrac{\boldsymbol{\beta}'(\mathbf{x}_i-\mathbf{x}_j)}{h}\right)}{\sum\nolimits_{j=1}^{n}\;\dfrac{1}{h}K\left(\dfrac{\boldsymbol{\beta}'(\mathbf{x}_i-\mathbf{x}_j)}{h}\right)}.$$

Two kernel functions are provided, the logistic function, $\Lambda(z)$ and the standard normal CDF, $\Phi(z)$.

As in the other semiparametric estimators, the bandwidth parameter is a crucial input. The program default is $n^{-(1/6)}$, which ranges from .3 to about .6 for n ranging from 30 to 1000. You may provide an alternative value.

E32.3.1 Command

The command for this estimator is

SEMIPARAMETRIC
; Lhs = dependent, binary variable
; Rhs = independent variables $

Do not include one on the Rhs list. The function itself is playing the role of the constant. Optional features include those specific to this model,

; Smooth = desired value for h
; Kernel = Normal – the logistic is standard

and the general ones available with other estimators,

; Partial Effects	
; Prob = name	to retain fitted probabilities
; Keep = name	to retain predictions
; Res = name	to retain residuals
; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**.

The semiparametric log likelihood function is a continuous function of the parameters which is maximized using *LIMDEP*'s standard tools for optimization. Thus, the options for controlling optimization are available,

; Maxit = n	to set maximum iterations
; Output = 1, 2, 3	to control intermediate output
; Alg = name	to select algorithm

Restrictions may be imposed and tested with

; Test: spec	to specify restriction (default = none)
; Rst = list	to specify fixed value and equality restrictions
; CML: spec	to specify other linear constraints

E32.3.2 Output

Output from this estimator includes the usual table of statistical results for a nonlinear estimator. Note that the estimator constrains the constant term to zero and also normalizes one of the slope coefficients to one for identification. This will be obvious in the results. Since probabilities which are a continuous function of the parameters are computed, you may also request marginal effects with

; Partial Effects (or ; Marginal Effects)

Marginal effects are computed using $P_n(\boldsymbol{\beta}'\mathbf{x}_i)$ and its derivatives (which are simple sums) computed at the sample means.

Results Kept by the Semiparametric Estimator

The model results kept by this estimator are

Matrices:	*b*	= final estimates of parameters
	varb	= mean squared deviation matrix for bootstrap estimates
Scalars:	*logl*	= log likelihood
	kreg	= number of Rhs variables
	nreg	= number of observations used to fit the function
	exitcode	= exit status for estimator

Last Model: The labels are *b_variable*

Last Function: None

E32.3.3 Application

The Klein and Spady estimator is computed with the binary logit model. We use only a small subset of the data, the observations that are observed only once. The complete lack of agreement of the two models is striking, though not unexpected.

```
REJECT   ; _groupti > 1 $
SEMI     ; Lhs = doctor
         ; Rhs = one,age,hhninc,hhkids,educ,married
         ; Partial Effects $
LOGIT    ; Lhs = doctor
         ; Rhs = one,age,hhninc,hhkids,educ,married
         ; Partial Effects $
```

```
-----------------------------------------------------------------------------
Semiparametric Binary Choice Model
Dependent variable                DOCTOR
Log likelihood function      -1001.96124
Restricted log likelihood    -1004.77427
Chi squared [   4 d.f.]          5.62607
Significance level                .22887
McFadden Pseudo R-squared       .0027997
Estimation based on N =   1525, K =   4
Inf.Cr.AIC  = 2011.922 AIC/N =    1.319
Hosmer-Lemeshow chi-squared = *********
P-value=  .00000 with deg.fr. =       8
Logistic kernel fn. Bandwidth =  .29475
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Odds Ratio       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Characteristics in numerator of Prob[Y = 1]
     AGE|     .98652         .02284     -.59  .5577      .94176   1.03128
  HHNINC|     .02962**       .04607    -2.26  .0236     -.06067    .11991
  HHKIDS|    3.16366        4.50864      .81  .4190    -5.67311  12.00042
    EDUC|     .96226         .11808     -.31  .7539      .73083   1.19368
 MARRIED|    2.71828    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Odds ratio = exp(beta); z is computed for the original beta
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|    -.00025       -.01488      -.59  .5523     -.00107    .00057
  HHNINC|    -.06479***    -.03782    -76.40  .0000     -.06645   -.06313
  HHKIDS|     .02120        .01063       .26  .7984     -.14148    .18388
    EDUC|    -.00071       -.01305      -.33  .7445     -.00497    .00355
 MARRIED|     .01841    .....(Fixed Parameter).....
```

```
--------------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
--------------------------------------------------------------------------

--------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable               DOCTOR
Log likelihood function      -996.30681
Restricted log likelihood   -1004.77427
Chi squared [   5 d.f.]        16.93492
Significance level               .00462
McFadden Pseudo R-squared      .0084272
Estimation based on N =   1525, K =   6
Inf.Cr.AIC  = 2004.614 AIC/N =    1.315
Hosmer-Lemeshow chi-squared =  10.56919
P-value=  .22732 with deg.fr. =       8
--------+-----------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     .46605          .34260     1.36  .1737     -.20544   1.13754
     AGE|     .00509          .00448     1.14  .2556     -.00369    .01387
  HHNINC|    -.49045*         .26581    -1.85  .0650    -1.01142    .03052
  HHKIDS|    -.36639***       .12639    -2.90  .0037     -.61410   -.11867
    EDUC|     .00783          .02419      .32  .7461     -.03957    .05523
 MARRIED|     .16046          .12452     1.29  .1975     -.08360    .40451
--------+-----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------

--------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
Average partial effects for sample obs.
--------+-----------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCTOR|      Effect    Elasticity      z    |z|>Z*         Interval
--------+-----------------------------------------------------------------
     AGE|     .00117        -.00127     1.14  .2554     -.00085    .00320
  HHNINC|    -.11304*        .00087    -1.85  .0648     -.23301    .00694
  HHKIDS|    -.08606***      .00019    -2.87  .0041     -.14476   -.02736   #
    EDUC|     .00180        -.00053      .32  .7461     -.00912    .01273
 MARRIED|     .03702        -.00057     1.29  .1971     -.01924    .09327   #
--------+-----------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------
```

E32.3.4 Technical Details

The log likelihood function,

$$\log L = \frac{1}{n}\sum_{i=1}^{n} [y_i \log P_n(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)\log(1-P_n(\boldsymbol{\beta}'\mathbf{x}_i))]$$

$$P_n(\boldsymbol{\beta}'\mathbf{x}_i) = \frac{\sum_{j=1}^{n} \frac{y_j}{h} K\left(\frac{\boldsymbol{\beta}'(\mathbf{x}_i - \mathbf{x}_j)}{h}\right)}{\sum_{j=1}^{n} \frac{1}{h} K\left(\frac{\boldsymbol{\beta}'(\mathbf{x}_i - \mathbf{x}_j)}{h}\right)}$$

is easily computed by simple summation for the value of h, $\boldsymbol{\beta}$, and the specified kernel functions,

$$K(.) = \Lambda(.) \quad \text{for the logistic model, with } K'(.) = \Lambda(.)[1 - \Lambda(.)], \text{ or}$$

$$= \Phi(.) \quad \text{for the normal distribution, with } K'(.) = \phi(.).$$

Let the numerator and denominator of $P_n(\boldsymbol{\beta}'\mathbf{x}_i)$ be denoted F_{i0} and F_{i1}, respectively, and let G_{i0} and G_{i1} denote the numerator and denominator computed at $K'(.)$ instead of $K(.)$. Then, let

$$\mathbf{d}_i = \mathbf{x}_i - \mathbf{x}_j.$$

Let

$$g_i = \frac{y_i}{P_n(\boldsymbol{\beta}'\mathbf{x}_i)} - \frac{1-y_i}{1-P_n(\boldsymbol{\beta}'\mathbf{x}_i)}$$

Then, collecting all terms, the vector of derivatives of the log likelihood is

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} g_i \left[\frac{G_{i0}}{F_{i1}} - \frac{F_{i0}G_{i1}}{F_{i1}^2}\right]\mathbf{d}_i = \sum_{i=1}^{n} \mathbf{w}_i$$

The estimator of the asymptotic covariance matrix is the BHHH estimator,

$$\text{Est.Asy.Var} = \left[\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i'\right]^{-1}$$

E32.4 Nonparametric Binary Choice Model

The kernel density estimator is a device used to describe the distribution of a variable nonparametrically, that is, without any assumption of the underlying distribution. This section describes an extension to a simple regression function. The kernel density function estimates any sufficiently smooth regression function, $F_\beta(z) = E[\delta|\boldsymbol{\beta}'\mathbf{x}=z]$, using the method of kernels, for any parameter vector $\boldsymbol{\beta}$. δ must be a response variable with bounded range [0,1]. In the special case in which δ is a binary response taking values 0/1, **NPREG** estimates the probability of a positive response conditional on the linear index $\boldsymbol{\beta}'\mathbf{x}$. With an appropriate choice of $\mathbf{x}$ and $\boldsymbol{\beta}$, and by rescaling the response, this estimator can estimate any sufficiently smooth univariate regression function with known bounded range. One simple approach is to assume that $\mathbf{x}$ is a single variable and $\boldsymbol{\beta}$ equals 1.0, in which case, the estimator describes $E[y_i|x_i]$. Alternatively, **NPREG** may be used with the estimated index function, $\boldsymbol{\beta}'\mathbf{x}_i$, from any binary choice estimator. The natural choice in this instance would be MSCORE, since MSCORE does not compute the probabilities (that is, the conditional mean). In principle, the estimated index function could come from any estimator, but from a probit or other parametric model, this would be superfluous.

The regression function computed is

$$F(z_j) = \frac{\sum_{i=1}^{N} y_i \frac{1}{h} K\left(\frac{z_j - z_i}{h}\right)}{\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{z_j - z_i}{h}\right)} ., \; j = 1,...,M, \; i = 1,..., \text{number of observations.}$$

The function is computed for a specified set of values z_j, j = 1,...,M. Note that each value requires a sum over the full sample of n values. The primary component of the computation is the kernel function, K[.]. Eight alternatives are provided:

1. Epanechnikov:	$K[z]$	$= .75(1 - .2z^2)$ / Sqr(5) if $\lvert z\rvert <= 5$, 0 else,
2. Normal:	$K[z]$	$= \phi(z)$ (normal density),
3. Logit:	$K[z]$	$= \Lambda(z)[1-\Lambda(z)]$ (default),
4. Uniform:	$K[z]$	$= .5$ if $\lvert z\rvert < 1$, 0 1 else,
5. Beta:	$Z[z]$	$= (1-z)(1+z)/24$ if $\lvert z\rvert < 1$, 0 1 else,
6. Cosine:	$K[z]$	$= 1 + \cos(2\pi z)$ if $\lvert z\rvert < .5$, 0 else,
7. Triangle:	$K[z]$	$= 1 - \lvert z\rvert$, if $\lvert z\rvert <= 1$, 0 else.
8. Parzen:	$K[z]$	$= 4/3 - 8z^2 + 8\lvert z\rvert^3$ if $\lvert z\rvert <= .5$, $8(1-\lvert z\rvert)^3$ else.

The other essential part of the computation is the smoothing (bandwidth) parameter, h. Large values of h stabilize the function, but tend to flatten it and reduce the resolution. Small values of h produce greater detail, but also cause the estimator to become less stable.

The basic command is

```
NPREG     ; Lhs = the dependent variable
          ; Rhs = the variable $
```

With no other options specified, the routine uses the logit kernel function, and uses a bandwidth equal to

$$h = .9Q/n^{0.2} \text{ where } Q = \min(\text{std.dev., range}/1.5)$$

You may specify the kernel function to be used with

; Kernel = one of the names of the eight types of kernels listed above

The bandwidth may be specified with

; Smooth = the bandwidth parameter

There is no theory for choosing the right smoothing parameter, λ. Large values will cause the estimated function to flatten at the average value of y_i. Values close to zero will cause the function to pass through the points z_i, y_i and to become computationally unstable elsewhere. A choice might be made on the basis of the *CVMSPE*. (See Wong (1983) for discussion.) A value that minimizes *CVMSPE*(λ) may work well in practice. Since *CVMSPE* is a saved result, you could compute this for a number of values of λ then retrieve the set of values to find the optimal one.

The default number of points specified is 100, with z_j = a partition of the range of the variable. You may specify the number of points, up to 200 with

; Pts = number of points to compute and plot

The range of values plotted is the equally spaced grid from min(x)-h to max(x)+h, with the number of points specified.

E32.4.1 Output from NPREG

Output from **KERNEL** is a set of points for an estimated function, several descriptive statistics, and a plot of the estimated regression function. The added specification

; List

displays the specific results, z_i for the sample observations and the associated estimated regression functions. These values are also placed in a two column matrix named *kernel* after estimation of the function.

The cross validation mean squared prediction error (*CVMSPE*) is a goodness of fit measure. Each observation, '*i*' is excluded in turn from the sample. Using the reduced sample, the regression function is reestimated at the point z_i in order to provide a point prediction for y_i. The average squared prediction error defines the *CVMSPE*. The calculation is defined by

$$F_i^*(z) = \frac{\sum_{j \neq i} \frac{1}{h} y_j K\left(\frac{x_j - x_i}{h}\right)}{\sum_{j \neq i} \frac{1}{h} K\left(\frac{x_j - x_i}{h}\right)}$$

Then, $$CVMSPE(h) = (1/n) \Sigma_i [y_i - F_i^*(x_i)]^2.$$

E32.4.2 Application

The following estimates the parameters of a regression function using **MSCORE**, then uses **NPREG** to plot the regression function.

```
REJECT      ; _groupti > 1 $
NAMELIST    ; x = one,age,hhninc,hhkids,educ,married $
MSCORE      ; Lhs = doctor ; Rhs =x $
CREATE      ; xb = x'b$
NPREG       ; Lhs = doctor ; Rhs = xb $
```

```
-----------------------------------------------------------------------------
Maximum Score Estimates of Linear Quantile
Regression Model from Binary Response Data
Quantile                   .500      Number of Parameters =     6
Observations input    =  1525      Maximum Iterations    =   500
End Game Iterations   =   100      Bootstrap Estimates   =    20
Check Ties?                  No
Save bootstraps?             No
Start values from MSCORE (normalized)
Normal exit after  100 iterations.
Score functions:     Naive   At theta(0)      Maximum
           Raw      .26033       .26033         .27738
    Normalized      .63016       .63016         .63869
Estimated MSEs from  20 bootstrap samples
(Nonconvergence in   0 cases)
Angular deviation (radians) of bootstraps from estimate
Mean square =  1.027841             Mean absolute =   .979001
Standard errors below are based on bootstrap mean squared
deviations.  These and the t-ratios are only approximations.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .42253         .63272      .67  .5043     -.81758   1.66263
     AGE|     .01146         .03120      .37  .7134     -.04969    .07261
  HHNINC|    -.20766         .45880     -.45  .6508    -1.10689    .69157
  HHKIDS|    -.82224         .65955    -1.25  .2125    -2.11494    .47045
    EDUC|     .01446         .07191      .20  .8406     -.12648    .15541
 MARRIED|     .31926         .35336      .90  .3663     -.37331   1.01183
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
+-----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when beta*x is greater than one, zero otherwise.        |
|Note, column or row total percentages may not sum to      |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |      23 (  1.5%)|     541 ( 35.5%)|     564 ( 37.0%)|
|  1   |      10 (   .7%)|     951 ( 62.4%)|     961 ( 63.0%)|
+------+----------------+----------------+----------------+
|Total |      33 (  2.2%)|    1492 ( 97.8%)|    1525 (100.0%)|
+------+----------------+----------------+----------------+
```

```
+----------------------------------------------------------+
|Crosstab for Binary Choice Model.  Predicted probability |
|vs. actual outcome. Entry = Sum[Y(i,j)*Prob(i,m)] 0,1.    |
|Note, column or row total percentages may not sum to      |
|100% because of rounding. Percentages are of full sample.|
+------+----------------------------------+----------------+
|Actual|      Predicted Probability        |                |
|Value |    Prob(y=0)        Prob(y=1)    | Total Actual   |
+------+----------------+-----------------+----------------+
| y=0  |    564 ( 37.0%)|      0 (    .0%)|    564 ( 37.0%)|
| y=1  |    961 ( 63.0%)|      0 (    .0%)|    961 ( 63.0%)|
+------+----------------+-----------------+----------------+
|Total |   1525 (100.0%)|      0 (    .0%)|   1525 (100.0%)|
+------+----------------+-----------------+----------------+
```

```
+---------------------------------------+
| Nonparametric Regression for DOCTOR   |
| Observations        =          1525   |
| Points plotted      =          1525   |
| Bandwidth           =       .090121   |
| Statistics for abscissa values----    |
| Mean                =       .854823   |
| Standard Deviation =        .433746   |
| Minimum             =      -.167791   |
| Maximum             =      1.662874   |
| ----------------------------------    |
| Kernel Function     =      Logistic   |
| Cross val. M.S.E.   =       .231635   |
| Results matrix      =        KERNEL   |
+---------------------------------------+
```

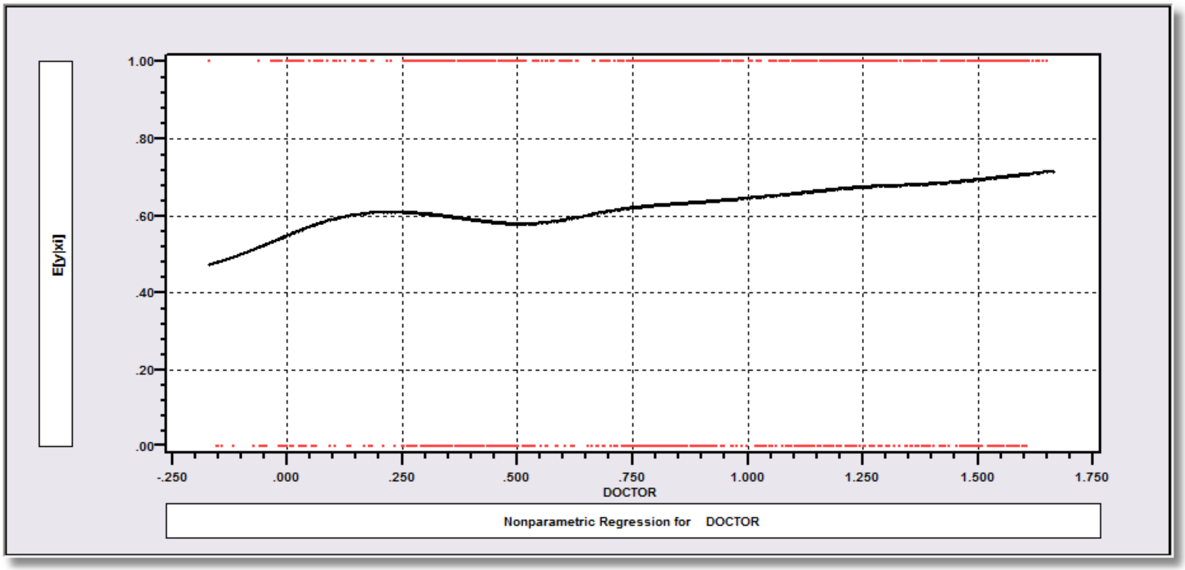


Figure E32.1 Nonparametric Regression

E33: Bivariate and Multivariate Probit and Partial Observability Models

E33.1 Introduction

The basic formulation of the models in this chapter is the bivariate probit model:

$$z_{i1} = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1},\ \ y_{i1} = 1 \text{ if } z_{i1} > 0,\ y_{i1} = 0 \text{ otherwise,}$$

$$z_{i2} = \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \varepsilon_{i2},\ \ y_{i2} = 1 \text{ if } z_{i2} > 0,\ y_{i2} = 0 \text{ otherwise,}$$

$[\varepsilon_{i1},\varepsilon_{i2}]$ ~ bivariate normal (BVN) $[0,0,1,1,\rho]$, $-1 < \rho < 1$,

individual observations on y_1 and y_2 are available for all i.

(This model is also available for grouped (proportions) data. See Section E33.2.3.) The model given above would be estimated using a complete sample on $[y_1, y_2, \mathbf{x}_1, \mathbf{x}_2]$ where y_1 and y_2 are binary variables and $\mathbf{x}_{ij}$ are sets of regressors. This chapter will describe estimation of this model and several variants:

- The disturbances in one or both equations may be heteroscedastic.
- The observation mechanism may be such that y_{i1} is not observed when y_{i2} equals zero.
- The observation mechanism may be such that only the product of y_{i1} and y_{i2} is observed. That is, we only observe the compound outcomes 'both variables equal one' or 'one or both equal zero.'
- The basic model is extended to as many as 20 equations as a *multivariate probit model*.

NOTE: It is not necessary for there to be different variables in the two (or more) equations. The Rh1 and Rh2 lists may be identical if your model specifies that. There is no issue of identifiability or of estimability of the model – the variable lists are unrestricted. This is not a question of identification by functional form. The analogous case is the SUR model which is also identified even if the variables in the two equations are the same.

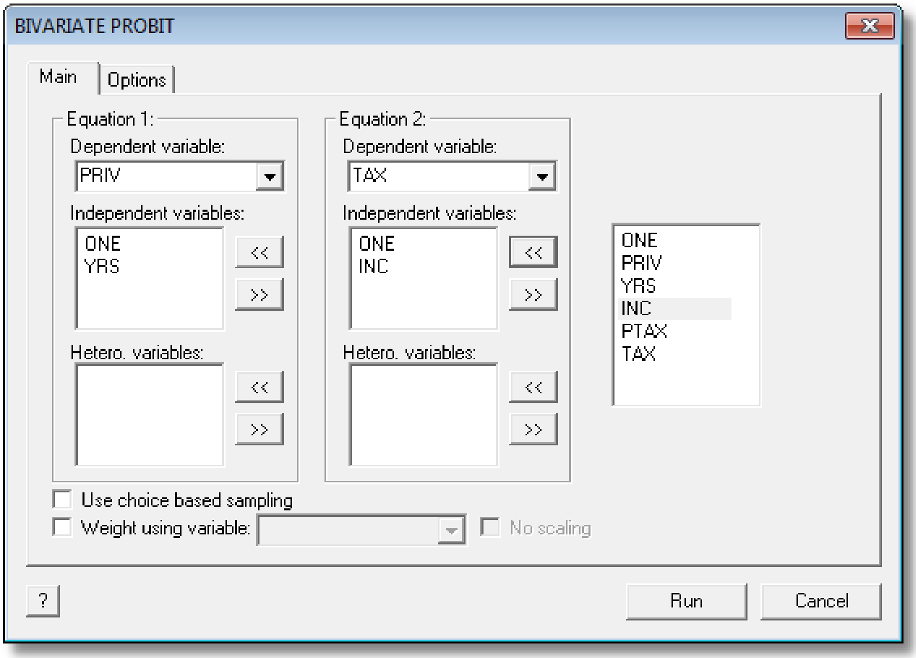
- Some extensions to a simultaneous equations model are easily programmed.
- The bivariate probit and partial observability models are extended to the random parameters modeling framework for panel data.

E33.2 Estimating the Bivariate Probit Model

The two equations can each be estimated consistently by individual single equation probit methods (see Chapter E27). However, this is inefficient and incomplete in that it ignores the correlation between the disturbances. Moreover, the correlation coefficient itself might be of interest. The comparison is analogous to that between OLS and GLS in the multivariate regression model. The model is estimated in *LIMDEP* using full information maximum likelihood. The essential command is

```
BIVARIATE PROBIT      ; Lhs = y1,y2
(or just BIVARIATE)   ; Rh1 = right hand side for equation 1
                      ; Rh2 = right hand side for equation 2 $
```

The command builder for this model is in Model:Binary Choice/Bivariate Probit. The two dependent variables and the right hand sides of the two equations are specified on the Main page. You can also specify a model with heteroscedasticity in either or both equations on this page. The Options page allows you to specify the model above (normal) or the sample selection or partial observability model.



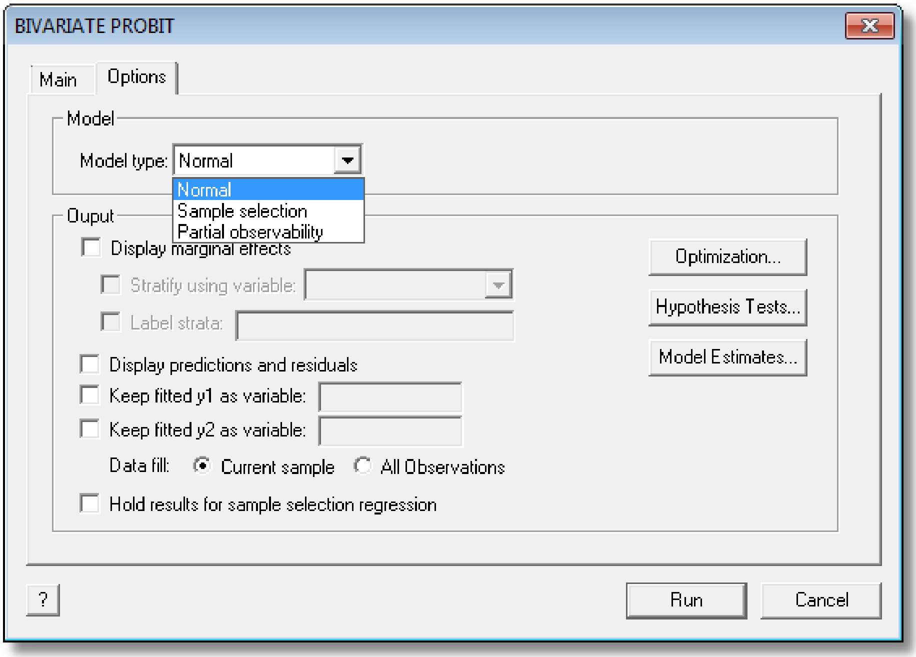


Figure E33.1 Command Builder for Bivariate Probit Models

E33.2.1 Options for the Bivariate Probit Model

Restrictions may be imposed both between and within equations by using

; Rst = list of specifications...

and **; CML: linear restrictions**

You might, for example, force the coefficients in the two equations to be equal as follows:

```
NAMELIST    ; x = ... $
CALC        ; k = Col(x) $
BIVARIATE   ; Lhs = y1,y2 ; Rh1 = x ; Rh2 = x ; Rst = k_b, k_b, r $
```

(The model *is* identified with the same variables in the two equations.)

NOTE: You should not use the name *rho* for ρ in your **; Rst** specification; *rho* is the reserved name for the scalar containing the most recently estimated value of ρ in whatever model estimated it. If it has not been estimated recently, it is zero. Either way, when **; Rst** contains the name *rho*, this is equivalent to *fixing* ρ at the value then contained in the scalar *rho*. That is, *rho* is a value, not a model parameter name such as *b*1. On the contrary, however, you might wish specifically to use *rho* in your **; Rst** specification. For example, to trace the maximized log likelihood over values of ρ, you might base the study on a command set that includes

```
PROCEDURE
BIVARIATE   ; ....  ; Rst = ..., rho $
...
ENDPROC
EXECUTE     ; rho = 0.0, .90, .10 $
```

This would estimate the bivariate probit model 10 times, with ρ fixed at 0, .1, .2, ..., .9. Presumably, as part of the procedure, you would be capturing the values of *logl* and storing them for a later listing or perhaps a plot of the values against the values of *rho*.

If you use the constraints option, the parameter specification includes ρ. As such, you can use this method to fix ρ to a particular value. For another example, consider the application in Section E33.2.8. This is a model for a voting choice and use of private schools:

$$vote = f_1(one, income, property\_taxes)$$
$$private = f_2(one, income, years, teacher).$$

Suppose it were desired to make the income coefficient the same in the two equations and, in a second model, fix *rho* at 0.4. The commands could be

```
BIVARIATE   ; Lhs  = tax,priv
            ; Rh1 = one,inc,ptax ; Rh2 = one,inc,yrs,tch
            ; Rst  = b10,bi,b12,b20,bi,b22,b23,r $
```

and

```
BIVARIATE   ; Lhs  = tax, priv
            ; Rh1 = one,inc,ptax ; Rh2 = one,inc,yrs,tch
            ; Rst  = b10,bi,b12,b20,bi,b22,b23,0.4 $
```

Choice Based Sampling

Any of the bivariate probit models may be estimated with choice based sampling. The feature is requested with

; Wts = the appropriate weighting variable
; Choice Based

For this model, your weighting variable will take four values, for the four cells (0,0), (0,1), (1,0), and (1,1);

w_{ij} = population proportion / sample proportion, i,j = 0,1.

The particular value corresponds to the outcome that actually occurs. You must provide the values. You can obtain sample proportions you need if you do not already have them by computing a crosstab for the two Lhs variables:

CROSSTAB ; Lhs = y1 ; Rhs = y2 $

The table proportions are exactly the proportions you will need. To use this estimator, it is assumed that you know the population proportions.

Robust Covariance Matrix with Correction for Clustering

The standard errors for all bivariate probit models may be corrected for clustering in the sample. Full details on the computation are given in Chapter R10, so we give only the final result here. Assume that the data set is partitioned into G clusters of related observations (like a panel). After estimation, let **V** be the estimated asymptotic covariance matrix which ignores the clustering. Let $\mathbf{g}_{ij}$ denote the first derivatives of the log likelihood with respect to all model parameters for observation (individual) i in cluster j. Then, the corrected asymptotic covariance matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{V}\left(\frac{G}{G-1}\right)\left[\sum_{i=1}^{G}\left(\sum_{j=1}^{n_i}\mathbf{g}_{ij}\right)\left(\sum_{j=1}^{n_i}\mathbf{g}_{ij}\right)'\right]\mathbf{V}$$

You specify the clusters with

; Cluster = either the fixed number of individuals in a group or the name of a variable which identifies the group membership

Any identifier which is common to all members in a cluster and different from other clusters may be used. The controls for stratified and clustered data may be used as well. These are as follows:

; Cluster = the number of observations in a cluster (fixed) or the name of a stratification variable which gives the cluster an identification. This is the setup that is described above.
; Stratum = the number of observations in a stratum (fixed) or the name of a stratification variable which gives the stratum an identification
; Wts = the name of the usual weighting variable for model estimation if weights are desired. This defines w_{ics}. This is the usual weighting setup that has been used in all previous versions of *LIMDEP*.

; FPC = the name of a variable which gives the number of clusters in the stratum. This number will be the same for all observations in a stratum – repeated for all clusters in the stratum. If this number is the same for all strata, then just give the number.
; Huber Use this switch to request h_s. If omitted, $h_s = 1$ is used.
; DFC Use this switch to request the use of d given above. If omitted, $d = 1$ is used.

Note, these corrections will generally lead to larger standard errors compared to the uncorrected results.

Standard Model Specifications for the Bivariate Probit Model

This is the full list of general options that are applicable to this model estimator.

Controlling Output from Model Commands

; OLS reports initial ordinary least squares estimates of parameters
; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Choice uses choice based sampling (sandwich with weighting) estimated matrix.
; Cluster = spec requests computation of the cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm (Newton is not available, avoid BHHH).
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values of y_1 as a new (or replacement) variable in data set.
; Res keeps a fitted value for y_2.
; Fill fills missing values (outside estimating sample) for fitted values.
; Prob keeps a fitted probability for observed joint y_1, y_2 outcome.
; Density keeps a fitted bivariate normal density for observed outcome.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Fitted values for the bivariate probit model are treated a bit differently from other single equation models. Note that the fitted value is the prediction for y_1 while the 'residual' is the prediction for y_2. Since this is a two equation model, there is no residual as such.

E33.2.2 Starting values

You may provide starting values for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. (A starting value for ρ is optional, even if you give the values for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.) Since estimation of this model is a bit more difficult than the univariate probit model – the log likelihood is not globally concave because of ρ – a good set of starting values may be helpful. Since the single equation estimators are consistent, one way to proceed would be

```
NAMELIST    ; x1 = the Rhs variables in equation 1
            ; x2 = the Rhs variables in equation 2 $
PROBIT      ; Lhs = y1 ; Rhs = x1 ... ; ... any other options $
MATRIX      ; b1 = b $
PROBIT      ; Lhs = y2 ; Rhs  = x2 ... ; ... any other options $
MATRIX      ; b2 = b $
BIVARIATE   ; Lhs = y1,y2 ; Rh1 = x1 ; Rh2 = x2 ; Start = b1,b2 $
```

where *y*1 and *y*2 are the two binary variables, *x*1 and *x*2 are lists of variable names for the two regressor vectors, and *b*1 and *b*2 are the two column vectors of starting values in your matrix work area. There must be an exact correspondence between the values in *b*1 and *b*2 and the variables in *x*1 and *x*2.

If a starting value for ρ is present, it must be last in your list of starting values. If you do not provide starting values for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the OLS results of regressing *y*1 on *x*1 and *y*2 on *x*2 are used. The starting value for ρ is obtained as follows: The conditional mean function $E[z_{i1}|y_{i2}, \mathbf{x}_{i2}]$ is

$$E[z_{i1}|y_{i2}, \mathbf{x}_{i2}] = \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \rho\lambda_{i2}$$

where $$\lambda_{i2} = (2y_{i2}-1)\phi(\boldsymbol{\beta}_1'\mathbf{x}_{i2}) / \Phi[(2y_{i2}-1)\boldsymbol{\beta}_1'\mathbf{x}_{i2}].$$

Thus, if z_{i1} and $\boldsymbol{\beta}_1$ were observed, ρ could be estimated by a linear regression of z_{i1} on $\mathbf{x}_{i2}$ and λ_{i2}. In order to approximate this result, we use y_{i1} for z_{i1} and the starting values for the parameters in this regression. The resulting estimator is inconsistent, but generally closer to the final result than the obvious alternative, zero.

E33.2.3 Proportions Data

Like other discrete choice models, this one may be fit with proportions data. Since this is a bivariate model, you must provide the full set of four proportions variables, in the order

; Lhs = p00, p01, p10, p11.

(You may use your own names). Proportions must be strictly between zero and one, and the four variables must add to 1.0.

NOTE: When you fit the model using proportions data, there is no cross tabulation of fitted and actual values produced, and no fitted values or 'residuals' are computed.

E33.2.4 Heteroscedasticity

All bivariate probit specifications, including the basic two equation model, the sample selection model (Section E33.4), and the Meng and Schmidt partial observability model (Section E33.7), may be fit with a multiplicative heteroscedasticity specification. The model is the same as the univariate probit model (Section E27.11);

$$\varepsilon_i \sim N\{0, [\exp(\boldsymbol{\gamma}_i'\mathbf{z}_i)]^2\},\ i = 1 \text{ and/or } 2.$$

Either or both equations may be specified in this fashion. Use

; Hf1 = list of variables if you wish to modify the first equation
; Hf2 = list of variables if you wish to modify the second equation

NOTE: Do not include *one* in either list. The model will become inestimable.

The model is unchanged otherwise, and the full set of options given earlier remains available. To give starting values with this modification, supply the following values in the order given:

$$\boldsymbol{\Theta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \rho].$$

As before, all starting values are optional, and if you do provide the slopes, the starting value for ρ is still optional. The internal starting values for the variance parameters are zero for both equations. (This produces the original homoscedastic model.)

E33.2.5 Specification Tests

Wald, LM, and LR tests related to the slope parameters would follow the usual patterns discussed in previous chapters. One might be interested in testing hypotheses about the correlation coefficient. The Wald test for the hypothesis that ρ equals zero is part of the standard output for the model – see the results below which include a 't' statistic for this hypothesis. Likelihood ratio and LM tests can be carried out as shown below:

The following routine will test the specification of the bivariate probit model against the null hypothesis that two separate univariate probits apply. The test of the hypothesis that ρ equals zero is sufficient for this. The first group of commands computes and saves the univariate probit coefficients and log likelihoods.

```
NAMELIST    ; x1 = ... Rhs for the first equation
            ; x2 = ... Rhs for the second equation $
PROBIT      ; Lhs = y1 ; Rhs = x1 $
MATRIX      ; b1 = b $
CALC        ; l1 = logl $
PROBIT      ; Lhs = y2 ; Rhs = x2 $
MATRIX      ; b2 = b $
CALC        ; l2 = logl $
```

To carry out the likelihood ratio test, we now fit the bivariate model, which is the unrestricted one. The restricted model, with ρ = 0, is the two univariate models. The restricted log likelihood is the sum of the two univariate values. The **CALC** command carries out the test. The **BIVARIATE** command also produces a t statistic in the displayed output for the hypothesis that ρ = 0. To automate the test, we can also use the automatically retained values *rho* and *varrho*. The second **CALC** command carries out this test.

```
BIVARIATE   ; Lhs  = y1,y2 ; Rh1 = x1 ; Rh2 = x2 $
CALC        ; lrtest = 2*(l1 + l2 - logl)
            ; pvalue = 1 - Chi(lrtest,1) $
CALC        ; waldtest = rho^2 / varrho
            ; pvalue = 1 - Chi(waldtest,1) $
```

The Lagrange multiplier test is also simple to carry out using the built in procedure, as we have already estimated the restricted model. The test is carried out with the model command that specifies the starting values from the restricted model and restricts the maximum iterations to zero.

```
NAMELIST    ; x1 = ... Rhs for the first equation
            ; x2 = ... Rhs for the second equation $
PROBIT      ; Lhs = y1 ; Rhs = x1 $
MATRIX      ; b1 = b $
PROBIT      ; Lhs = y2 ; Rhs = x2 $
MATRIX      ; b2 = b $
BIVARIATE   ; Lhs = y1,y2 ; Rh1 = x1 ; Rh2 = x2
            ; Start = b1,b2,0 ; Maxit = 0 $
```

You can test the heteroscedasticity assumption by any of the three classical tests as well. The LM test will be the simplest since it does not require estimation of the model with heteroscedasticity. You can carry out the LM test as follows:

```
NAMELIST    ; x1 = ... ; x2  = ... ; z1  = ... ; z2  = ... $
BIVARIATE   ; Lhs = ...  ; Rh1 = x1 ; Rh2 = x2 $
CALC        ; h1 = Col(z1) ; h2 = Col(z2)
            ; k1 = Col(x1) ; k2 = Col(x2) ; k12 = k1+k2 $
MATRIX      ; b1_b2 = b(1:k12) $
BIVARIATE   ; Lhs  = ...
            ; Rh1 = x1 ; Rh2 = x2 ? specify the two probit equations
            ; Hf1  = z1 ; Hf2 = z2  ? variables in the two variances
            ; Start = b1_b2, h1_0, h2_0, rho
            ; Maxit = 0 $
```

In this instance, the starting value for *rho* is the value that was estimated by the first model, which is retained as a scalar value.

E33.2.6 Model Results for the Bivariate Probit Model

The initial output for the bivariate probit models consists of the ordinary least squares results if you request them with

; OLS

Final output includes the log likelihood value and the usual statistical results for the parameter estimates.

The last output, requested with

; Summary

is a joint frequency table for four cells, with actual and predicted values shown. The predicted outcome is the cell with the largest probability. Cell probabilities are computed using

$$P_{i00} = 1 - P_{i11} - P_{i10} - P_{i01} \qquad P_{i01} = \Phi\ [\boldsymbol{\beta}_2'\mathbf{x}_{i2}] - P_{i11}$$

$$P_{i10} = \Phi\ [\boldsymbol{\beta}_1'\mathbf{x}_{i1}] - P_{i11} \qquad P_{i11} = \Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_{i1}, \boldsymbol{\beta}_2'\mathbf{x}_{i1}, \rho]$$

A table which assesses the success of the model in predicting the two variables is presented as well. An example appears below. The predictions and residuals are a bit different from the usual setup (because this is a two equation model):

; Keep = name to retain the predicted y_1
; Res = name to retain the predicted y_2
; Prob = name to retain the probability for observed y_1, y_2 outcome
; Density = fitted bivariate normal density for observed outcome

Matrix results kept in the work areas automatically are *b* and *varb*. An extra matrix named *b_bprobt* is also created. This is a two column matrix that collects the coefficients in the two equations in a parameter matrix. The number of rows is the larger of the number of variables in *x*1 and *x*2. The coefficients are placed at the tops of the respective columns with the shorter column padded with zeros.

NOTE: There is no correspondence between the coefficients in any particular row of *b_bprobt*. For example, in the second row, the coefficient in the first column is that on the second variable in *x*1 and the coefficient in the second column is that on the second variable in *x*2. These may or may not be the same.

The results saved by the binary choice models are:

Matrices: *b* = estimate of $(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2',\rho)'$
varb = asymptotic covariance matrix

Scalars: *kreg* = number of parameters in model
nreg = number of observations
logl = log likelihood function

Variables: *logl_obs* = individual contribution to log likelihood

Last Model: *b1_variables*, *b2_variables*, *c1_variables*, *c1_variables*, *r12*

Last Function: $\text{Prob}(y_1 = 1, y_2 = 1|\mathbf{x}_1, x_2) = \Phi_2(\mathbf{b}_1'\mathbf{x}_1, \mathbf{b}_2'\mathbf{x}_2, r)$

The saved scalars are *nreg*, *kreg*, *logl*, *rho*, *varrho*. The *Last Model* labels are *b_variables* and *b*2*_variables*. If the heteroscedasticity specification is used, the additional coefficients are *c*1*_variables* and *c*2*_variables*. To extract a vector that contains only the slopes, and not the correlation, use

```
MATRIX      ; {kb = kreg-1} ; b1b2 = b(1:kb) $
```

To extract the two parameter vectors separately, after defining the namelists, you can use

```
MATRIX      ; {k1 = Col(x1) ; k12 = k1+1 ; kb = kreg-1}
            ; b1 = b(1:k1) ; b2 = b(k12:kb) $
```

You may use other names for the matrices. (Note that the **MATRIX** commands contain embedded **CALC** commands contained in {}.) If the model specifies heteroscedasticity, similar constructions can be used to extract the three or four parts of *b*.

E33.2.7 Partial Effects

Because it is a two equation model, it is unclear what should be an appropriate 'marginal effect' in the bivariate probit model. (This is one of our frequently asked questions, as users are often uncertain about what it is that they are looking for when they seek the 'partial effects' in the model – effect of what? on what?) The literature is not necessarily helpful in this regard. The one published result in the econometrics literature, Christofides, Stengos and Swidinsky (1997), plus an error correction in a later issue, focuses on the joint probability of the two outcome variables equaling one – which is not a conditional mean. The probability might be of interest. It can be examined with the **PARTIAL EFFECTS** program. An example appears in Section E33.2.8. The *marginal* means in the model are the univariate probabilities that the two variables equal one. These are also not necessarily interesting, but, in any event, they can be computed using the univariate models.

LIMDEP analyzes the conditional mean function

$$E[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2] = \text{Prob}[y_1 = 1, y_2 = 1 \mid \mathbf{x}_1, \mathbf{x}_2, \rho) \,/\, \text{Prob}[y_2 = 1 \mid \mathbf{x}_1].$$

This is the function analyzed in the bivariate probit marginal effects processor. The bivariate probit estimator in *LIMDEP* allows either or both of the latent regressions to be heteroscedastic. The reported effects for this model include the decomposition of the marginal effect into all four terms, the regression part and the variance part, in each of the two latent models.

The computations of the following marginal effects in the bivariate probit model are included as an option with the estimator. There are two models, the base case of y_1,y_2 a pair of correlated probit models, and $y_1|y_2 = 1$, the bivariate probit with sample selection. (See Section E33.4 below.) The conditional mean computed for these two models would be identical,

$$\text{E}[y_1|y_2 = 1] = \Phi_2\,[w_1, w_2\,, \rho\,] \,/\, \Phi(\,w_2\,)$$

where Φ_2 is the bivariate normal CDF and Φ is the univariate normal CDF. This model allows multiplicative heteroscedasticity in either or both equations, so

$$w_1 = \boldsymbol{\beta}_1'\mathbf{x}_1 \,/\, \exp(\boldsymbol{\gamma}_1'\mathbf{z}_1)$$

and likewise for w_2. In the homoscedastic model, $\boldsymbol{\gamma}_1$ and/or $\boldsymbol{\gamma}_2$ is a zero vector. Four full sets of marginal effects are reported, for $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{z}_1$, and $\mathbf{z}_2$. Note that the last two may be zero. The four vectors may also have variables in common. For any variable which appears in more than one of the parts, the marginal effect is the sum of the individual terms. A table is reported which displays these total effects for every variable which appears in the model, along with estimated standard errors and the usual statistical output. Formulas for the parts of these marginal effects are given below with the technical details. For further details, see Greene (2011). Commands and suggestions for how to do these computations are given in Section E33.2.8.

Note that you can get marginal effects for $y_2|y_1$ just by respecifying the model with y_1 and y_2 reversed (y_2 now appears first) in the Lhs list of the command. You can also trick *LIMDEP* into giving you marginal effects for $y_1|y_2 = 0$ (instead of $y_2 = 1$) by computing $z_1 = 1\text{-}y_1$ and $z_2 = 1\text{-}y_2$, and fitting the same bivariate probit model but with Lhs = z_1,z_2. *You must now reverse the signs of the marginal effects (and all slope coefficients) that are reported.*

The example below was produced by a sampling experiment: Note that the model specifies heteroscedasticity in the second equation though, in fact, there is none.

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-500 $
CREATE      ; u1 = Rnn(0,1)  ; u2 = u1 + Rnn(0,1)
            ; z  = Rnu(.2,.4) ; x1 = Rnn(0,1) ; x2 = Rnn(0,1)
            ; x3 = Rnn(0,1)  ; y1 = (x1 + x2 + u1) > 0 ; y2 = (x1 + x3 + u2) > 0 $
BIVARIATE   ; Lhs  = y1,y2
            ; Rh1 = one,x1,x2 ; Rh2 = one,x1,x3
            ; Hf2  = z ; Partial Effects $
```

The first set of results is the model coefficients.

```
-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable                  Y1Y2
Log likelihood function       -416.31350
Estimation based on N =    500, K =   8
Inf.Cr.AIC  =  848.627 AIC/N =    1.697
Disturbance model is multiplicative het.
Var. Parms follow   6 slope estimates.
For e(2),  1 estimates follow X3
--------+--------------------------------------------------------------------
      Y1|                  Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for Y1
Constant|    -.04292         .07362     -.58  .5599     -.18721    .10137
      X1|    1.09235***      .08571    12.74  .0000      .92435   1.26035
      X2|    1.06802***      .08946    11.94  .0000      .89268   1.24337
        |Index    equation for Y2
Constant|     .01017         .06432      .16  .8744     -.11590    .13623
      X1|     .82908**       .37815     2.19  .0283      .08792   1.57024
      X3|     .70123**       .30512     2.30  .0215      .10321   1.29925
        |Variance equation for Y2
       Z|    -.05575        1.45449     -.04  .9694    -2.90651   2.79500
        |Disturbance correlation
RHO(1,2)|     .66721***      .07731     8.63  .0000      .51568    .81874
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the decomposition of the marginal effects for the four possible contributors to the effect.

```
+--------------------------------------------------------+
|              Partial Effects for Ey1|y2=1              |
+----------+---------------------+---------------------+
|          | Regression Function |  Heteroscedasticity |
|          +---------------------+---------------------+
|          |   Direct | Indirect |   Direct | Indirect |
| Variable | Efct  x1 | Efct  x2 | Efct  h1 | Efct  h2 |
+----------+----------+----------+----------+----------+
|       X1 |   .48383 |  -.17370 |   .00000 |   .00000 |
|       X2 |   .47305 |   .00000 |   .00000 |   .00000 |
|       X3 |   .00000 |  -.14691 |   .00000 |   .00000 |
|        Z |   .00000 |   .00000 |   .00000 |   .00092 |
+----------+----------+----------+----------+----------+
```

A table of the specific effects is produced for each contributor to the marginal effects. This first table gives the total effects. The values here are the row total in the table above.

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .661053
Observations used for means are  All Obs.
Total effects reported = direct+indirect.
--------+--------------------------------------------------------------------
      Y1|     Partial      Standard            Prob.      95% Confidence
      Y2|      Effect         Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
      X1|     .31013***      .04356     7.12  .0000      .22476    .39550
      X2|     .47305***      .04338    10.91  .0000      .38804    .55807
      X3|    -.14691***      .02853    -5.15  .0000     -.20283   -.09099
       Z|     .00092         .02404      .04  .9694     -.04620    .04804
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The direct effects are the marginal effects of the variables ($\mathbf{x}_1$ and $\mathbf{z}_1$) that appear in the first equation.

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .435447
Observations used for means are  All Obs.
These  are the  direct marginal  effects.
--------+--------------------------------------------------------------------
     TAX|     Partial      Standard            Prob.      95% Confidence
    PRIV|      Effect         Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
     INC|     .67814***      .24487     2.77  .0056      .19820   1.15807
    PTAX|    -.83030**       .38146    -2.18  .0295    -1.57794   -.08266
     YRS|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
------------------------------------------------------------------------------
```

The indirect effects are the effects of the variables that appear in the other (second) equation.

```
------------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .661053
Observations used for means are  All Obs.
These are the indirect  marginal effects.
--------+---------------------------------------------------------------------
      Y1|     Partial      Standard            Prob.      95% Confidence
E[y1|x,z|      Effect        Error       z    |z|>Z*         Interval
--------+---------------------------------------------------------------------
      X1|    -.17370***      .03250    -5.34  .0000     -.23740   -.11000
      X2|        0.0    .....(Fixed Parameter).....
      X3|    -.14691***      .02853    -5.15  .0000     -.20283   -.09099
       Z|     .00092         .02404      .04  .9694     -.04620    .04804
--------+---------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
------------------------------------------------------------------------------
```

The marginal effects processor in the bivariate probit model detects when a regressor is a dummy variable. In this case, the marginal effect is computed using differences, not derivatives. The model results will contain a specific description. To illustrate this computation, we revisit the German health care data. A description appears in Chapter E2. Here, we analyze the two health care utilization variables, *doctor* = 1(*docvis* > 0) and *hospital* = 1(*hospvis* > 0) in a bivariate probit model. The model command is

SAMPLE ; All $
CREATE ; doctor = docvis > 0 ; hospital = hospvis > 0 $
BIVARIATE ; Lhs = doctor,hospital
; Rh1 = one,age,educ,hhninc,hhkids
; Rh2 = one,age,hhninc,hhkids
; Partial Effects $

The variable *hhkids* is a binary variable for whether there are children in the household. The estimation results are as follows. This is similar to the preceding example. The final table contains the result for the binary variable. In fact, the explicit treatment of the binary variable results in very little change in the estimate.

```
-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable              DOCHOS
Log likelihood function    -25552.65886
Estimation based on N =  27326, K =  10
Inf.Cr.AIC  =51125.318 AIC/N =    1.871
--------+--------------------------------------------------------------------
  DOCTOR|                 Standard            Prob.      95% Confidence
HOSPITAL|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for DOCTOR
Constant|     .13653**       .05618     2.43  .0151      .02642    .24663
     AGE|     .01353***      .00076    17.84  .0000      .01205    .01502
    EDUC|    -.02675***      .00345    -7.75  .0000     -.03352   -.01998
  HHNINC|    -.10245**       .04541    -2.26  .0241     -.19144   -.01345
  HHKIDS|    -.12299***      .01670    -7.37  .0000     -.15571   -.09027
        |Index    equation for HOSPITAL
Constant|   -1.54988***      .05325   -29.10  .0000    -1.65426  -1.44551
     AGE|     .00510***      .00100     5.08  .0000      .00313    .00707
  HHNINC|    -.05514         .05510    -1.00  .3169     -.16314    .05285
  HHKIDS|    -.02682         .02392    -1.12  .2622     -.07371    .02006
        |Disturbance correlation
RHO(1,2)|     .30251***      .01381    21.91  .0000      .27545    .32958
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+----------------------------------+
|  Partial Effects for Ey1|y2=1  |
+----------+----------+----------+
|          |   Direct | Indirect |
| Variable | Efct  x1 | Efct  x2 |
+----------+----------+----------+
|      AGE |   .00367 |  -.00036 |
|     EDUC |  -.00726 |   .00000 |
|   HHNINC |  -.02779 |   .00385 |
|   HHKIDS |  -.03336 |   .00187 |
+----------+----------+----------+

-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .822131
Observations used for means are  All Obs.
Total effects reported = direct+indirect.
--------+--------------------------------------------------------------------
  DOCTOR|     Partial     Standard            Prob.      95% Confidence
HOSPITAL|      Effect       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .00332***      .00023    14.39  .0000      .00286    .00377
    EDUC|    -.00726***      .00096    -7.58  .0000     -.00913   -.00538
  HHNINC|    -.02394*        .01225    -1.95  .0507     -.04796    .00008
  HHKIDS|    -.03149***      .00471    -6.69  .0000     -.04072   -.02226
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .822131
Observations used for means are  All Obs.
These  are the  direct marginal  effects.
--------+--------------------------------------------------------------------
  DOCTOR|     Partial      Standard            Prob.      95% Confidence
HOSPITAL|      Effect         Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|     .00367***      .00022    16.44  .0000       .00323    .00411
    EDUC|    -.00726***      .00096    -7.58  .0000     -.00913   -.00538
  HHNINC|    -.02779**       .01232    -2.25  .0241     -.05195   -.00364
  HHKIDS|    -.03336***      .00460    -7.26  .0000     -.04237   -.02436
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .822131
Observations used for means are  All Obs.
These are the indirect  marginal effects.
--------+--------------------------------------------------------------------
  DOCTOR|     Partial      Standard            Prob.      95% Confidence
E[y1|x,z|      Effect         Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|    -.00036***   .7075D-04    -5.03  .0000     -.00049   -.00022
    EDUC|        0.0    .....(Fixed Parameter).....
  HHNINC|     .00385         .00385     1.00  .3167     -.00369    .01140
  HHKIDS|     .00187         .00167     1.12  .2620     -.00140    .00515
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

```
+------------------------------------------------------------+
| Analysis of dummy variables in the model. The effects are |
| computed using E[y1|y2=1,d=1] - E[y1|y2=1,d=0] where d is |
| the variable. Variances use the delta method.  The effect |
| accounts for all appearances of the variable in the model.|
+------------------------------------------------------------+
|Variable      Effect   Standard error     t ratio          |
+------------------------------------------------------------+
 HHKIDS      -.031829     .004804           -6.625
```

E33.2.8 Application

The following are a subset of the variables and observations of a data set given by Pindyck and Rubinfeld (1991). The variables in the data set are:

priv = decision whether to have at least one child in private school.
yrs = years lived in the community.
inc = log of income. Read in as a code, then recoded.
ptax = log of property taxes paid. Read in as a code, then recoded.
tax = vote (0=no) on a property tax.

The data were entered and transformed as follows:

READ ; Nvar = 5 ; Nobs = 80 ; By Variables
; Names = priv,yrs,inc,ptax,tax $

```
0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0
0 0 0 0 0 0
42 5 10 4 4 11 5 35 3 16 7 5 11 3 2 2 2 2 4 2 3 3 2 10 2 2 3 3 3 6 2 26 18
4 6 12 49 6 18 5 6 20 1 3 5 2 5 18 20 14 3 17 20 3 2 5 35 10 8 12 7 3 25 5
4 2 5 3 2 6 3 12 3 3 3 3 3 5 35 3
4 6 5 6 6 7 6 4 7 5 7 4 4 4 6 5 3 1 7 5 6 6 6 5 6 5 8 5 5 5 5 4 6 4 5 5 3
7 4 5 4 3 4 5 7 5 8 3 4 2 3 3 5 5 5 6 4 5 4 4 6 7 6 4 6 5 7 4 8 2 4 3 4 5
5 6 5 5 2 7
3 4 4 4 3 4 4 3 5 3 4 3 4 3 4 3 3 4 4 4 5 4 6 4 4 4 6 4 4 3 4 3 6 3 4 3 3
5 4 4 1 4 2 3 4 4 5 3 1 2 6 3 4 4 4 4 4 5 4 4 3 3 3 3 4 5 3 4 6 1 4 2 3 4
3 5 4 4 1 6
1 1 0 1 1 1 1 1 1 1 0 1 0 0 1 1 0 0 1 1 0 1 1 0 1 0 0 1 1 1 1 0 0 0 0 1 1
1 0 1 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 1 1 0 1 0 1 1 0 1 1 1 1 0 1 0 1 1 1 1
1 1 1 1 1 0
```

RECODE ; inc
; 1 = 8.294 ; 2 = 8.9227 ; 3 = 9.4335 ; 4 = 9.77
; 5 = 10.021 ; 6 = 10.222 ; 7 = 10.463 ; 8 = 10.820 $
RECODE ; ptax
; 1 = 5.9915 ; 2 = 6.3969 ; 3 = 6.7452 ; 4 = 7.0475
; 5 = 7.2793 ; 6 = 7.4955 $

A bivariate probit model using these data was estimated by Greene (1984). The following is a version of that application. We fit the model

$$vote = f_1(income, property\ taxes),$$
$$private = f_2(income, years, property\ taxes).$$

In the first model, the coefficients are unrestricted. In the second, the income coefficients in the two equations are forced to be equal.

NAMELIST ; y = tax,priv
; x1 = one,inc,ptax ; x2 = one,inc,yrs,ptax $
BIVARIATE ; Lhs = y ; Rh1 = x1 ; Rh2 = x2 ; OLS ; Summary
; Partial Effects $

```
-----------------------------------------------------------------------------
OLS Starting Estimates for Bivariate Probit
--------+--------------------------------------------------------------------
     TAX|                  Standard            Prob.      95% Confidence
    PRIV|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.28967        1.30849     -.22  .8248    -2.85426   2.27492
     INC|     .47407***      .14186     3.34  .0008      .19602    .75211
    PTAX|    -.54751***      .18403    -2.98  .0029     -.90819   -.18682
Constant|    -.56240        1.06063     -.53  .5959    -2.64120   1.51639
     INC|     .08378         .10800      .78  .4379     -.12789    .29546
     YRS|    -.00129         .00411     -.31  .7532     -.00935    .00677
    PTAX|    -.01966         .13886     -.14  .8874     -.29181    .25249
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               TAXPRI
Log likelihood function       -74.32308
Estimation based on N =     80, K =   8
Inf.Cr.AIC  =  164.646 AIC/N =    2.058
Model estimated: Jun 16, 2011, 07:58:43
--------+--------------------------------------------------------------------
     TAX|                  Standard            Prob.      95% Confidence
    PRIV|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for TAX
Constant|   -2.04203        4.36803     -.47  .6401   -10.60321   6.51915
     INC|    1.64616***      .59722     2.76  .0058      .47563   2.81670
    PTAX|   -2.01554**       .91996    -2.19  .0285    -3.81863   -.21245
        |Index    equation for PRIV
Constant|   -3.86036        4.41911     -.87  .3824   -12.52166   4.80094
     INC|     .35280         .80209      .44  .6600    -1.21926   1.92486
     YRS|    -.01622         .04913     -.33  .7413     -.11252    .08008
    PTAX|    -.09948        1.04226     -.10  .9240    -2.14227   1.94331
        |Disturbance correlation
RHO(1,2)|    -.32379         .29914    -1.08  .2791     -.91009    .26251
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+-------------------------------------------------------+
| Joint Frequency Table for Bivariate Probit Model      |
| Predicted cell is the one with highest probability    |
+-------------------------------------------------------+
|                          PRIV                         |
+-------------+-----------------------------------------+
|  TAX        |       0             1          Total    |
|-------------+-------------+------------+------------+
|        0    |         24  |         5  |        29  |
|    Fitted   |   (     19) |  (      0) |  (     19) |
|-------------+-------------+------------+------------+
|        1    |         46  |         5  |        51  |
|    Fitted   |   (     61) |  (      0) |  (     61) |
|-------------+-------------+------------+------------+
|     Total   |         70  |        10  |        80  |
|    Fitted   |   (     80) |  (      0) |  (     80) |
|-------------+-------------+------------+------------+
```

```
+---------------------------------------------------------+
| Bivariate Probit Predictions for TAX      and PRIV      |
| Predicted cell (i,j) is cell with largest probability   |
| Neither TAX      nor PRIV     predicted correctly       |
|                             4 of      80 observations   |
| Only    TAX      correctly predicted                    |
|         TAX      = 0:       2 of      29 observations   |
|         TAX      = 1:       4 of      51 observations   |
| Only    PRIV     correctly predicted                    |
|         PRIV     = 0:      18 of      70 observations   |
|         PRIV     = 1:       4 of      10 observations   |
| Both    TAX      and PRIV     correctly predicted       |
|         TAX      = 0 PRIV     = 0:       11 of      24  |
|         TAX      = 1 PRIV     = 0:       41 of      46  |
|         TAX      = 0 PRIV     = 1:        0 of       5  |
|         TAX      = 1 PRIV     = 1:        0 of       5  |
+---------------------------------------------------------+
```

The partial effects are as follows:

```
+--------------------------------+
|  Partial Effects for Ey1|y2=1  |
+----------+----------+----------+
|          |   Direct | Indirect |
| Variable | Efct  x1 | Efct  x2 |
+----------+----------+----------+
|      INC |   .67814 |   .03880 |
|     PTAX |  -.83030 |  -.01094 |
|      YRS |   .00000 |  -.00178 |
+----------+----------+----------+
```

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .435447
Observations used for means are  All Obs.
Total effects reported = direct+indirect.
--------+--------------------------------------------------------------------
     TAX|     Partial      Standard            Prob.      95% Confidence
    PRIV|      Effect         Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     INC|     .71693***      .24000     2.99  .0028      .24654   1.18732
    PTAX|    -.84124**       .37127    -2.27  .0235    -1.56892   -.11356
     YRS|    -.00178         .00669     -.27  .7897     -.01489    .01132
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .435447
Observations used for means are  All Obs.
These  are the  direct marginal  effects.
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
     TAX|     Partial      Standard            Prob.      95% Confidence
    PRIV|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     INC|     .67814***      .24487     2.77  .0056      .19820   1.15807
    PTAX|    -.83030**       .38146    -2.18  .0295    -1.57794   -.08266
     YRS|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .435447
Observations used for means are  All Obs.
These are the indirect  marginal effects.
--------+--------------------------------------------------------------------
     TAX|     Partial      Standard            Prob.      95% Confidence
E[y1|x,z|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     INC|     .03880         .07495      .52  .6047     -.10810     .18570
    PTAX|    -.01094         .10889     -.10  .9200     -.22435     .20247
     YRS|    -.00178         .00669     -.27  .7897     -.01489     .01132
--------+--------------------------------------------------------------------
```

The preceding examined the conditional mean function, $\Phi_2(\mathbf{b}_1'\mathbf{x}_1,\mathbf{b}_2'\mathbf{x}_2,\rho)/\Phi(\mathbf{b}_2'\mathbf{x}_2)$. The **PARTIAL EFFECTS** (or just **PARTIALS**) command will produce effects for the joint probability instead. The default computation is the average partial effect. The following shows the computation at the data means for comparison.

PARTIALS ; Effects: ptax $
PARTIALS ; Effects: ptax ; Means $

```
---------------------------------------------------------------------
Partial Effects  Analysis for Bivariate Probit Prob. Function
---------------------------------------------------------------------
Effects on function with respect to PTAX
Results are computed by average over sample observations
Partial effects for continuous PTAX     computed by differentiation
Effect is computed as derivative     = df(.)/dx
---------------------------------------------------------------------
df/dPTAX           Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
APE. Function      -.10264     .11275     .91     -.32363      .11834
--> partials ; effects: ptax; means $
---------------------------------------------------------------------
```

```
----------------------------------------------------------------------
Partial Effects  Analysis for Bivariate Probit Prob. Function
----------------------------------------------------------------------
Effects on function with respect to PTAX
Results are computed at sample means of all variables
Partial effects for continuous PTAX     computed by differentiation
Effect is computed as derivative     = df(.)/dx
----------------------------------------------------------------------
df/dPTAX             Partial    Standard
(Delta method)       Effect      Error     |t|  95% Confidence Interval
----------------------------------------------------------------------
PE.Func(means)      -.10940     .11724     .93     -.33918      .12039
```

The function in the **PARTIALS** command can be changed, for example, to analyze the same conditional mean function as above. We would do this as follows:

```
PARTIALS    ; Effects: ptax
            ; Function = bx1 = b1+b2*inc+b3*ptax |
                         bx2 = c1+c2*inc+c3*yrs+c4*ptax |
                         Bvn(bx1,bx2,ro)/Phi(bx2)
            ; Labels = b1,b2,b3,c1,c2,c3,c4,ro  $
```

Adding

```
            ; Means $
```

would then replicate the computations done with the **; Partial Effects** specification in the model command. The -.84124 and standard error of .37127 appear in the earlier table of total effects for the bivariate probit model.

```
----------------------------------------------------------------------
Partial Effects  Analysis for User Specified Function
----------------------------------------------------------------------
Effects on function with respect to PTAX
Results are computed by average over sample observations
Partial effects for continuous PTAX     computed by differentiation
Effect is computed as derivative     = df(.)/dx
----------------------------------------------------------------------
df/dPTAX             Partial    Standard
(Delta method)       Effect      Error     |t|  95% Confidence Interval
----------------------------------------------------------------------
APE. Function       -.72099     .26544    2.72    -1.24124     -.20074

----------------------------------------------------------------------
Partial Effects  Analysis for User Specified Function
----------------------------------------------------------------------
Effects on function with respect to PTAX
Results are computed at sample means of all variables
Partial effects for continuous PTAX     computed by differentiation
Effect is computed as derivative     = df(.)/dx
----------------------------------------------------------------------
df/dPTAX             Partial    Standard
(Delta method)       Effect      Error     |t|  95% Confidence Interval
----------------------------------------------------------------------
PE.Func(means)      -.84124     .37127    2.27    -1.56892     -.11356
```

The advantage of the latter computations is that the partial effect can be computed for a variety of values of the variable of interest. For example,

```
PARTIALS    ; Effects: ptax & ptax = 6(.1)8 ; Plot(ci)
            ; Function = bx1 = b1+b2*inc+b3*ptax |
                         bx2 = c1+c2*inc+c3*yrs+c4*ptax |
                         Bvn(bx1,bx2,ro)/Phi(bx2)
            ; Labels = b1,b2,b3,c1,c2,c3,c4,ro $
```

```
----------------------------------------------------------------------
Partial Effects  Analysis for User Specified Function
----------------------------------------------------------------------
Effects on function with respect to PTAX
Results are computed by average over sample observations
Partial effects for continuous PTAX     computed by differentiation
Effect is computed as derivative     = df(.)/dx
----------------------------------------------------------------------
df/dPTAX           Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
----------------------------------------------------------------------
APE. Function      -.72099     .26544    2.72    -1.24124     -.20074
PTAX    =  6.00    -.20817     .15743    1.32     -.51673      .10040
PTAX    =  6.10    -.26309     .16808    1.57     -.59252      .06634
PTAX    =  6.20    -.32621     .17380    1.88     -.66686      .01444
PTAX    =  6.30    -.39555     .17715    2.23     -.74277     -.04834
PTAX    =  6.40    -.46780     .18345    2.55     -.82737     -.10823
PTAX    =  6.50    -.53839     .19785    2.72     -.92619     -.15060
PTAX    =  6.60    -.60194     .22038    2.73    -1.03389     -.16998
PTAX    =  6.70    -.65286     .24492    2.67    -1.13290     -.17281
PTAX    =  6.80    -.68619     .26308    2.61    -1.20182     -.17056
PTAX    =  6.90    -.69840     .26843    2.60    -1.22453     -.17227
PTAX    =  7.00    -.68796     .25902    2.66    -1.19564     -.18029
PTAX    =  7.10    -.65567     .23829    2.75    -1.12272     -.18863
PTAX    =  7.20    -.60451     .21462    2.82    -1.02517     -.18384
PTAX    =  7.30    -.53914     .19825    2.72     -.92772     -.15056
PTAX    =  7.40    -.46519     .19459    2.39     -.84658     -.08379
PTAX    =  7.50    -.38838     .19962    1.95     -.77964      .00287
PTAX    =  7.60    -.31385     .20452    1.53     -.71471      .08701
PTAX    =  7.70    -.24554     .20263    1.21     -.64268      .15161
PTAX    =  7.80    -.18603     .19158     .97     -.56152      .18946
PTAX    =  7.90    -.13653     .17234     .79     -.47432      .20126
PTAX    =  8.00    -.09708     .14767     .66     -.38652      .19236
```

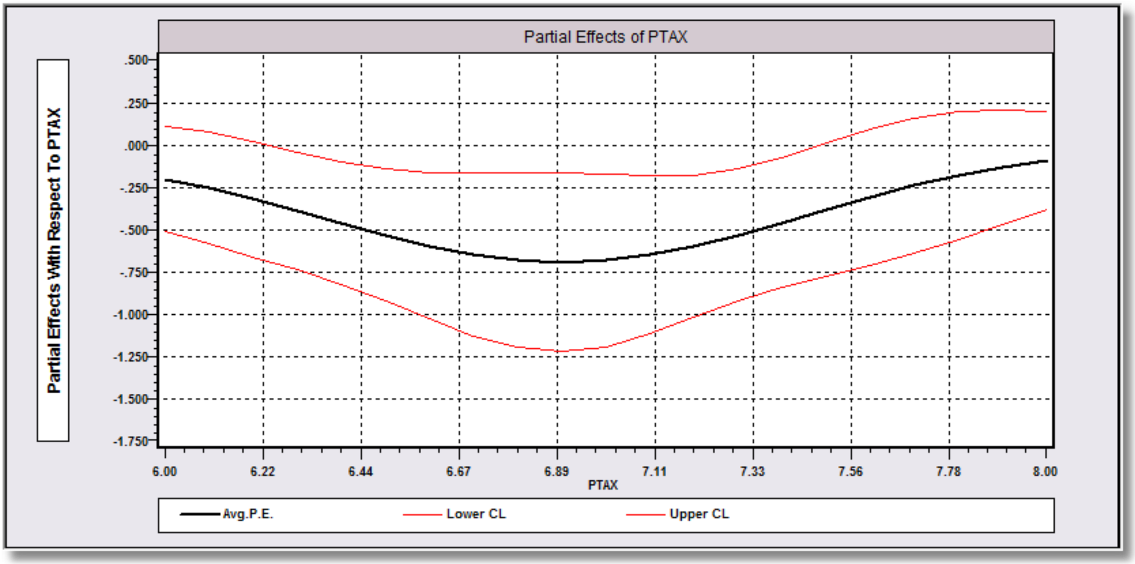


Figure E33.2 Partial Effects with Confidence Bounds

BIVARIATE ; Lhs = y ; Rh1 = x1 ; Rh2 = x2
; Rst = b01,bi,b21,b02,bi,b22,b23,r $

```
-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               TAXPRI
Log likelihood function       -75.50379
Estimation based on N =     80, K =   7
Inf.Cr.AIC  =  165.008 AIC/N =    2.063
--------+--------------------------------------------------------------------
     TAX|                Standard            Prob.      95% Confidence
    PRIV|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for TAX
Constant|     .14105       4.45802      .03  .9748     -8.59651    8.87860
     INC|    1.08018***     .39721     2.72  .0065       .30167    1.85869
    PTAX|   -1.51763**      .70085    -2.17  .0304     -2.89127    -.14398
        |Index    equation for PRIV
Constant|   -7.57707**     3.76951    -2.01  .0444    -14.96518    -.18897
     INC|    1.08018***     .39721     2.72  .0065       .30167    1.85869
     YRS|    -.01039        .03813     -.27  .7852      -.08512     .06434
    PTAX|    -.62120        .65264     -.95  .3412     -1.90036     .65795
        |Disturbance correlation
RHO(1,2)|    -.31361        .28024    -1.12  .2631      -.86287     .23564
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E33.2.9 Technical Details

Let

$$q_{im} = 2y_{im} - 1,\ m = 1,2.$$

The log likelihood function for the bivariate probit model is

$$\log L = \Sigma_i \log \Phi_2[q_{i1}\boldsymbol{\beta}_1'\mathbf{x}_{i1},\ q_{i2}\boldsymbol{\beta}_2'\mathbf{x}_{i2},\ q_{i1}q_{i2}\rho],$$

where we use Φ_2 to denote the bivariate standard normal CDF. We will also use $\phi_2[.,.,.]$ to denote the bivariate normal density function. We use ϕ and Φ, without subscripts, to denote the univariate standard normal density and CDF, respectively. For convenience in what follows, we will drop the observation subscript. Let

$$z_m = \boldsymbol{\beta}_m'\mathbf{x}_m,\ m = 1,2,$$

$$w_m = q_m z_m,\ m = 1,2,$$

$$\rho_* = q_1 q_2 \rho \text{ (note that the sign is the same if } y_1 = y_2\text{)},$$

$$g_1 = \phi(w_1)\Phi[(w_2 - \rho_* w_1)/(1 - \rho_*^2)^{1/2}],$$

$$g_2 = \phi(w_2)\Phi[(w_1 - \rho_* w_2)/(1 - \rho_*^2)^{1/2}].$$

Then, $\partial \log L/\partial\boldsymbol{\beta}_j = \Sigma_i (q_j g_j/\Phi_2)\mathbf{x}_j,\ j = 1,2$

and $\partial \log L/\partial\rho = \Sigma_i q_1 q_2 \phi_2/\Phi_2.$

NOTE: A corollary to this result is the marginal effect for the conditional mean function. Define **x** to be the union of $\mathbf{x}_1$ and $\mathbf{x}_2$, and define $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ conformably with zeros in the appropriate places so that $z_1 = \boldsymbol{\theta}_1'\mathbf{x} = \boldsymbol{\beta}_1'\mathbf{x}_1$ and $z_2 = \boldsymbol{\theta}_2'\mathbf{x} = \boldsymbol{\beta}_2'\mathbf{x}_2$. Then,

$$\frac{\partial\Phi_2(z_1,z_2,\rho)/\Phi(z_2)}{\partial\mathbf{x}} = \frac{g_1}{\Phi(z_2)}\boldsymbol{\theta}_1 + \left[\frac{g_2}{\Phi(z_2)} - \frac{\Phi_2(z_1,z_2,\rho)\phi(z_2)}{[\Phi(z_2)]^2}\right]\boldsymbol{\theta}_2$$

These are the parts that appear in the tables in the earlier applications, with zeros placed appropriately. (The computation is similar, albeit much more tedious, for a model with heteroscedasticity.)

During estimation, we use a transformation of ρ to avoid problems resulting from invalid correlation coefficients in computing the log likelihood. We define

$$\tau = \log[(1 + \rho)/(1 - \rho)]$$

Then, $\rho = [\exp(\tau) - 1] / [\exp(\tau) + 1].$

The range of τ is unrestricted. The model is viewed as a function of τ. During any computation of the log likelihood or its derivatives, we compute ρ from τ, then use ρ. Derivatives are then corrected to accommodate the transformation. The end result for your estimation of the model is that you will not receive diagnostics about ρ going out of the allowable range, again, since τ is unrestricted. Of course, it is possible for τ to become extremely large or small, which would imply that the model is gravitating toward the polar values of ρ. This signals a problem with your model, such as when the two Lhs variables are too highly correlated, or if an independent variable in one equation is a perfect predictor of the Lhs variable in the other. One point to note is that if your model command contains **; Output = 3**, the displayed output will show you the transient value of τ, not of ρ. For example, in estimating the unrestricted model above, the technical output at the last iteration shows

```
1st derivs.      .32279D-07   .33001D-06   .24128D-06   .29297D-07   .30245D-06
   .77385D-07   .21848D-06  -.48524D-08
Parameters:     -.20420D+01   .16462D+01  -.20155D+01  -.38604D+01   .35280D+00
  -.16221D-01  -.99484D-01  -.67176D+00
```

The last parameter in the list is τ, which appears above as the value -0.67176. But, the model output shows

```
RHO(1,2)|    -.32379          .29914    -1.08  .2791     -.91009    .26251
```

The value given is [exp(-.67176) - 1]/[exp(-.67176) + 1] = -.3237943.

Some of these results are produced as part of the output from your estimation. But, you may wish to do these computations for other purposes. The **CALC** and **CREATE** functions Bvn, Bvd, Bv1, and Bv2 are provided for this purpose. To use these, you must first compute either the variables z_1 and z_2 or the scalars, we'll call them c_1 and c_2, and obtain the value of ρ that you wish to use, which we'll call r. Then, use

```
NAMELIST   ; z = z1,z2 $
CREATE     ; phi2 = Bvn(z,r)       ? to compute the probability
           ; f2  = Bvd(z,r)        ? to compute the density
           ; g1 = Bv1(z,r)         ? to compute the g1 function
           ; g2 = Bv2(z,r)         $ to compute the g2 function
```

The same functions are available in **CALC**, except that in **CALC**, instead of the namelist with two names, you give both arguments. Thus,

```
CALC       ; cphi2 = Bvn(z1,z2,r)  ? to compute the probability
           ; cf2 = Bvf(z1,z2,r)    ? to compute the density
           ; cg1= Bv1(z1,z2,r)     ? to compute the g1 function
           ; cg2= Bv2(z1,z2,r)     $ to compute the g2 function
```

computes these functions for the single values given. Here are three applications. In all cases, we precede the computations with

```
NAMELIST   ; x1 =  ... Rhs for equation 1
           ; x2 =  ... Rhs for equation 2 $
```

Conditional Mean Predictions

```
BIVARIATE   ; Lhs = y1 ; Rh1 = x1 ; Rh2 = x2 $
CALC        ; k1 = Col(x1) ; k11 = k1+1 ; k12 = k1 + Col(x2) $
MATRIX      ; b1 = b(1,k1)  ; b2 = b(k11,k12) $
CREATE      ; z1 = b1'x1 ; z2 = b2'z2 $
NAMELIST    ; z = z1,z2 $
CREATE      ; ey1_y2 = Bvn(z,rho) / Phi(z2) $
```

Scale Factor for Marginal Effects, at the Means

```
BIVARIATE   ; Lhs = y1 ; Rh1 = x1 ; Rh2 = x2 $
CALC        ; k1 = Col(x1) ; k11 = k1+1 ;  k12 = k1 + Col(x2) $
MATRIX      ; b1 = b(1:k1) ; b2 = b(k11:k12) $
CREATE      ; z1 = b1'x1 ; z2 = b2'z2 $
CALC        ; cz1 = Xbr(z1) ?  Scale factors for derivatives
            ; cz2 = Xbr(z2) ?  of E[y1|y2=1] wrt x1 and x2
            ; me1 =  Bv1(cz1,cz2,rh0) / Phi(cz2)
            ; me2 = (Bv2(cz1,cz2,rho)  -
                      Bvn(cz1,cz2,rho) * N01(cz2) / Phi(cz2) ) / Phi(cz2) $
```

Lambda Variables for the Sample Selection Model

(This is another frequently asked question.) Section E55.3.3 describes the following sample selection model:

(y_1, y_2) determined by the bivariate probit model of this chapter

$$y = \delta'\mathbf{x} + u$$

$\text{Corr}(u,\varepsilon_1) = \rho_{u1}$, $\text{Corr}(u,\varepsilon_2) = \rho_{u2}$. But, $(y,\mathbf{x})$ are only observed when $(y_1 = 1, y_2 = 1)$. Estimation of this model is done by a two step extension of Heckman's method for a single probit selection model. The linear regression is computed using the observed data, with regression of y on $\mathbf{x}$, λ_1 and λ_2 where the two 'lambda' variables are, in fact, g_1/Φ_2 and g_2/Φ_2 as defined above. These variables are computed internally during estimation, but not retained anywhere accessible. We are often asked how these can be computed and, moreover, can they be computed for the 'nonselected' observations. Using what is already done above, the computation is actually simple. The full set of computations would be as follows: (This is generic. Only the first command would be specific to any application.)

```
CREATE      ; y1 = equation 1 Lhs variable ; y2 = equation 2 Lhs variable $
BIVARIATE   ; Lhs = y1, y2 ; Rh1 = x1 ; Rh2 = x2 $
CREATE      ; q1 = 2*y1 – 1; q2 = 2*y2 - 1 $
CALC        ; k1 = Col(x1)  ; k21 = k1 + 1 ; kvar = Col(b) $
MATRIX      ; b1 = b(1:k1)  ; b2 = b(k21:kvar) $
CREATE      ; v1 = q1*x1'b1 ; v2 = q2 * x2'b2 ; rs  = q1*q2*rho $
NAMELIST    ; v = v1,v2 $
CREATE      ; lambda1 = q1*Bv1(v,rs) / Bvn(v,rs)
            ; lambda2 = q2*Bv2(v,rs) / Bvn(v,rs) $
```

Finally, let

$$\delta = 1 / (1 - \rho_*^2)^{1/2},$$

$$v_1 = \delta(w_2 - \rho_* w_1) \text{ so } g_1 = \phi(w_1)\Phi(v_1)$$

and $$v_2 = \delta(w_1 - \rho_* w_2) \text{ so } g_2 = \phi(w_2)\Phi(v_2).$$

Then, $$\partial^2 \ln L/\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1' = \Sigma_i (-1/\Phi_2)[w_1 g_1 + \rho_*\phi_2 + g_1^2/\Phi_2]\mathbf{x}_1\mathbf{x}_1'$$

and likewise for $\boldsymbol{\beta}_2$. The mixed derivatives are:

$$\partial^2 \ln L/\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2' = \Sigma_i (q_1 q_2/\Phi_2)[\phi_2 - g_1 g_2/\Phi_2]\mathbf{x}_1\mathbf{x}_2',$$

$$\partial^2 \ln L/\partial\boldsymbol{\beta}_1\partial\rho = \Sigma_i\, q_2(\phi_2/\Phi_2)[\rho_*\delta v_1 - w_1 - g_1/\Phi_2]\mathbf{x}_1,$$

and $$\partial^2 \ln L/\partial\rho^2 = \Sigma_i (\phi_2/\Phi_2)[\delta^2\rho_*(1-\delta^2(w_1^2 + w_2^2 - 2\rho_* w_1 w_2)) + \delta^2 w_1 w_2 - \phi_2/\Phi_2].$$

The derivatives for a model with heteroscedasticity can be easily obtained by modification of the preceding. In the derivation above, $g_j = \partial\Phi_2(h_1,h_2,\rho)/\partial h_j$, so by making the appropriate modification of h_j, derivatives of the extended model can be obtained by differentiating h_j.

The bivariate normal CDF is approximated with a 15 point Gauss-Laguerre quadrature. The procedure used is as follows. We compute the upper (not the lower) tail area bivariate integral thusly:

$$\text{BVN}'(x_1, x_2, \rho) = \int_{x_1}^{\infty}\int_{x_2}^{\infty} \phi_2(x_1, x_2, \rho)\,dx_2\,dx_1$$

Let $$d_1 = 0.0 \text{ if } x_1 > 0.0, \text{ otherwise, } d_1 = 1,\ d_3 = 1.0 - 2.0 d_1$$

$$d_2 = 0.0 \text{ if } x_2 > 0.0, \text{ otherwise, } d_2 = 1,\ d_4 = 1.0 - 2.0 d_2$$

Then, $$B = \frac{1}{\sqrt{2\pi}}\sum_{h=1}^{15} w_h\Phi(a_h)\exp[z_h - .5(z_h + d_3 x_1)^2]$$

where z_h and w_h = the nodes and weights for the quadrature, and

$$a_h = d_4[\rho(x_1 + d_3 z_h) - x_2] / \sqrt{1-\rho^2}$$

Then, $$\text{BVN}'(x_1, x_2, \rho) \approx d_3 d_4 B - d_1 d_2 + d_1\Phi(-x_2) + d_2\Phi(-x_1).$$

The complementary CDF, integrating from $-\infty$ to the argument, is obtained just by sending $-x_1$ and/or $-x_2$ to this computation.

E33.3 Tetrachoric Correlation

The tetrachoric correlation is a measure of the correlation between two binary variables. The familiar Pearson, product moment correlation is inappropriate as it is used for continuous variables. The tetrachoric correlation coefficient is equivalent to the correlation coefficient in the following bivariate probit model:

$$y_1{}^* = \mu + \varepsilon_1, \qquad y_1 = 1(y_1{}^* > 0)$$

$$y_2{}^* = \mu + \varepsilon_2, \qquad y_2 = 1(y_2{}^* > 0)$$

$$(\varepsilon_1,\varepsilon_2) \sim N_2[(0,0),(1,1,\rho)]$$

The applicable literature contains a number of approaches to estimation of this correlation coefficient, some a bit ad hoc. We proceed directly to the implied maximum likelihood estimator. You can fit this 'model' with

BIVARIATE ; Lhs = y1,y2 ; Rh1 = one ; Rh2 = one $

The reported estimate of ρ is the desired estimate. *LIMDEP* notices if your model does not contain any covariates in the equation, and notes in the output that the estimator is a tetrachoric correlation. The results below based on the German health care data show an example.

```
-----------------------------------------------------------------------------
FIML Estimation of Tetrachoric Correlation
Dependent variable                DOCHOS
Log likelihood function    -25898.27183
Estimation based on N =  27326, K =   3
Inf.Cr.AIC  =51802.544 AIC/N =    1.896
--------+--------------------------------------------------------------------
  DOCTOR|                  Standard            Prob.      95% Confidence
HOSPITAL|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Estimated alpha for P[DOCTOR  =1] = F(alpha)
Constant|     .32949***      .00773    42.61  .0000      .31433    .34465
        |Estimated alpha for P[HOSPITAL=1] = F(alpha)
Constant|   -1.35540***      .01074  -126.15  .0000    -1.37646  -1.33434
        |Tetrachoric Correlation between DOCTOR   and HOSPITAL
RHO(1,2)|     .31106***      .01357    22.92  .0000      .28446    .33766
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The preceding suggests an interpretation for the bivariate probit model; the correlation coefficient reported is the *conditional* (on the independent variables) tetrachoric correlation.

The computation in the preceding can be generalized to a set of M binary variables, $y_1,...,y_M$. The tetrachoric correlation matrix would be the $M\times M$ matrix, **R**, whose off diagonal elements are the ρ_{mn} coefficients described immediately above. There are several ways to do this computation, again, as suggested by a literature that contains recipes. Once again, the maximum likelihood estimator turns out to be a useful device.

A direct approach would involve expanding the latent model to

$$y_1^* = \mu + \varepsilon_1, \quad y_1 = 1(y_1^* > 0)$$

$$y_2^* = \mu + \varepsilon_2, \quad y_2 = 1(y_2^* > 0)$$

$$\ldots$$

$$y_M^* = \mu + \varepsilon_M, \quad y_M = 1(y_M^* > 0)$$

$$(\varepsilon_1,\varepsilon_2,...,\varepsilon_M) \sim N_M[\mathbf{0},\mathbf{R}]$$

The appropriate estimator would be *LIMDEP*'s multivariate probit estimator, **MPROBIT**, which can handle up to $M = 20$. The correlation matrix produced by this procedure is precisely the full information MLE of the tetrachoric correlation matrix. However, for any M larger than two, this requires use of the GHK simulator to maximize the simulated log likelihood, and is extremely slow. The received estimators of this model estimate the correlations pairwise, as shown earlier. For this purpose, the FIML estimator is unnecessary. The matrix can be obtained using bivariate probit estimates. The following procedure would be useable:

```
NAMELIST    ; y = y1,y2,...,ym $
CALC        ; m = Col(y) $
MATRIX      ; r = Iden(m) $
PROCEDURE $
DO FOR      ; 20 ; i = 2,m $
CALC        ; i1 = i - 1 $
DO FOR      ; 10 ; j = 1,i1 $
BIVARIATE   ; Lhs = y:i, y:j ; Rh1 = one ; Rh2 = one $
MATRIX      ; r(i,j) = rho $
MATRIX      ; r(j,i) = rho $
ENDDO       ; 10 $
ENDDO       ; 20 $
ENDPROC $
EXECUTE     ; Quietly $
```

A final note, the preceding approach is not fully efficient. Each bivariate probit estimates (μ_m,μ_n) which means that μ_m is estimated more than once when $m > 1$. A minimum distance estimator could be used to reconcile these after all the bivariate probit estimates are computed. But, since the means are nuisance parameters in this model, this seems unlikely to prove worth the effort.

E33.4 Bivariate Probit Model with Sample Selection

In the bivariate probit setting, data on y_1 might be observed only when y_2 equals one. For example, in modeling loan defaults with a sample of applicants, default will only occur among applicants who are granted loans. Thus, in a bivariate probit model for the two outcomes, the observed default data are nonrandomly selected from the set of applicants. The model is

$$z_{i1} = \boldsymbol{\beta}'\mathbf{x}_{i1} + \varepsilon_{i1},\ y_{i1} = \text{sgn}(z_{i1}),$$

$$z_{i2} = \boldsymbol{\beta}'\mathbf{x}_{i2} + \varepsilon_{i2},\ y_{i2} = \text{sgn}(z_{i2}),$$

$$\varepsilon_{i1},\varepsilon_{i2} \sim \text{BVN}(0,0,1,1,\rho),$$

$$(y_{i1},\mathbf{x}_{i1}) \text{ is observed only when } y_{i2} = 1.$$

This is a type of sample selectivity model. The estimator was proposed by Wynand and van Praag (1981). An extensive application which uses choice based sampling as well is Boyes, Hoffman, and Low (1989). (See also Greene (1992 and 2011).) The sample selection model is obtained by adding

; Selection (or just ; Sel)

to the **BIVARIATE PROBIT** command. All other options and specifications are the same as before. Except for the diagnostic table which indicates that this model has been chosen, the results for the selection model are the same as for the basic model.

E33.4.1 Technical Details

The log likelihood for the bivariate probit model with selection is

$$\begin{aligned}\text{Log-}L = &\sum\nolimits_{y_2=1,y_1=1} \log \Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_{i1}, \boldsymbol{\beta}_2'\mathbf{x}_{i2}, \rho] \\ &+ \sum\nolimits_{y_2=1,y_1=0} \log \Phi_2[-\boldsymbol{\beta}_1'\mathbf{x}_{i1}, \boldsymbol{\beta}_2'\mathbf{x}_{i2}, -\rho] \\ &- \sum\nolimits_{y_2=0} \log \Phi[-\boldsymbol{\beta}_2'\mathbf{x}_{i2}].\end{aligned}$$

The necessary first and second derivatives are given in Section E33.6.

NOTE: This is one of several sample selection models estimated by maximum likelihood with *LIMDEP*. In this setting, there is no 'lambda' variable as there is in the regression model with sample selection (see Chapter E52). Heckman's (1979) selection correction variable applies to the linear regression model estimated with two step least squares, but not generally to models fit by maximum likelihood. For testing for selection effects, the appropriate approach is to test the hypothesis of no effects, which results if ρ equals zero.

NOTE: You may code y_1 as 0.0 for the nonselected (nonobserved) observations in this model. The correct value to use (or ignore) is determined by the program during estimation.

Further details on this model, with an application and technical background appear in Section E33.2.9.

The following carries out a sampling experiment that conforms exactly to the assumptions of the model. The lhs variables y1 and y2 are governed by a bivariate probit model with coefficient vectors $\beta_1 = \beta_2 = (0,1,1)$ and $\rho = .5$. However, y1s is missing when y2 equals zero, so the appropriate approach is the selection model. As seen below, estimation proceeds routinely. Partial effects and subsequent analysis would be the same as for the bivariate probit model prior to the selection. The force of the revision of the estimator is to use an approach that produces consistent estimators of the model parameters. It is not a fundamentally different model. For comparison, the full sample results are shown as well. Not surprisingly, they are essentially the same.

```
SAMPLE      ; 1-1000 $
CREATE      ; x1 = Rnn(0,1) ; x2 = Rnn(0,1) ; x3 = Rnn(0,1) $
CREATE      ; u1 = Rnn(0,1) ; u2 = .5*(u1 + Rnn(0,1)) $
CREATE      ; y1 = (x1 + x3 + u1) > 0 ; y2 = (x2 + x3 + u2) > 0 $
BIVARIATE   ; Lhs = y1,y2 ; Rh1=one,x1,x3 ; Rh2 = one,x2,x3 $
CREATE      ; y1s = y1 ; If(y2 = 0)y1s = -999 $
BIVARIATE   ; Lhs = y1s,y2 ; Rh1 = one,x1,x3 ; Rh2 = one,x2,x3 ; Selection $
```

```
-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               Y1SY2
Selection model based on Y2
Selected sample:   481, Nonselected:   519
--------+--------------------------------------------------------------------
      Y1|                 Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for Y1
Constant|    -.04304         .11125     -.39  .6989     -.26109    .17502
      X1|    1.06002***      .11067     9.58  .0000      .84311   1.27694
      X3|    1.05358***      .11942     8.82  .0000      .81953   1.28763
        |Index    equation for Y2
Constant|    -.04318         .05642     -.77  .4441     -.15375    .06740
      X2|    1.43944***      .09268    15.53  .0000     1.25779   1.62108
      X3|    1.35190***      .08580    15.76  .0000     1.18373   1.52007
        |Disturbance correlation
RHO(1,2)|     .57713***      .12950     4.46  .0000      .32332    .83095
--------+--------------------------------------------------------------------
(Full Sample Results)
--------+--------------------------------------------------------------------
        |Index    equation for Y1
Constant|    -.01822         .05000     -.36  .7156     -.11622    .07978
      X1|    1.05590***      .07082    14.91  .0000      .91709   1.19471
      X3|     .96638***      .07133    13.55  .0000      .82656   1.10619
        |Index    equation for Y2
Constant|    -.04284         .05571     -.77  .4419     -.15203    .06635
      X2|    1.41857***      .09039    15.69  .0000     1.24140   1.59573
      X3|    1.34578***      .08422    15.98  .0000     1.18070   1.51085
        |Disturbance correlation
RHO(1,2)|     .54221***      .06755     8.03  .0000      .40981    .67460
--------+--------------------------------------------------------------------
```

E33.5 Simultaneity in the Binary Variables

A simultaneous equations sort of model would appear as

$$z_{i1} = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \gamma_1 y_{i2} + \varepsilon_{i1},\ \ y_{i1} = 1 \text{ if } z_{i1} > 0,\ y_{i1} = 0 \text{ otherwise},$$

$$z_{i2} = \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \gamma_1 y_{i1} + \varepsilon_{i2},\ \ y_{i2} = 1 \text{ if } z_{i2} > 0,\ y_{i2} = 0 \text{ otherwise},$$

$$[\varepsilon_{i1},\varepsilon_{i2}] \sim \text{bivariate normal (BVN) } [0,0,1,1,\rho],\ -1 < \rho < 1,$$

individual observations on y_1 and y_2 are available for all i.

It would follow from the construction that

$$\text{Prob}[y_1 = 1,\ y_2 = 1] = \Phi_2\,(\boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_1 y_2,\ \boldsymbol{\beta}_2'\mathbf{x}_2 + \gamma_2 y_1,\ \rho]$$

and likewise for the other cells, where y_1 and y_2 are two binary variables. Unfortunately, the model as stated is not internally consistent, and is inestimable. Ultimately, it is not identifiable. As a practical matter, you can verify this by attempting to devise a way to simulate a sample of observations that conforms exactly to the assumptions of the model. In this case, there is none because there is no linear reduced form for this model. (The approach suggested by Maddala (1983) is not consistent.) *LIMDEP* will detect this condition and decline to attempt to do the estimation. For example:

BIVARIATE PROBIT ; Lhs = y1,y2 ; Rh1 = one,x1,x3,y2 ; Rh2 = one,x2,x3,y1 $

produces a diagnostic,

```
Error   809: Fully simultaneous BVP model is not identified
```

NOTE: Unlike the case in linear simultaneous equations models, nonidentifiability does not prevent ‘estimation’ in this model. (2SLS estimates cannot be computed when there are too few instrumental variables, which is the signature of nonidentifiability in a linear context.) With the ‘fully simultaneous bivariate probit model,’ it is possible to maximize what purports to be a log likelihood function – numbers will be produced that might even look reasonable. However, as noted, the model itself is nonsensical – it lacks internal coherency.

To illustrate the effect, the following program attempts to estimate a fully simultaneous bivariate probit model. In the first version, the optimizer appears to find a solution, though the theoretical result is that the results are not meaningful. In the second version, the coefficient on $y1$ in the second equation is constrained to equal zero. This produces the generally useable recursive model described in the next section. We use the built in **MAXIMIZE** command to construct our own log likelihood maximizer for this model, as *LIMDEP* will refuse it. The optimization trace for the model is punctuated with error messages. But, ultimately a set of ordinary looking results is produced. The correlation coefficient of .99334 is suspiciously large, however. (This application also demonstrates using **MAXIMIZE** to construct an estimator. **MAXIMIZE** is described in Chapter E66.)

The commands are:

```
NAMELIST   ; y = tax, priv
           ; x1 = one,inc,ptax ; x2  = one,inc,yrs,ptax $
CREATE     ; y1 = tax ; y2 = priv
           ; q1 = 2*y1-1 ; q2 = 2*y2-1 ; q12 = q1*q2 $
PROBIT     ; Lhs = y1 ; Rhs = x1,y2 $
MATRIX     ; bc1 = b $
PROBIT     ; Lhs = y2 ; Rhs = x2,y1 $
MATRIX     ; bc2 = b $
MAXIMIZE   ; Labels = b11,b12,b13,c1,b21,b22,b23,b24,c2,ro
           ; Start = bc1,bc2,0
           ; Fcn = bx1 = q1*(b11'x1+c1*y2) |
                   bx2 = q2*(b21'x2+c2*y1) |
                   r12 = q12*ro |
                   Log(Bvn(bx1,bx2,r12))
           ; Output = 3 $
```

```
Itr 13 F=  .7412D+02 gtHg=  .3234D+01 chg.F=  .7508D-01 max|db|=  .6318D+01
  Error   590: Obs.=      1 Cannot compute function: BadFnPrm
Warning   137: Iterations: function not computable at crnt.trial estimates
1st derivs.     .88960D+01   .88460D+02   .63332D+02  -.13309D+01  -.38917D+01
  -.38549D+02  -.52808D+02  -.26737D+02  -.58411D+01  -.19929D+02
Parameters:    -.47503D+01   .15252D+01  -.13869D+01  -.19755D+01  -.51988D+01
   .13039D+01  -.72843D-01  -.10489D+01  -.22773D+01   .52288D+00
Itr 14 F=  .7405D+02 gtHg=  .1469D+03 chg.F=  .7650D-01 max|db|=  .1246D+05
  Error   590: Obs.=      1 Cannot compute function: BadFnPrm
1st derivs.     .10269D+02   .10260D+03   .73035D+02  -.18254D+01  -.50748D+01
  -.50777D+02  -.64088D+02  -.34603D+02  -.79427D+01  -.29606D+02
Parameters:    -.66086D+01   .16254D+01  -.12459D+01  -.28014D+01  -.61176D+01
   .13582D+01  -.98235D-01  -.90757D+00  -.30863D+01   .89237D+00
Itr 15 F=  .7246D+02 gtHg=  .1505D+01 chg.F=  .1589D+01 max|db|=  .9591D+00
1st derivs.     .56869D+01   .58117D+02   .40923D+02  -.29843D+01  -.30518D+01
  -.31419D+02  -.53961D+02  -.19614D+02  -.66064D+01  -.28128D+02
Parameters:    -.78376D+01   .18961D+01  -.14702D+01  -.27742D+01  -.67497D+01
   .95559D+00  -.95663D-01  -.25752D+00  -.28599D+01   .87896D+00
Itr 16 F=  .7166D+02 gtHg=  .2355D+00 chg.F=  .8009D+00 max|db|=  .9301D-01
1st derivs.     .55771D+01   .57096D+02   .40217D+02  -.28408D+01  -.38132D+01
  -.39006D+02  -.56699D+02  -.25002D+02  -.68968D+01  -.27174D+02
Parameters:    -.78699D+01   .19107D+01  -.14889D+01  -.27437D+01  -.67786D+01
   .96026D+00  -.93481D-01  -.26820D+00  -.28316D+01   .86761D+00
Itr 17 F=  .7165D+02 gtHg=  .1715D+00 chg.F=  .1205D-01 max|db|=  .4563D-01
  Error   590: Obs.=      1 Cannot compute function: BadFnPrm
1st derivs.     .49681D+01   .50995D+02   .35963D+02  -.29773D+01  -.36462D+01
  -.37220D+02  -.55215D+02  -.23852D+02  -.67399D+01  -.26763D+02
Parameters:    -.78501D+01   .19252D+01  -.15148D+01  -.27175D+01  -.67916D+01
   .97156D+00  -.91622D-01  -.28746D+00  -.27932D+01   .84912D+00
Itr 18 F=  .7169D+02 gtHg=  .1660D+01 chg.F=  .4775D-01 max|db|=  .2396D+01
  Error   590: Obs.=      1 Cannot compute function: BadFnPrm
1st derivs.     .93073D+01   .94865D+02   .66198D+02  -.11718D+01  -.51859D+01
  -.53983D+02  -.67518D+02  -.34423D+02  -.87711D+01  -.36230D+02
Parameters:    -.80342D+01   .18018D+01  -.12967D+01  -.29355D+01  -.67084D+01
   .92192D+00  -.10657D+00  -.18805D+00  -.31096D+01   .99177D+00
```

```
Itr 19 F=  .7112D+02 gtHg=  .3194D+00 chg.F=  .5769D+00 max|db|=  .4324D+00
1st derivs.     .84502D+01   .86125D+02   .60156D+02  -.19133D+01  -.36120D+01
  -.38132D+02  -.61097D+02  -.23430D+02  -.78864D+01  -.37154D+02
Parameters:    -.79335D+01   .17807D+01  -.12790D+01  -.29658D+01  -.66513D+01
   .95284D+00  -.10767D+00  -.23455D+00  -.31267D+01   .99334D+00
Itr 20 F=  .7109D+02 gtHg=  .4565D+00 chg.F=  .3112D-01 max|db|=  .2227D+00
Itr 20 F=  .7109D+02 gtHg=  .1353D+03 chg.F=  .8417D-03 max|db|=  .5675D+03
Line search at iteration 20 does not improve fn. Exiting optimization.
Function=  .73442696419D+02, at entry,  .71165594038D+02 at exit
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function        71.16559
Restricted log likelihood        .00000
Chi squared [  10 d.f.]       142.33119
Significance level               .00000
Estimation based on N =     80, K =   0
Inf.Cr.AIC  = -142.331 AIC/N =   -1.779
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     B11|   -7.93350*      4.59041    -1.73  .0839   -16.93055   1.06355
     B12|    1.78070***     .59633     2.99  .0028      .61193   2.94948
     B13|   -1.27900        .89894    -1.42  .1548    -3.04089    .48290
      C1|   -2.96584***     .68236    -4.35  .0000    -4.30323  -1.62845
     B21|   -6.65132       6.27517    -1.06  .2892   -18.95042   5.64778
     B22|     .95284        .76396     1.25  .2123     -.54449   2.45017
     B23|    -.10767***     .04116    -2.62  .0089     -.18834   -.02700
     B24|    -.23455        .83708     -.28  .7793    -1.87519   1.40609
      C2|   -3.12667***     .65501    -4.77  .0000    -4.41046  -1.84289
      RO|     .99334***     .10776     9.22  .0000      .78213   1.20454
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

We now attempt the same optimization, but force the coefficient on one of the endogenous variables to equal zero. This identifies the model, and leads to a reasonable set of estimates. No error or warning messages occur during the optimization.

```
? Constrain coefficient on y1 in equation 2 to equal zero.

MATRIX      ; bc2(5) = 0 $
MAXIMIZE    ; Labels = b11,b12,b13,c1,b21,b22,b23,b24,c2,ro
            ; Start = bc1,bc2,0
            ; Fcn = bx1 = q1*(b11'x1+c1*y2) |
                    bx2 = q2*(b21'x2+c2*y1) |
                    r12 = q12*ro |
                    Log(Bvn(bx1,bx2,r12))
            ; Fix = c2     ? forces c2 to be fixed at the starting value
            ; Output = 3 $
```

```
Itr 23 F=  .7421D+02 gtHg=  .1049D-05 chg.F=  .4283D-09 max|db|=  .5631D-05
Itr 23 F=  .7421D+02 gtHg=  .4083D-04 chg.F=  .1265D-11 max|db|=  .4575D-03
1st derivs.     .41475D-06   .55241D-05   .46484D-05   .44285D-06  -.50979D-06
  -.36037D-05   .10967D-06  -.30851D-05   .26917D-06
Parameters:    -.68059D+00   .12277D+01  -.16316D+01   .98177D+00  -.28146D+01
   .16264D+00  -.34840D-01   .46046D-01  -.83118D+00
Itr 24 F=  .7421D+02 gtHg=  .9836D-06 chg.F=  .1307D-11 max|db|=  .5609D-05
Itr 24 F=  .7421D+02 gtHg=  .8680D-05 chg.F=  .7105D-13 max|db|=  .6700D-04
Line search at iteration 24 does not improve fn. Exiting optimization.
Function=  .76322747822D+02, at entry,  .74211794755D+02 at exit

-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function         74.21179
Restricted log likelihood         .00000
Chi squared [   9 d.f.]        148.42359
Significance level                .00000
Estimation based on N =      80, K =   0
Inf.Cr.AIC  = -148.424 AIC/N =   -1.855
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     B11|    -.68059        4.05342      -.17  .8667    -8.62515   7.26396
     B12|    1.22768         .81424      1.51  .1316     -.36820   2.82356
     B13|   -1.63161         .99597     -1.64  .1014    -3.58368    .32047
      C1|     .98177         .95912      1.02  .3060     -.89808   2.86162
     B21|   -2.81455        5.51612      -.51  .6099   -13.62594   7.99684
     B22|     .16264         .76312       .21  .8312    -1.33304   1.65832
     B23|    -.03484         .04247      -.82  .4120     -.11808    .04840
     B24|     .04605         .98275       .05  .9626    -1.88011   1.97220
      C2|        0.0    .....(Fixed Parameter).....
      RO|    -.83118         .57072     -1.46  .1453    -1.94977    .28741
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

E33.6 Recursive Bivariate Probit Model

A slight modification of the model in the previous section is identified and used in many recent applications. Consider the model for the probability of the event y_1 = 0/1 and y_2 = 0/1 assuming $\gamma_2 = 0$.

$$\text{Prob}[y_1 = 1, y_2 = 1 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_1, \; \boldsymbol{\beta}_2'\mathbf{x}_2, \rho)$$

$$\text{Prob}[y_1 = 1, y_2 = 0 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1, \; -\boldsymbol{\beta}_2'\mathbf{x}_2, -\rho)$$

$$\text{Prob}[y_1 = 0, y_2 = 1 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(-\boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_1, \; \boldsymbol{\beta}_2'\mathbf{x}_2, -\rho)$$

$$\text{Prob}[y_1 = 0, y_2 = 0 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(-\boldsymbol{\beta}_1'\mathbf{x}_1, \; -\boldsymbol{\beta}_2'\mathbf{x}_2, \rho)$$

This is a recursive simultaneous equations model. Surprisingly enough, it can be estimated by full information maximum likelihood *ignoring the simultaneity* in the system;

```
BIVARIATE   ; Lhs =  y1, y2
            ; Rh1 =  x1,y2 ; Rh2 = x2 $
```

(A proof of this result is suggested in Maddala (1983, p. 123) and pursued in Greene (1998).) An application of the result to the gender economics study is given in Greene (1998). Some extensions are presented in Greene (2003, 2011).

This model presents the same ambiguity in the conditional mean function and marginal effects that were noted earlier in the bivariate probit model. The conditional mean for y_1 is

$$E[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2] \;=\; \Phi_2\,(\boldsymbol{\beta}_1'\mathbf{x}_1 + \;\gamma_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho)\;/\;\Phi(\boldsymbol{\beta}_2'\mathbf{x}_2)$$

for which derivatives were given earlier. Given the form of this result, we can identify *direct* and *indirect* effects in the conditional mean:

$$\frac{\partial E[y_1 \mid y2 = 1, \mathbf{x}_1, \mathbf{x}_2]}{\partial \mathbf{x}_1} = \frac{g_1}{\Phi(\boldsymbol{\beta}'\mathbf{x}_2)}\boldsymbol{\beta}_1 \;=\; \text{direct effects}$$

$$\frac{\partial E[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2]}{\partial \mathbf{x}_2} = \left[\frac{g_2}{\Phi(\boldsymbol{\beta}'\mathbf{x}_2)} - \frac{\Phi_2(\boldsymbol{\beta}'\mathbf{x}_1, \boldsymbol{\beta}'\mathbf{x}_2, \rho)\phi(z_2)}{\left[\Phi(\boldsymbol{\beta}'\mathbf{x}_2)\right]^2}\right]\boldsymbol{\beta}_2 \;=\; \text{indirect effects}$$

The unconditional mean function is

$$\begin{aligned} \mathrm{E}[y_1 \mid \mathbf{x}_1, \mathbf{x}_2] &= \Phi(\boldsymbol{\beta}_2'\mathbf{x}_2)\,\mathrm{E}[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2] + [1\text{-}\Phi(\boldsymbol{\beta}_2'\mathbf{x}_2)]\,\mathrm{E}[y_1 \mid y_2 = 0, \mathbf{x}_1, \mathbf{x}_2] \\ &= \Phi_2\,(\boldsymbol{\beta}_1'\mathbf{x}_1 + \;\gamma_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho) + \Phi_2\,(\boldsymbol{\beta}_1'\mathbf{x}_1, -\boldsymbol{\beta}_2'\mathbf{x}_2, -\rho) \end{aligned}$$

Derivatives for marginal effects can be derived using the results given earlier. Analysis appears in Greene (1998). The decomposition is done automatically when you specify a recursive bivariate probit model – one in which the second Lhs variable appears in the Rhs of the first equation.

The following demonstrates this by extending the model in Section E33.2.8. Note the appearance of *priv* on the Rhs of the first equation, *x1*.

```
NAMELIST    ; y = tax, priv
            ; x1 = one,inc,ptax,priv ; x2 = one,inc,yrs,ptax $
BIVARIATE   ; Lhs = tax,priv ; Rh1 = x1 ; Rh2 = x2 ; Partial Effects $
```

```
--------------------------------------------------------------------------------
FIML - Recursive Bivariate Probit Model
Dependent variable                PRITAX
Log likelihood function        -74.21179
Estimation based on N =      80, K =   9
Inf.Cr.AIC  =  166.424 AIC/N =    2.080
--------+-----------------------------------------------------------------------
    PRIV|                  Standard            Prob.      95% Confidence
     TAX|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Index    equation for PRIV
Constant|   -2.81454        5.51612     -.51  .6099   -13.62594   7.99687
     INC|     .16264         .76312      .21  .8312    -1.33304   1.65832
     YRS|    -.03484         .04247     -.82  .4120     -.11808    .04840
    PTAX|     .04605         .98275      .05  .9626    -1.88011   1.97220
        |Index    equation for TAX
Constant|    -.68059        4.05341     -.17  .8667    -8.62513   7.26394
     INC|    1.22768         .81424     1.51  .1316     -.36820   2.82356
    PTAX|   -1.63160         .99598    -1.64  .1014    -3.58368    .32047
    PRIV|     .98178         .95912     1.02  .3060     -.89807   2.86162
        |Disturbance correlation
RHO(1,2)|    -.83119         .57072    -1.46  .1453    -1.94977    .28740
--------+-----------------------------------------------------------------------

---------------------------------------------------------------
Decomposition of Partial Effects for Recursive Bivariate Probit
Model is     PRIV = F(x1b1), TAX      = F(x2b2+c*PRIV    )
Conditional mean function is E[TAX     |x1,x2] =
            Phi2(x1b1,x2b2+gamma,rho) + Phi2(-x1b1,x2b2,-rho)
Partial effects for continuous variables are derivatives.
Partial effects for dummy variables (*) are first differences.
Direct effect is wrt x2, indirect is wrt x1, total is the sum.
---------------------------------------------------------------
Variable   Direct Effect   Indirect Effect    Total Effect
---------+---------------+----------------+-------------------
     INC |   .4787001        .0169062          .4956064
    PTAX |  -.6362002        .0047864         -.6314138
     YRS |   .0000000       -.0036217         -.0036217
---------+-----------------------------------------------------
```

The decomposition of the partial effects accounts for the direct and indirect influences. Note that there is no partial effect given for *priv* because this variable is endogenous. It does not vary ‘partially.’

E33.7 Bivariate Probit Models with Partial Observability

We consider a bivariate probit model in which, instead of observing both y_{i1} and y_{i2}, we observe the product, $y_i = y_{i1}y_{i2}$. The situation arises when we observe the final outcome of two decision processes which lead to a single conclusion. Basic references are Poirier (1980), Abowd and Farber (1982), and Meng and Schmidt (1985). There are three variants available:

Poirier

In the Poirier model, y_1 and y_2 are simultaneously determined, and ε_1 and ε_2 are correlated. Then,

$$\text{Prob}[y = 1] = \Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_1,\boldsymbol{\beta}_2'\mathbf{x}_2,\rho],$$

$$\text{Prob}[y = 0] = 1 - \text{Prob}[y = 1].$$

As an example, Poirier cites the case of a joint decision made by two people each of whom has veto power.

Abowd and Farber

In the Abowd and Farber model, y_1 and y_2 are determined sequentially, and ε_1 and ε_2 are uncorrelated. The model is

$$\text{Prob}[y = 1] = \text{Prob}[y_1 = 1]\ \text{Prob}[y_2 = 1] = \Phi(\boldsymbol{\beta}_1'\mathbf{x}_1)\ \Phi(\boldsymbol{\beta}_2'\mathbf{x}_2),$$

$$\text{Prob}[y = 0] = 1 - \text{Prob}[y = 1].$$

The Abowd and Farber variant results from the Poirier model when ρ equals zero. Abowd's example is that of an individual who decides to enter a queue, then subsequently decides whether or not to accept an offer upon reaching his or her turn in the queue. *LIMDEP* produces full information maximum likelihood estimates of all parameters in both of these models. Since they have only a single dependent variable (the product, $y_1 \times y_2$), these partial observability models are estimated as probit models, not bivariate probit models. The Poirier variant is requested simply by adding the second list of exogenous variables to the **PROBIT** command. I.e.,

PROBIT ; Lhs = y ; Rh1 = x1list ; Rh2 = x2list $

The Abowd and Farber variant is requested by adding

; Selection

to the Poirier variant.

Starting values for both of these models are the ordinary least squares estimates and ρ equal to zero for the Poirier variant. As always, you may provide your own starting values if you like. If so, you must provide a value of ρ for the Poirier variant. In both models, the full set of parameters involves $[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2]$. You may use **; Rst** in the example at the end of Section E33.2.8, to impose both within and cross equation restrictions on the models. For the listing of predictions and residuals, **; Keep** and **; Res**, the same prediction rule as in the univariate probit model is used. That is, for each observation, we compute Prob[y = 1], then predict y = 1 if the probability is greater than .5.

Meng and Schmidt

In the Meng and Schmidt model, y_1 and y_2 are defined by separate probit models;

if $y_1 = 1$, both y_1 and y_2 are observed,

if $y_1 = 0$, then only $y_1 \times y_2$ is observed.

The setup involves both Lhs variables, so it is estimated as a bivariate probit model.

NOTE: When y_1 is zero, you should code y_2 as zero also.

The command for the Meng and Schmidt model is

```
BIVARIATE   ; Lhs  = y1,y2
            ; Rh1 = Rhs for first equation
            ; Rh2 = Rhs for second equation
            ; Model = Partial $
```

All other options are the same as for other bivariate probit models.

NOTE: The Meng and Schmidt model is identical to the bivariate probit model with sample selection, with the two variables reversed.

E33.7.1 Example

The following experiment will illustrate the computations in the partial observability models: The data are simulated, and correspond exactly to the assumptions of the models.

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-500 $
CREATE      ; x1 = Rnn(0,1) ; x2 = Rnn(0,1)
            ; y1 = x1 + Rnn(0,1) ; y1 = y1 > 0
            ; y2 = x2 + Rnn(0,1) ; y2 = y2 > 0 ; y2 = y2*y1 $
CREATE      ; y = y1*y2 $
```

Estimate the Meng and Schmidt model.

```
BIVARIATE   ; Lhs = y1,y2 ; Rh1 = one,x1 ; Rh2 = one,x2 ; Model = Partial $
```

Estimate the Poirier model.

```
PROBIT      ; Lhs = y ; Rh1 =one,x1 ; Rh2 = one,x2 ; Partial Effects $
```

Estimate the Abowd and Farber model.

```
PROBIT      ; Lhs = y ; Rh1 =one,x1 ; Rh2  = one,x2
            ; Selection ; Partial Effects $
```

Meng and Schmidt Model

```
-----------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable                   Y1Y2
Log likelihood function        -371.68472
Estimation based on N =    500, K =   5
Inf.Cr.AIC  =  753.369 AIC/N =    1.507
Meng & Schmidt Partial Observability Model
--------+--------------------------------------------------------------------
      Y1|                Standard            Prob.      95% Confidence
      Y2|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for Y1
Constant|    -.03899        .06445     -.60  .5452     -.16532    .08733
      X1|    1.00245***     .09074    11.05  .0000      .82460   1.18030
        |Index    equation for Y2
Constant|     .04903        .16503      .30  .7664     -.27442    .37247
      X2|    1.05905***     .12965     8.17  .0000      .80494   1.31316
        |Disturbance correlation
RHO(1,2)|    -.15595        .22990     -.68  .4976     -.60655    .29465
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Poirier Model

Note, the appearance of an estimate of ρ indicates the Poirier model.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                      Y
Log likelihood function        -207.51923
Restricted log likelihood      -268.42420
Chi squared [   4 d.f.]         121.80995
Significance level                 .00000
McFadden Pseudo R-squared        .2268982
Estimation based on N =    500, K =   5
Inf.Cr.AIC  =  425.038 AIC/N =     .850
Partial Observability Model
Hosmer-Lemeshow chi-squared = *********
P-value=  .00000 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .13167        .33321      .40  .6927     -.52140    .78474
      X1|     .99518***     .25503     3.90  .0001      .49532   1.49504
Constant|    -.07516        .27003     -.28  .7808     -.60441    .45410
      X2|     .90165***     .20472     4.40  .0000      .50041   1.30289
Rho(1,2)|    -.24122        .41392     -.58  .5600   -1.05249    .57005
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
       Y|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|     .00708         .00123      .07  .9468     -.20067    .21482
      X2|    -.00910         .00170     -.15  .8782     -.12558    .10737
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Abowd and Farber Model

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y
Log likelihood function      -207.69977
Restricted log likelihood    -268.42420
Chi squared [   3 d.f.]       121.44886
Significance level               .00000
McFadden Pseudo R-squared      .2262256
Estimation based on N =    500, K =   4
Inf.Cr.AIC  =  423.400 AIC/N =     .847
Model estimated: Jun 16, 2011, 10:29:40
Partial Observability Model
Hosmer-Lemeshow chi-squared = *********
P-value=  .00000 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard             Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .06042         .27377      .22  .8253     -.47615    .59699
      X1|    1.05877***      .25029     4.23  .0000      .56821   1.54933
Constant|    -.14912         .22029     -.68  .4985     -.58089    .28265
      X2|     .93298***      .19656     4.75  .0000      .54774   1.31823
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
       Y|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|     .17887***      .02975     5.75  .0000      .11789    .23985
      X2|     .19756***     -.03534     6.31  .0000      .13618    .25893
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E33.7.2 Technical Details

The log likelihood for Poirier's variant of the partial observability model is

$$\text{Log } L = \Sigma_{y=1} \log \Phi_2 (\boldsymbol{\beta}_1'\mathbf{x}_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho) + \Sigma_{y=0} \log [1 - \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho)].$$

The log likelihood for Abowd and Farber's variant of the partial observability model is derived from Poirier's by setting $\rho = 0$. The bivariate CDF then factors into the product of two univariate CDFs. The derivatives of this function are given above in Section E33.2.9. For the Poirier and Abowd/Farber models, the conditional mean function is

$$E[y \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2 (\boldsymbol{\beta}_1'\mathbf{x}_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho)$$

Using the results from Section E33.2.9 once again, it follows that the marginal effects are

$$\boldsymbol{\delta} = \partial E[y \mid \mathbf{x}_1, \mathbf{x}_2] / \partial \mathbf{x}$$

$$= g_1 \boldsymbol{\gamma}_1 + g_2 \boldsymbol{\gamma}_2$$

where

$\mathbf{x}$ = the union of $\mathbf{x}_1$ and $\mathbf{x}_2$

$\boldsymbol{\gamma}_m = \boldsymbol{\beta}_m$ augmented with zeros to correspond to $\mathbf{x}$

$$w_m = \boldsymbol{\gamma}_m' \mathbf{x},\ m = 1,2$$

$$a_1 = (w_2 - \rho w_1)/(1 - \rho^2)^{1/2}$$

$$a_2 = (w_1 - \rho w_2)/(1 - \rho^2)^{1/2}$$

$$g_1 = \phi(w_1)\Phi[(w_2 - \rho w_1)/(1 - \rho^2)^{1/2}] = \phi(w_1)\Phi(a_1)$$

$$g_2 = \phi(w_2)\Phi[(w_1 - \rho w_2)/(1 - \rho^2)^{1/2}] = \phi(w_2)\Phi(a_2)$$

The Abowd and Farber case is produced by setting $\rho = 0$. To compute standard errors for the marginal effects, we use the delta method. The necessary derivatives are as follows: We will require

$$\phi'(w_1) = -w_1\phi(w_1) \text{ and likewise for } w_2.$$

Then,

$$\frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\gamma}_1} = g_1\mathbf{I} + \left(\phi'(w_1)\Phi(a_1) + \frac{-\rho\phi(w_1)\phi(a_1)}{\sqrt{1-\rho^2}} \right)\boldsymbol{\gamma}_1\mathbf{x}' + \phi(w_2)\phi(a_2)\frac{1}{\sqrt{1-\rho^2}}\boldsymbol{\gamma}_2\mathbf{x}'$$

$$\partial\boldsymbol{\delta}/\partial\rho = \boldsymbol{\gamma}_1\, \phi(w_1)\phi(a_1)\partial a_1/\partial\rho + \gamma_2\phi(w_2)\phi(a_2)\partial a_2/\partial\rho$$

$$\partial a_1/\partial\rho = (\rho a_1/(1 - \rho^2)^{1/2} - w_1) / (1 - \rho^2)^{1/2}$$

The remaining derivatives, $\partial\boldsymbol{\delta}/\partial\boldsymbol{\gamma}_2$ and $\partial a_2/\partial\rho$ are obtained by reversing subscripts in the preceding.

For the Meng and Schmidt model, the log likelihood is

$$\begin{aligned}\text{Log } L \;=\;& \Sigma_{y=1,y2=1} \log \Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_1,\boldsymbol{\beta}_2'\mathbf{x}_2,\rho] && \text{(both variables observed)}\\ &+ \Sigma_{y1=1,y2=0} \log \Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_1,-\boldsymbol{\beta}_2'\mathbf{x}_2,-\rho] && \text{(both variables observed)}\\ &+ \Sigma_{y1=0} \log [1-\Phi(\boldsymbol{\beta}_1'\mathbf{x}_1)] && \text{(only } y_1 \text{ observed)}\end{aligned}$$

(Note that save for a reversal of subscripts and a minor change in interpretation, the Meng and Schmidt log likelihood is the same as that for the bivariate probit with sample selection – in fact, the models are identical.) The various first and second derivatives can be obtained from the terms given earlier. Since there are two outcomes and no natural conditional mean function, marginal affects are not computed for the Meng and Schmidt model.

For all of these models, the BFGS method is used for estimation. BHHH and Newton (based on the BHHH estimator of the Hessian) will probably perform very poorly. We have also found that iteration with the Hessian as opposed to the BHHH estimator for the bivariate probit models performs, likewise, very poorly. The choice based sampling estimator uses the Hessian in order to construct the adjusted covariance matrix.

E33.8 Panel Data Bivariate Probit Models

The four bivariate probit models, bivariate probit, bivariate probit with selection, Poirier's partial observability and Abowd's partial observability model have all been extended to the random parameters form of the panel data models. (The fixed effects and latent class models are not available.) Use of the random parameters formulation is described in detail in Chapter R24. We will only sketch the extension here. The commands for the models are as follows, where [...] indicates an optional part of the specification:

```
        BIVARIATE   ; Lhs  = y1, y2              ? Bivariate probit
                    ; Rh1  = Rhs for equation 1
                    ; Rh2  = Rhs for equation 2
                    [ ; Selection ]              ? Partial observability

or      PROBIT      ; Lhs  = y                   ? Probit model
                    ; Rh1  = Rhs for equation 1
                    ; Rh2  = Rhs for equation 2 ? Partial observability (Poirier)
                    [ ; Selection ]              ? Abowd and Farber

Then,               ; RPM [ = list for heterogeneity in the mean ]
                    ; Pds  = panel specification ? Optional if cross section
                    [ ; Pts = number of replications ]
                    [ ; Halton and other controls for the estimation ]
                    ; Fcn  = designation of random parameters $
```

For the random parameters specification, use

; name (distribution) distribution = *n*, *u*, *t*, *l*, *c* for the first equation

or **; name [distribution]** for the second equation.

Note that random parameters in the second equation are designated by square brackets rather than parentheses. This is necessary because the same variables can appear in both equations. Two other specifications should be useful

; Cor allows the random parameters to be correlated.
; AR1 allows the random terms to evolve according to an AR(1) process rather than be time invariant.

The two equation random parameters save the matrices *b* and *varb* and the scalar logl after estimation. No other variables, partial effects, etc. are provided internally to the command. But, you can use the estimation results directly in the **SIMULATION**, **PARTIAL EFFECTS** commands, and so on. An example appears after the results of the simulation below.

E33.8.1 Application

To demonstrate this model, we will fit a true random effects model for a bivariate probit outcome. Each equation has its own random effect, and the two are correlated. The model structure is

$$z_{it1} = \boldsymbol{\beta}_1'\mathbf{x}_{it1} + \varepsilon_{it1} + u_{i1},\quad y_{it1} = 1 \text{ if } z_{it1} > 0,\ y_{it1} = 0 \text{ otherwise,}$$

$$z_{it2} = \boldsymbol{\beta}_2'\mathbf{x}_{it2} + \varepsilon_{it2} + u_{i2},\quad y_{it2} = 1 \text{ if } z_{it2} > 0,\ y_{it2} = 0 \text{ otherwise,}$$

$$[\varepsilon_{it1},\varepsilon_{it2}] \sim \text{Bivariate normal (BVN) } [0,0,1,1,\rho],\ -1 < \rho < 1,$$

$$[u_{i1},u_{i2}] \sim \text{Bivariate normal (BVN) } [0,0,1,1,\theta],\ -1 < \theta < 1,$$

Individual observations on y_1 and y_2 are available for all i. Note, in the structure, the idiosyncratic ε_{itj} creates the bivariate probit model, whereas the time invariant common effects, u_{ij} create the random effects (random constants) model. Thus, there are two sources of correlation across the equations, the correlation between the unique disturbances, ρ, and the correlation between the time invariant disturbances, θ. The data are generated artificially according to the assumptions of the model.

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-200 $
CREATE      ; x1 = Rnn(0,1) ; x2 = Rnn(0,1) ; x3 = Rnn(0,1) $
MATRIX      ; u1i = Rndm(20) ; u2i = .5* Rndm(20) + .5* u1i $
CREATE      ; i  = Trn(10,0) ; u1 = u1i(i) ; u2 = u2i(i) $
CREATE      ; e1 = Rnn(0,1) ; e2 = .7*Rnn(0,1) + .3*e1 $
CREATE      ; y1 = (x1+e1 + u1) > 0
            ; y2 = (x2+x3+e2+u2) > 0  ; y12 = y1*y2 $
BIVARIATE   ; Lhs = y1,y2 ; Rh1 = one,x1 ; Rh2 = one,x2,x3
            ; RPM ; Pds = 10 ; Pts = 25 ; Cor ; Halton
            ; Fcn = one(n), one[n] $
PROBIT      ; Lhs = y12 ; Rh1 = one,x1 ; Rh2 = one,x2,x3
            ; RPM ; Pds = 10 ; Pts = 25 ; Cor ; Halton
            ; Fcn  = one(n), one[n] ; Selection $
PROBIT      ; Lhs  = y12 ; Rh1 = one,x1 ; Rh2 = one,x2,x3
            ; RPM ; Pds = 10 ; Pts = 25 ; Cor ; Halton
            ; Fcn  = one(n), one[n] $
```

Note that by construction, most of the cross equation correlation comes from the random effects, not the disturbances. The second model is the Abowd/Farber version of the partial observability model. The Poirier model is not estimable for this setup. It is easy to see why. The correlations in the Poirier model are overspecified. Indeed, with **; Cor** for the random effects, the Poirier model specifies two separate sources of cross equation correlation. This is a weakly identified model. The implication can be seen in the results below, where the estimator failed to converge for the probit model, and at the exit, the estimate of ρ was nearly -1.0. This is the signature of a weakly identified (or unidentified) model.

These are the estimates of the Meng and Schmidt model.

```
-----------------------------------------------------------------------------
Probit   Regression Start Values for Y1
Dependent variable                  Y1
Log likelihood function     -114.32973
--------+--------------------------------------------------------------------
      Y1|                Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|     .65214***      .10287     6.34  .0000      .45052    .85375
Constant|    -.12214         .09617    -1.27  .2041     -.31062    .06634
--------+--------------------------------------------------------------------
Probit   Regression Start Values for Y2
Dependent variable                  Y2
Log likelihood function      -83.99189
--------+--------------------------------------------------------------------
      Y1|                Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X2|     .96584***      .14838     6.51  .0000      .67503   1.25665
      X3|    1.00421***      .14562     6.90  .0000      .71880   1.28961
Constant|     .17104         .11176     1.53  .1259     -.04801    .39009
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Random Coefficients  BivProbt Model
Dependent variable                  Y1
Log likelihood function     -163.43468
Estimation based on N =    200, K =   9
Inf.Cr.AIC  =  344.869 AIC/N =    1.724
Sample is 10 pds and     20 individuals
Bivariate Probit model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
      Y1|                Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    X1_1|    1.08374***      .19408     5.58  .0000      .70335   1.46412
    X2_2|    1.18264***      .22213     5.32  .0000      .74727   1.61800
    X3_2|    1.18893***      .18946     6.28  .0000      .81758   1.56027
        |Means for random parameters
   ONE_1|    -.05021         .12427     -.40  .6862     -.29377    .19335
   ONE_2|     .27827*        .15481     1.80  .0723     -.02514    .58169
```

```
        |Diagonal elements of Cholesky matrix
   ONE_1|    1.08131***      .17778     6.08  .0000      .73288   1.42975
   ONE_2|     .42491***      .15811     2.69  .0072      .11503    .73480
        |Below diagonal elements of Cholesky matrix
lONE_ONE|    -.45867**       .17845    -2.57  .0102     -.80842   -.10892
        |Unconditional cross equation correlation
lONE_ONE|    -.17471         .17798     -.98  .3263     -.52355    .17413
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------

Implied covariance matrix of random parameters
Var_Beta|             1             2
--------+----------------------------
       1|       1.16924      -.495965
       2|      -.495965       .390927

Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       1.08131
       2|       .625242

Implied correlation matrix of random parameters
Cor_Beta|             1             2
--------+----------------------------
       1|       1.00000      -.733586
       2|      -.733586       1.00000
```

These are the estimates of the Abowd and Farber model.

```
-----------------------------------------------------------------------
Probit   Regression Start Values for Y12
Dependent variable                 Y12
Log likelihood function      -103.81770
--------+-------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
      X1|     .52842***      .10360     5.10  .0000      .32537    .73147
Constant|    -.66498***      .10303    -6.45  .0000     -.86692   -.46304
--------+-------------------------------------------------------------
Probit   Regression Start Values for Y12
Dependent variable                 Y12
Log likelihood function      -102.69669
--------+-------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
      X2|     .50336***      .11606     4.34  .0000      .27588    .73084
      X3|     .38430***      .11126     3.45  .0006      .16622    .60237
Constant|    -.64606***      .10368    -6.23  .0000     -.84927   -.44286
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Coefficients  PrshlObs Model
Dependent variable                  Y12
Log likelihood function       -72.83435
Restricted log likelihood    -102.69669
Chi squared [   3 d.f.]        59.72467
Significance level               .00000
McFadden Pseudo R-squared      .2907819
Estimation based on N =    200, K =   8
Inf.Cr.AIC  =  161.669 AIC/N =     .808
Sample is 10 pds and     20 individuals
Partial observability probit model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    X1_1|    1.09511***      .23019     4.76  .0000      .64394   1.54629
    X2_2|    2.26279***      .79573     2.84  .0045      .70319   3.82239
    X3_2|    1.90015***      .70892     2.68  .0074      .51070   3.28960
        |Means for random parameters
   ONE_1|     .09219         .22240      .41  .6785    -.34370     .52809
   ONE_2|    -.06872         .36077     -.19  .8489    -.77581     .63837
        |Diagonal elements of Cholesky matrix
   ONE_1|     .59436**       .23215     2.56  .0105      .13935   1.04937
   ONE_2|    1.98257***      .73799     2.69  .0072      .53614   3.42900
        |Below diagonal elements of Cholesky matrix
lONE_ONE|    -.91612**       .41168    -2.23  .0261   -1.72299   -.10925
        |Unconditional cross equation correlation
lONE_ONE|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

```
Implied covariance matrix of random parameters
Var_Beta|            1             2
--------+----------------------------
       1|       .353265      -.544507
       2|      -.544507       4.76987

Implied standard deviations of random parameters
S.D_Beta|            1
--------+--------------
       1|       .594361
       2|       2.18400

Implied correlation matrix of random parameters
Cor_Beta|            1             2
--------+----------------------------
       1|       1.00000      -.419469
       2|      -.419469       1.00000
```

These are the estimates of the Poirier model.

```
-----------------------------------------------------------------------------
Probit   Regression Start Values for Y12
Dependent variable                 Y12
Log likelihood function      -103.81770
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|      .52842***      .10360     5.10  .0000       .32537    .73147
Constant|     -.66498***      .10303    -6.45  .0000     -.86692   -.46304
-----------------------------------------------------------------------------
Probit   Regression Start Values for Y12
Dependent variable                 Y12
Log likelihood function      -102.69669
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X2|      .50336***      .11606     4.34  .0000       .27588    .73084
      X3|      .38430***      .11126     3.45  .0006       .16622    .60237
Constant|     -.64606***      .10368    -6.23  .0000     -.84927   -.44286
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
Random Coefficients  PrshlObs Model
Dependent variable                 Y12
Log likelihood function       -70.16147
Sample is 10 pds and     20 individuals
Partial observability probit model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     Y12|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    X1_1|      .95923***      .21311     4.50  .0000       .54154   1.37692
    X2_2|     1.02185***      .28212     3.62  .0003       .46890   1.57480
    X3_2|      .77643***      .23096     3.36  .0008       .32376   1.22910
        |Means for random parameters
   ONE_1|      .41477         .32108     1.29  .1964     -.21454   1.04407
   ONE_2|      .08625         .31520      .27  .7844     -.53153    .70402
        |Diagonal elements of Cholesky matrix
   ONE_1|      .42395         .28240     1.50  .1333     -.12955    .97744
   ONE_2|      .98957***      .29127     3.40  .0007       .41869   1.56044
        |Below diagonal elements of Cholesky matrix
lONE_ONE|     -.62399**       .31020    -2.01  .0443    -1.23197   -.01601
        |Unconditional cross equation correlation
lONE_ONE|     -.99693***      .01079   -92.41  .0000    -1.01808   -.97579
--------+--------------------------------------------------------------------
```

```
Implied covariance matrix of random parameters
Var_Beta|             1             2
--------+----------------------------
       1|       .179731      -.264539
       2|      -.264539       1.36861

Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       .423947
       2|       1.16988

Implied correlation matrix of random parameters
Cor_Beta|             1             2
--------+----------------------------
       1|       1.00000      -.533382
       2|      -.533382       1.00000
```

E33.8.2 Simulation and Partial Effects

This is the model estimated at the beginning of the previous section.

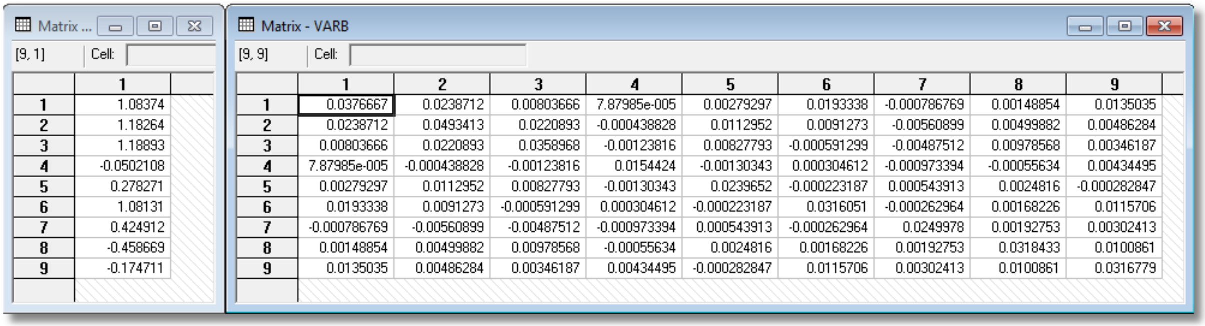
$$y1^* = a1 + b11\,x1 + u1 + e1$$

$$y2^* = a2 + b22\,x2 + b23\,x3 + u2 + e2.$$

The random effects, $u1$ and $u2$, are time invariant – the same value appears in each of the 10 periods of the data. The model command is

BIVARIATE ; Lhs = y1,y2
; Rh1 = one,x1 ; Rh2 = one,x2,x3
; RPM ; Pds = 10 ; Pts = 25 ; Cor ; Halton
; Fcn = one(n), one[n] $

```
-----------------------------------------------------------------------------
Random Coefficients  BivProbt Model
Bivariate Probit model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
      Y1|                 Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    X1_1|    1.08374***      .19408     5.58  .0000      .70335   1.46412
    X2_2|    1.18264***      .22213     5.32  .0000      .74727   1.61800
    X3_2|    1.18893***      .18946     6.28  .0000      .81758   1.56027
        |Means for random parameters
   ONE_1|    -.05021         .12427     -.40  .6862    -.29377    .19335
   ONE_2|     .27827*        .15481     1.80  .0723    -.02514    .58169
        |Diagonal elements of Cholesky matrix
   ONE_1|    1.08131***      .17778     6.08  .0000      .73288   1.42975
   ONE_2|     .42491***      .15811     2.69  .0072      .11503    .73480
        |Below diagonal elements of Cholesky matrix
lONE_ONE|    -.45867**       .17845    -2.57  .0102    -.80842   -.10892
        |Unconditional cross equation correlation
lONE_ONE|    -.17471         .17798     -.98  .3263    -.52355    .17413
--------+--------------------------------------------------------------------
```


Matrix ... [9, 1] Cell:

	1
1	1.08374
2	1.18264
3	1.18893
4	-0.0502108
5	0.278271
6	1.08131
7	0.424912
8	-0.458669
9	-0.174711

Matrix - VARB [9, 9] Cell:

	1	2	3	4	5	6	7	8	9
1	0.0376667	0.0238712	0.00803666	7.87985e-005	0.00279297	0.0193338	-0.000786769	0.00148854	0.0135035
2	0.0238712	0.0493413	0.0220893	-0.000438828	0.0112952	0.0091273	-0.00560899	0.00499882	0.00486284
3	0.00803666	0.0220893	0.0358968	-0.00123816	0.00827793	-0.000591299	-0.00487512	0.00978568	0.00346187
4	7.87985e-005	-0.000438828	-0.00123816	0.0154424	-0.00130343	0.000304612	-0.000973394	-0.00055634	0.00434495
5	0.00279297	0.0112952	0.00827793	-0.00130343	0.0239652	-0.000223187	0.000543913	0.0024816	-0.000282847
6	0.0193338	0.0091273	-0.000591299	0.000304612	-0.000223187	0.0316051	-0.000262964	0.00168226	0.0115706
7	-0.000786769	-0.00560899	-0.00487512	-0.000973394	0.000543913	-0.000262964	0.0249978	0.00192753	0.00302413
8	0.00148854	0.00499882	0.00978568	-0.00055634	0.0024816	0.00168226	0.00192753	0.0318433	0.0100861
9	0.0135035	0.00486284	0.00346187	0.00434495	-0.000282847	0.0115706	0.00302413	0.0100861	0.0316779

Figure E33.3 Matrix Results

The estimator does not support predictions or partial effects. But, we can use the template **SIMULATE** and **PARTIAL EFFECTS** programs to create our own by supplying our function and estimates.. We will use the model exactly as shown in the results, with labels for the estimates in order of their appearance: **b11,b22,b23,a1,a2,c11,c22,c21,ro.** For purposes of the exercise, we will examine the bivariate normal probability P(y1=1,y2=1). With all the parts in place, other functions, such as the conditional means, can be examined by making minor changes in the function definition. For example, in the program below, partial effects are obtained simply by changing the command to **PARTIALS** and changing **; Scenario: to ; Effects: x1**.

? Create time invariant random effects. Used to create correlated u1 and u2

```
MATRIX      ; mv1 = Rndm(20,1) ; mv2 = Rndm(20,1) $
CREATE      ; index = Trn(10,0) $
CREATE      ; v1 = mv1(index) ; v2 = mv2(index) $
```

? Simulate the joint probability and examine its behavior as x1 varies

```
SIMULATE    ; Labels = b11,b22,b23,a1,a2,c11,c22,c21,ro
            ; Parameters = b
            ; Covariance = varb
            ; Function = xb1 = a1+b11*x1+c11*v1 |
                         xb2 = a2+b22*x2+b23*x3+c21*v1+c22*v2 |
                         Bvn(xb1,xb2,ro)
            ; Scenario: & x1 = -3(.2)3 ; Plot $
```

```
----------------------------------------------------------------------
Model Simulation Analysis for User Specified Function
----------------------------------------------------------------------
Simulations are computed by average over sample observations
----------------------------------------------------------------------
User Function       Function   Standard
(Delta method)       Value      Error      |t|  95% Confidence Interval
----------------------------------------------------------------------
Avrg. Function       .23829     .02576    9.25      .18780       .28878
X1       = -3.00     .00645     .00464    1.39    -.00266       .01555
X1       = -2.80     .00870     .00567    1.54    -.00240       .01981
(rows omitted)
X1       =  2.80     .51118     .03121   16.38      .45001       .57235
X1       =  3.00     .51513     .03049   16.90      .45538       .57488
```

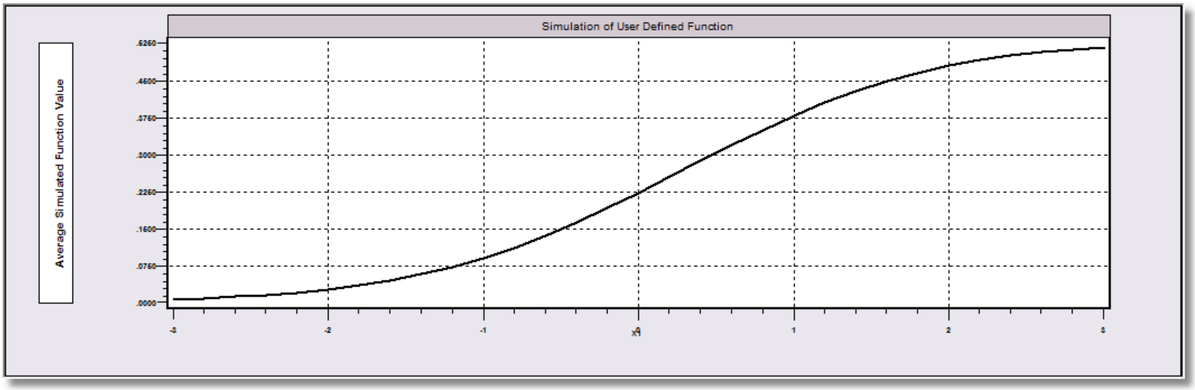


Figure E33.4 Simulation of Estimated Model

E33.9 Simultaneous Equations Models

Sections E33.5 and E33.6 suggested some results for simultaneous equations models involving binary variables. For present purposes, the important feature is the appearance of the actual binary outcomes in the structural equations. A number of models have been developed that involve not the outcome variables, but the underlying, unobserved continuous utilities. A familiar source for some of these is Maddala (1983, Chapter 8). His structural model is

$$y_1^* = \gamma_1 y_2^* + \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1,\ y_1 = 1(y_1^* > 0) \text{ in the probit formulations,}$$

$$y_2^* = \gamma_2 y_1^* + \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2,\ y2 = 1(y2^* > 0) \text{ in the probit formulations,}$$

$$[\varepsilon_1,\varepsilon_2] \sim \text{BVN}[(0,0), \sigma_1^2, \sigma_2^2, \rho] \ (\rho \text{ is the correlation, } \sigma_j \text{ is a standard deviation).}$$

Maddala's Section 8.8 lists numerous permutations of the model based on different censoring mechanisms, that is, whether y_j^* or y_j is observed, or perhaps somewhere between, a censored version, $y_j^+ = y_j y_{j^*}$. A crucial element of this model is that the underlying structure involves the latent, uncensored variables. One might easily imagine that the observed data, rather than the latent variables appear in the model. Consider Greene's (1998) model (based on Burnett (1997)) of whether a liberal arts college offers a gender economics course. Its structure has the form:

$$\text{Prob[Women's studies program]} = F_1(\boldsymbol{\beta}_1'\mathbf{x}_1)$$

$$\text{Prob[Gender economics course]} = F_2(\boldsymbol{\beta}_2'\mathbf{x}_2 + \gamma\text{Women's studies program).}$$

This is a simultaneous equations (albeit a recursive one), but with the crucial difference that the observed binary variable from the first equation, not the latent continuous variable, appears in the second equation. We reconsider this model below. Estimation of the two formulations must be handled differently. The second case was developed in Sections E33.5 and E33.6.

E33.9.1 Maddala's Models

The fundamental aspect of Maddala's models (which may or may not be appropriate in a given application) is that there exists a reduced form

$$y_1^* = \pi_1'\mathbf{x} + v_1,$$

$$y_2^* = \pi_2'\mathbf{x} + v_2,$$

$$[v_1,v_2] \sim \text{BVN}[(0,0), \theta_1^2, \theta_2^2, \tau] \text{ (τ is the correlation, θ_j is a standard deviation)},$$

where $\mathbf{x}$ is the union of the variables in $\mathbf{x}_1$ and $\mathbf{x}_2$. The observation mechanism that translates the latent variables into the observed ones thus operates on the reduced form as well as on the structure. This turns out to be a (very) large convenience. We consider two of the models given in Maddala (1983). The others can be handled similarly. We note, most of Maddala's formulations of these estimation problems, although correct, are probably more complicated than necessary. The manipulation of variance parameters that are unidentified, rather than normalization of them to one at the outset, unnecessarily complicates some of the analyses as well. Also, in some cases, the Murphy-Topel result for correcting covariance matrices for two step estimators can be used to some advantage to simplify Maddala's results. (Interestingly, Maddala (1983) gives, on page 243 the basic result to generate Murphy and Topel's (1985) formulas, but he does not take full advantage of it in his derivations.) Also, it is important to note that given the way these models are formulated, all can be estimated by full information maximum likelihood. Imposition of the overidentifying restrictions will complicate the development, which is a reasonable motivation for the two step methods considered here. In these cases, the models are estimated using the method of moments and two step estimators sometimes using single equation MLE at the first step.

E33.9.2 Model 3: y_1* Observed Directly, y_2* Observed as Binary y_2

The steps in the estimation involve fitting the two equations of the reduced form, by least squares for the first equation and univariate probit for the second. Fitted values for the latent variables from the two equations are then inserted into the structural forms. The structures are then fit by least squares and maximum likelihood probit, respectively, and the asymptotic covariance matrices are corrected after estimation. This *LIMDEP* routine estimates Maddala's Model 3, page 245. A few algebraic errors are corrected in what follows. Also, Maddala carries σ_2 around his derivation, but σ_2 must equal one. We just impose the normalization outright. The implied equation system is

$$y_1^* = \gamma_1\, y_2^* + \boldsymbol{\beta}_1'\mathbf{x}_1 + u_1,\;\; y_1 = y_1^*,$$

$$y_2^* = \gamma_2\, y_1^* + \boldsymbol{\beta}_2'\mathbf{x}_2 + u_2,\;\; y_2 = \mathbf{1}(y_2^* > 0).$$

The first variable, y_1* is directly observed, but only the sign of y_2* is observed. The variance of u_1 is σ_1^2 squared. Since y_2* is not observed, save for sign, σ_2^2 squared is normalized to 1.0. The correlation is ρ_{12}.

In the following, $\mathbf{x}$ is the union of $\mathbf{x}_1$ and $\mathbf{x}_2$. The reduced form for the structural system is

$$y_1^* = \boldsymbol{\pi}_1'\mathbf{x} + v_1,\ \mathrm{Var}[v_1] = \theta_1^2 = (\sigma_1^2 + \gamma_1^2 + 2\rho_{12}\gamma_1\sigma_1) / (1 - \gamma_1\gamma_2)^2$$

$$y_2^* = \boldsymbol{\pi}_2'\mathbf{x} + v_2,\ \mathrm{Var}[v_2] = \theta_2^2 = (\ 1\ + \gamma_2\sigma_1^2 + 2\rho\gamma_2\ ) /\ (1 - \gamma_1\gamma_2)^2$$

We note before proceeding that there is a huge gap in Maddala's derivation. The parameter ρ_{12} is crucial in the derivation, and Maddala offhandedly claims that σ_{12}/σ_2 is estimable. In fact, σ_2 is not estimable – it must be normalized to 1.0. That leaves σ_{12} which is the needed calculation, since $\rho_{12} = \sigma_{12}/\sigma_1$. But, in his derivation, he never presents the estimator for σ_{12} and it is far from obvious where one should get it. We see two possibilities: (1) Since σ_1, γ_1 and θ_1^2 are estimated in the OLS regression of the first reduced form, the value of ρ_{12} that equates the left and right hand sides can be used. (2) Since the variance in a probit equation must equal one, we can find the ρ_{12} that makes θ_2^2 equal one. Both of these candidates are method of moments estimators that can produce nonsense estimates of ρ_{12}. Based on his specification, there is no other way out of this quandary.

Only this setup is needed for use of this program. The remainder is generic. This defines the exogenous variables in the structural equations. For convenience, we also make copies of the two left hand side variables.

```
NAMELIST    ; x1 = ... ; x2 = ... $
CREATE      ; y1 = ... ; y2 = ... $
```

The procedure will now be generic. This can be called by name, supplying only the names of the left and right hand side variables.

```
PROC = Model3(y1,y2,x1,x2) $
NAMELIST    ; xm3 = OR(x1,x2) $ Forms the union of the two right hand sides.
```

This is the first reduced form regression. Regress y_1 on $\mathbf{x}$. This estimates $\boldsymbol{\pi}_1$ and θ_1^2. The fitted values are $\boldsymbol{\pi}_1'\mathbf{x}$. Retrieve the reduced form variance, s^2; this is the estimate of θ_1^2.

```
REGRESS     ; Lhs = y1 ; Rhs = xm3 ; Keep = p1x $
CALC        ; t1sq = ssqrd $
```

The second reduced form is the probit equation to estimate $\boldsymbol{\pi}_2$. We use these to compute $\boldsymbol{\pi}_2'\mathbf{x}$.

```
PROBIT      ; Lhs = y2 ; Rhs = xm3 $
CREATE      ; p2x = xm3'b $
MATRIX      ; v0 = varb $
```

The fitted values are used on the right hand sides of the structural equation estimators. Estimate the first structural equation by least squares regression of y_1 on $(\boldsymbol{\pi}_2'\mathbf{x})$ and $\mathbf{x}_1$. Pick up the estimates of γ_1 and σ_1^2 and the full coefficient vector $\alpha_1 = (\gamma_1,\boldsymbol{\beta}_2)$. Estimate the second structural equation by probit estimation of y_2 on $\boldsymbol{\pi}_1'\mathbf{x}$ and x_2. Pick up the estimate of γ_2 and $\boldsymbol{\alpha}_2 = (\gamma_2,\boldsymbol{\beta}_2)$

```
NAMELIST    ; z1 = p2x,x1 ; z2 = p1x,x2 $
REGRESS     ; Lhs = y1 ; Rhs = z1 $  First structural equation
MATRIX      ; alpha1 = b ; vols = varb $
CALC        ; gamma1 = b(1) ; sigmasq1 = ssqrd ; sigma1 = s $
PROBIT      ; Lhs = y2 ; Rhs = z2 $  Second structural equation
MATRIX      ; alpha2 = b ; vprobit = varb  $
CALC        ; gamma2 = b(1) $
```

Use the estimates of θ_1, γ_1, γ_2, and σ_1^2 to estimate ρ_{12}. We first compute the possibly problematic method of moments estimator of ρ_{12}. Then compute the two scalars used in the computation, $c = \sigma_1^2 - 2\sigma_1\gamma_1\rho_{12}$ and $d = \gamma_2^2\sigma_1^2 - 2\sigma_1\gamma_2\rho_{12}$.

```
CALC        ; rho12 = ((1-gamma1*gamma2)^2*t1sq - gamma1^2 - sigmasq1)
                          / (2*gamma1*sigma1)
            ; c = sigmasq1 - 2* sigma1*gamma1*rho12
            ; d = gamma2^2 * sigmasq1 - 2*sigma1*gamma2*rho12 $
```

Now compute the adjusted asymptotic covariance matrices and correct the results. He carries σ_2 through all the computations, but σ_2 must equal one. Finish the covariance matrices and show the results.

```
MATRIX      ; z1z1i  =<z1'z1> ; va1 = c*z1z1i
                + {gamma1^2} * z1z1i * z1'XM3 * v0 * xm3'z1 * z1z1i
            ; z2xvi  = z2'xm3 * v0 * xm3'z2 ; z2xvi  = <z2xvi>
            ; va2 = z2xvi + d*z2xvi * z2'xm3 * v0 * <xm3'xm3> * v0 * xm3'z2 * z2xvi
            ; Stat(alpha1,va1,z1) ; Stat (alpha1,vols,z1)
            ; Stat(alpha2,va2,z2) ; Stat(alpha2,vprobit,z2) $
ENDPROCEDURE $
```

To use the procedure, the command is where the arguments are the variables and namelists that you have defined earlier.

```
EXECUTE     ; Proc = Model3(y1, y2, x1, x2) $
```

E33.9.3 Model 6: Both y_1^* and y_2^* Observed as Binary y_1 and y_2

This model involves only the binary variables. Neither structural variance is identified, so the only variance parameter the model produces is the correlation coefficient, ρ. The setup for the two sets of variables in the system is the same as above in the Model 3 procedure.

```
PROCEDURE  = Model6(y1,y2,x1,x2) $
NAMELIST    ; xm6 = OR(x1,x2) $
```

Estimate the reduced form probits and save the coefficients

```
PROBIT      ; Lhs = y1 ; Rhs = xm6 $
MATRIX      ; pi1 = b ; v1 = varb  $
PROBIT      ; Lhs = y2 ; Rhs = xm6 $
MATRIX      ; pi2 = b ; v2 = varb  $
CREATE      ; y1f = x'pi1 ; y2f = xm6'pi2 $
```

Estimate the structural probits using the fitted values from above.

```
NAMELIST    ; z1 = y2f,x1 ; z2 = y1f,x2 $
PROBIT      ; Lhs = y1 ; Rhs = z1 $
MATRIX      ; alpha1 = b ; vp1 = varb $
PROBIT      ; Lhs = y2 ; Rhs = z2 $
MATRIX      ; alpha2 = b ; vp2 = varb $
```

Calculate the covariance matrices, first for α_1 then, symmetrically, for α_2

```
CREATE      ; q1= z1'alpha1 ; q2 = xm6'pi2
            ; a1 = Lmp(q1)/Phi(q1) ; a2 = Lmp(q2)/Phi(q2)
            ; capa1 = N01(q1) * a1
            ; u1 = y1 - Phi(q1) ; u2 = y2 - Phi(q2) ; v = a1 * a2 * u1 * u2 $
MATRIX      ; w1 = z1'[capa1]z1 ; w2 = v1 ; w3 = alpha1(1) * z1'[capa1]xm6
            ; w4 = xm6'[v]z1 ; w324 = w3 * w2 * w4  ; w323 = w3 * w2 * w3'
            ; va1 = w1 - w324 - w324' + w323 ; va1 = <w1> * va1 * <w1> $
CREATE      ; q3 = z2'alpha2 ; q4 = xm6'pi1
            ; a3 = Lmp(q3)/Phi(q3) ; a4 = Lmp(q4)/Phi(q4)
            ; capa3 = N01(q3) * a3
            ; u3 = y2 - Phi(q3) ; u4 = y1 - Phi(q4) ; v = a3 * a4 * u3 * u4 $
MATRIX      ; w1 = z2'[capa3]z2 ; w2 = v2 ; w3 = alpha2(1) * z2'[capa3]xm6
            ; w4 = x '[v]z2  ; w324 = w3 * w2 * w4 ; w323 = w3 * w2 * w3'
            ; va2 =  w1 - w324 - w324' + w323 ; va2 = <w1> * va2 * <w1> $
MATRIX      ; Stat(alpha1,va1,z1) ; Stat(alpha1,vp1,z1)
            ; Stat(alpha2,va2,z2) ; Stat(alpha2,vp2,z2) $
```

To use the procedure, the command is where the arguments are the variables and namelists that you have defined earlier.

```
EXECUTE     ; Proc = Model6(y1, y2, x1, x2) $
```

E33.10 Multivariate Probit Model

The multivariate probit model is the extension to M equations of the bivariate probit model

$$y_{im}^* = \boldsymbol{\beta}_m'\mathbf{x}_{im} + \varepsilon_{im},\ m = 1,\ldots,M$$

$$y_{im} = 1 \text{ if } y_{im}^* > 0, \text{ and 0 otherwise.}$$

$$\varepsilon_{im},\ m = 1,\ldots,M \sim \text{MVN}\,[\mathbf{0},\mathbf{R}]$$

where R is the correlation matrix. Each individual equation is a standard probit model. This generalizes the bivariate probit model for up to M = 20 equations. Specify the model with the same command structure as the SURE model, using the command **MPROBIT**,

MPROBIT ; Lhs = y1,y2,...,ym (list of up to 20 variables)
; Eq1 = list of Rhs variables in the first equation
; Eq2 = list of Rhs variables in the second equation
...
; EqM = list of Rhs variables for Mth equation $

The data for this model must be individual, not proportions and not frequencies. You may use

; Wts = name

as usual. Other options specific for this model in addition to the standard output options are

; Prob = name

which requests the estimator to save the predicted probability for the observed joint outcome, and

; Utility = name

where '**name**' is an existing *namelist* to save the estimated utilities, $\mathbf{X}_m\boldsymbol{\beta}_m$. Restrictions can be imposed with

; Rst = list
and **; CML: specification for constraints**

Note that either of these can be used to specify the correlation matrix. The list for **; Rst** includes the $M(M\text{-}1)/2$ below diagonal elements of **R**. You can use this to force correlations to equal each other, or zero, or other values.

E33.10.1 Other Options

Standard Model Specifications for the Multivariate Probit Model

This is the full list of general that are applicable to this model estimator.

Controlling Output from Model Commands

; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Prob = name saves probabilities as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E33.10.2 Retrievable Results

This model keeps the following retrievable results:

Matrices:	*b*	= estimate of $(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2',\ldots,\boldsymbol{\beta}_M')'$ = vector of slopes only
	varb	= asymptotic covariance matrix
	omega	= M×M correlation matrix of disturbances
Scalars:	*kreg*	= number of parameters in model
	nreg	= number of observations
	logl	= log likelihood function
Variables:	*logl_obs*	= individual contribution to log likelihood
Last Model:	None	
Last Function:	None	

E33.10.3 Marginal Effects

You can obtain marginal effects for this model of the following form: The expected value of y_1 given that all other ys equal one is

$$E[y_1|y_2=1,...,y_M=1] = \text{Prob}(y_1=1,...,y_M=1)/\text{Prob}(y_2=1,...,y_M=1) = P_{1...M} / P_{2...M} = E_1.$$

The derivatives of this function are constructed as follows: Let $\mathbf{x}$ equal the union of all of the regressors that appear in the model, and let $\boldsymbol{\gamma}_m$ be such that $\mathbf{z}_m = \mathbf{x}'\boldsymbol{\gamma}_m = \boldsymbol{\beta}_m'\mathbf{x}_m$. ($\boldsymbol{\gamma}_m$ will usually have some zeros in it unless all regressors appear in all equations.) Then,

$$\frac{\partial E_1}{\partial \mathbf{x}} = \sum_{m=1}^{M}\left(\frac{1}{P_{2...M}}\frac{\partial P_{1...M}}{\partial \mathbf{z}_m}\right)\boldsymbol{\gamma}_m - E_1\sum_{m=2}^{M}\left(\frac{1}{P_{2...M}}\frac{\partial P_{2...M}}{\partial \mathbf{z}_m}\right)\boldsymbol{\gamma}_m$$

The relevant parts of this combination of the coefficient vectors are then extracted and reported for the specific equations. Standard errors are obtained using the delta method, and all derivatives are approximated numerically. All effects are computed at the means of the Rhs variables. Use

; Partial Effects

to request this computation. In the display of these results, derivatives with respect to the constant term are set to zero.

Standard errors for these marginal effects cannot be computed directly. We report a bootstrapped approximation computed as follows: Let the estimated set of marginal effects be denoted **d**. This is computed using the parameter estimates from the model as given earlier. Let **V** denote the estimated asymptotic covariance matrix for the coefficient estimates. An estimate of the variance of the estimator of the marginal effects is obtained as the mean squared deviation of 50 random draws from the distribution of the underlying slope parameters. You can set the number of bootstrap replications to use with

; Nbt = number of replications.

The draws are based on the asymptotic normal distribution with mean **b** and variance **V**. (The estimated correlation parameters are taken as fixed.) Thus, the marginal effects at the data means are computed 50 additional times with these new parameters, using

$$Est.Var[d_j] = \frac{1}{50}\sum_{r=1}^{50}\left(d_{jr} - d_j\right)^2$$

Note that the sums are centered at the original estimated marginal effect, not at the means of the random draws.

E33.10.4 Technical Details

The probabilities that enter the log likelihood, its derivatives, and so on are computed using the GHK simulation method described in Section R26.8. The approximation is based on averaging R draws from a certain multivariate normal distribution, for each observation. Each observation has its own seed for the random number generator, so for identical parameter values and fixed R, the draws are repeatable. Increasing R brings greater accuracy, but at the cost of greatly increased computation time. Note, as well, that all derivatives for this model are computed numerically, so it is very time consuming. However, one useful result is that although the amount of time needed to compute the function and the derivatives varies with R and the number of equations, for a given number of equations, the number of right hand side variables has only a very minor influence on the amount of time needed to compute the model. You can control the number of draws with

; Pts = R

where R is the number you desire.

The log likelihood for this model is accumulated as the sum of the logs of the probabilities of the observed outcomes. These are computed using the following construction:

$$\text{Prob}[y_1,y_2,...,y_M|\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_M] = \text{Mvn}\ (\ \mathbf{Tz}\ ,\ \mathbf{TRT}'\ )$$

where $\mathbf{z}$ = the vector of utilities, $z_m = \boldsymbol{\beta}_m'\mathbf{x}_{im}$, $\mathbf{R}$ is the correlation matrix, and $\mathbf{T}$ is a diagonal matrix with $t_{mm} = 2y_m - 1$ (i.e. $t_{mm} = 1$ if $y_m = 1$ and $t_{mm} = -1$ if $y_m = 0$).

E33.10.5 Example

The following example demonstrates estimation of a four equation model. The correlations are actually zero, so in principle, this could be fit with individual probit equations. But, normally, that would not be known a priori.

```
SAMPLE     ; 1-200 $
CALC       ; Ran (12345) $
CREATE     ; x1 = Rnn(0,1) ; x2 = Rnn(0,1) ; x3 = Rnn(0,1) ; x4 = Rnn(0,1) $
CREATE     ; u1 = Rnn(0,1); u2 = Rnn(0,1); u3 = Rnn(0,1); u4 = Rnn(0,1) $
CREATE     ; y1 = (x1+u1) > 0
           ; y2 = (x2+x3+u2) > 0
           ; y3 = (x1+x4+u3) > 0
           ; y4 = (x2+x4+u4) > 0 $
MPROBIT    ; Lhs = y1,y2,y3,y4
           ; Eq1 = one,x1
           ; Eq2 = one,x2,x3
           ; Eq3 = one,x1,x4
           ; Eq4 = one,x2,x4
           ; Pts = 10 ; Output = 4
           ; Partial Effects $
```

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator=  10
Start values:    .12360D+00   .70740D-01   .12269D+00   .55103D-01   .63230D-01
   .13232D+00   .56338D-01   .64745D-01   .12034D+00   .65986D-01   .58720D-01
   .00000D+00   .00000D+00   .00000D+00   .00000D+00   .00000D+00   .00000D+00
1st derivs.      .84798D+01  -.76676D+02   .19921D+02  -.63304D+02  -.73836D+02
   .14141D+02  -.63575D+02  -.76630D+02   .31436D+02  -.71270D+02  -.73015D+02
   .19690D+02  -.17021D+02   .13362D+02   .13334D+02  -.30084D+02  -.36003D+02
Parameters:      .12360D+00   .70740D-01   .12269D+00   .55103D-01   .63230D-01
   .13232D+00   .56338D-01   .64745D-01   .12034D+00   .65986D-01   .58720D-01
   .00000D+00   .00000D+00   .00000D+00   .00000D+00   .00000D+00   .00000D+00
Itr  1 F=  .5272D+03 gtHg=  .2014D+03 chg.F=  .5272D+03 max|db|=  .3600D+08
Try =  0 F=  .5272D+03 Step=  .0000D+00 Slope= -.2014D+03
Try =  1 F=  .5078D+03 Step=  .1000D+00 Slope= -.1867D+03
Try =  2 F=  .3623D+03 Step=  .1369D+01 Slope= -.5825D+02
Try =  3 F=  .3391D+03 Step=  .1945D+01 Slope= -.2370D+02
Try =  4 F=  .3337D+03 Step=  .2476D+01 Slope=  .2739D+01
1st derivs.      .32766D+01   .43567D+01  -.41789D+01  -.91291D+00  -.99307D+00
  -.82940D+01  -.91181D+01  -.75208D+01  -.17834D+02  -.11640D+02  -.65515D+00
  -.50245D+00   .16244D+02  -.81459D+01  -.94190D+01   .16572D+02   .17481D+02
Parameters:      .19343D-01   .10134D+01  -.12222D+00   .83338D+00   .97099D+00
  -.41531D-01   .83795D+00   .10069D+01  -.26614D+00   .94220D+00   .95639D+00
  -.24207D+00   .20926D+00  -.16428D+00  -.16393D+00   .36986D+00   .44264D+00
Itr  2 F=  .3337D+03 gtHg=  .4136D+02 chg.F=  .1935D+03 max|db|=  .1997D+03
Try =  0 F=  .3337D+03 Step=  .0000D+00 Slope= -.4136D+02
Try =  1 F=  .4215D+03 Step=  .2476D+01 Slope=  .1999D+03
Try =  2 F=  .3261D+03 Step=  .1089D+01 Slope=  .1681D+02
Try =  3 F=  .3225D+03 Step=  .6287D+00 Slope= -.1412D+01
```

```
1st derivs.    -.14478D+01  -.23864D+00  -.22465D+01  -.20622D+01  -.22485D+01
  -.24826D+01  -.60476D+01  -.49056D+01   .50617D+01  -.18036D+01  -.81122D+01
  -.41065D+01   .39010D+01   .29559D+01   .75582D+00   .61893D+00  -.24286D+01
Parameters:    -.30461D-01   .94719D+00  -.58704D-01   .84725D+00   .98608D+00
   .84539D-01   .97655D+00   .11212D+01   .49424D-02   .11191D+01   .96634D+00
  -.23443D+00  -.37641D-01  -.40463D-01  -.20757D-01   .11795D+00   .17692D+00
Itr 20 F=  .3191D+03 gtHg=  .1340D-04 chg.F=  .4547D-11 max|db|=  .6587D-03
Try =  0 F=  .3191D+03 Step=  .0000D+00 Slope= -.4783D-04
Try =  1 F=  .3191D+03 Step=  .5017D-06 Slope= -.1911D-04
Try =  2 F=  .3191D+03 Step=  .8357D-06 Slope= -.2119D-08
1st derivs.    -.22084D-06   .29391D-07   .31594D-07   .31861D-07  -.52226D-08
   .12163D-07   .17487D-07   .47632D-08  -.30642D-07  -.21781D-07  -.13777D-07
   .39684D-07  -.60782D-07  -.34254D-07  -.68164D-08  -.40873D-08   .23301D-07
Parameters:    -.20652D-01   .95692D+00  -.38949D-01   .93476D+00   .10500D+01
   .14372D+00   .12490D+01   .14330D+01  -.33649D-01   .12912D+01   .12151D+01
  -.11798D+00  -.13321D+00  -.92436D-01   .16791D-02   .93420D-01   .22953D+00
Itr 21 F=  .3191D+03 gtHg=  .2945D-07 chg.F=  .1955D-10 max|db|=  .2424D-06
                                                              * Converged
Normal exit:  21 iterations. Status=0, F=    319.0703
Function=  .52721306618D+03, at entry,  .31907030043D+03 at exit

-----------------------------------------------------------------------------
Multivariate Probit Model:  4 equations.
Dependent variable                MVProbit
Log likelihood function         -319.07030
Estimation based on N =     200, K =  17
Inf.Cr.AIC  =  672.141 AIC/N =     3.361
Replications for simulated probs. =  10
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
MVProbit|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for Y1
Constant|     -.02065         .10695     -.19  .8469     -.23028    .18897
      X1|      .95692***      .14188     6.74  .0000      .67883   1.23501
        |Index function for Y2
Constant|     -.03895         .11830     -.33  .7420     -.27082    .19292
      X2|      .93476***      .19402     4.82  .0000      .55448   1.31504
      X3|     1.05004***      .16917     6.21  .0000      .71847   1.38160
        |Index function for Y3
Constant|      .14372         .13389     1.07  .2831     -.11871    .40614
      X1|     1.24904***      .25474     4.90  .0000      .74975   1.74832
      X4|     1.43296***      .27672     5.18  .0000      .89059   1.97533
        |Index function for Y4
Constant|     -.03365         .13220     -.25  .7991     -.29277    .22547
      X2|     1.29116***      .22791     5.67  .0000      .84447   1.73785
      X4|     1.21507***      .19898     6.11  .0000      .82507   1.60507
        |Correlation coefficients
R(01,02)|     -.11798         .16431     -.72  .4727     -.44002    .20406
R(01,03)|     -.13321         .17566     -.76  .4482     -.47750    .21108
R(02,03)|     -.09244         .17993     -.51  .6074     -.44510    .26023
R(01,04)|      .00168         .19945      .01  .9933     -.38924    .39260
R(02,04)|      .09342         .18562      .50  .6148     -.27039    .45723
R(03,04)|      .22953         .21032     1.09  .2751     -.18268    .64174
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+-----------------------------------------------+
| Partials of E[y1|other vars=1,X] wrt X        |
| Computed at the means of all RHS vars.        |
| Conditional mean is Prob[Y1      =1] given    |
| Y2       through Y4       all equal 1.000.    |
| Estimate of conditional mean =  .49399        |
+-----------------------------------------------+
+--------+---------+--------------------------------------------------+---------+
|        |Mean of  |---------- Coefficient in Equation --------------|Marginal |
|Variable|Variable | Y1      | Y2      | Y3      | Y4      |         |Effect   |
+--------+---------+---------+---------+---------+---------+---------+---------+
|ONE     | 1.000000| -.020652| -.038949|  .143717| -.033649|  .000000|  .000000|
|X1      |  .108196|  .956919|  .000000| 1.249038|  .000000|  .000000|  .419524|
|X2      |  .029540|  .000000|  .934758|  .000000| 1.291157|  .000000| -.046952|
|X3      | -.048523|  .000000| 1.050036|  .000000|  .000000|  .000000|  .013699|
|X4      | -.187948|  .000000|  .000000| 1.432961| 1.215068|  .000000| -.021750|
+--------+---------+---------+---------+---------+---------+---------+---------+

-----------------------------------------------------------------------------
Std.Errors are based on  50 bootstrap reps.
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
MVProbit|      Effect         Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for Y1
      X1|     .41952***      .05019     8.36  .0000       .32116     .51789
        |Index function for Y2
      X2|    -.04695         .08341     -.56  .5735     -.21044     .11653
      X3|     .01370         .03955      .35  .7290     -.06381     .09121
        |Index function for Y3
      X1|     .41952***      .05019     8.36  .0000       .32116     .51789
      X4|    -.02175         .09006     -.24  .8092     -.19827     .15477
        |Index function for Y4
      X2|    -.04695         .08341     -.56  .5735     -.21044     .11653
      X4|    -.02175         .09006     -.24  .8092     -.19827     .15477
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E33.10.6 Sample Selection Model

There are two modifications of the multivariate probit model built into the estimator. The first is a multivariate version of the selection model in Section E33.4. The model structure is

$$
\begin{aligned}
y_{i1}^* &= \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1},\\
y_{i2}^* &= \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \varepsilon_{i2},\\
&\ldots\\
y_{i,M-1}^* &= \boldsymbol{\beta}_{M-1}'\mathbf{x}_{i,M-1} + \varepsilon_{I,M-1},\\
y_{iM}^* &= \boldsymbol{\beta}_M'\mathbf{x}_{iM} + \varepsilon_{iM},\\
y_{im} &= 1 \text{ if } y_{im}^* > 0 \text{, and 0 otherwise.}\\
\varepsilon_{im}, m &= 1,\ldots,M \ \sim \text{MVN}\,[\mathbf{0},\mathbf{R}]
\end{aligned}
$$

$y_{i,1}, y_{i,2}, \ldots, y_{i,M\text{-}1}$ only observed when $y_{iM} = 1$.

In the same fashion as earlier, the log likelihood is built up from the laws of probability. The different terms in the likelihood function are

$$\text{Prob}(y_{iM} = 1|\mathbf{x}_{im})$$

for the nonselected case, then

$$\text{Prob}(Y_{i1} = y_{i1},\ldots,Y_{i,M\text{-}1} = y_{i,M\text{-}1}\ , y_{iM} = 1|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iM}).$$

The last equation is the selection mechanism. This produces a difference in the likelihood that is maximized (and, to some degree, in the interpretation of the model), but no essential difference in the estimation results.

This form of the model is requested by adding

; Selection

to the **MVPROBIT** command. There are no other changes in the model specification, or the data. Missing data may be coded as zeros or as missing.

E33.10.7 Sequential Selection or Attrition

A second form of the multivariate probit model accommodates exogenous attrition. In this form, the M equations would be a sequence of probit outcomes, in the form of an M period panel. The feature produced here is that the individual is present only for the first T_i of the M periods; T_i might equal M, but could be fewer. For this form of the model, the structure is exactly as above, for all M periods. However, for individual *i*, only a T_i-variate probit model applies. To request this form of the model, use

MVPROBIT ; … all as before
; Pds = the variable that provides T_i $

The remaining features of the model are, once again, all as before. A (probably obvious) restriction is that at least some individuals must be present for all M periods in order for the model to be estimable.

E34: Ordered Choice Models

E34.1 Introduction

The basic ordered choice model is based on the following specification: There is a latent regression,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ \mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0,\ \mathrm{Var}[\varepsilon_i|\mathbf{x}_i] = 1,$$

The observation mechanism results from a complete censoring of the latent dependent variable as follows:

$$\begin{aligned} y_i &= 0 \text{ if } y_i \le \mu_0,\\ &= 1 \text{ if } \mu_0 < y_i \le \mu_1,\\ &= 2 \text{ if } \mu_1 < y_i \le \mu_2,\\ &\ \ldots\\ &= J \text{ if } y_i > \mu_{J\text{-}1}. \end{aligned}$$

The latent 'preference' variable, y_i^* is not observed. The observed counterpart to y_i^* is y_i. Five stochastic specifications are provided for the basic model shown above. The *ordered probit* model based on the normal distribution was developed by Zavoina and McElvey (1975). It applies in applications such as surveys, in which the respondent expresses a preference with the above sort of ordinal ranking. The variance of ε_i is assumed to be one, since as long as y_i^*, $\boldsymbol{\beta}$, and ε_i are unobserved, no scaling of the underlying model can be deduced from the observed data. (The assumption of homoscedasticity is arguably a strong one. We will relax that assumption in Section E35.2.) Since the μs are free parameters, there is no significance to the unit distance between the set of observed values of y. They merely provide the coding. Estimates are obtained by maximum likelihood. The probabilities which enter the log likelihood function are

$$\mathrm{Prob}[y_i = j] = \mathrm{Prob}[y_i^* \text{ is in the } j\text{th range}].$$

The model may be estimated either with individual data, with $y_i = 0, 1, 2, \ldots$ or with grouped data, in which case each observation consists of a full set of J+1 proportions, $p_{0i},\ldots,p_{Ji}$.

NOTE: If your data are not coded correctly, this estimator will abort with one of several possible diagnostics – see below for discussion. Your dependent variable must be coded 0,1,...,J. We note that this differs from some other econometric packages which use a different coding convention.

There are numerous variants and extensions of this model which can be estimated. The underlying mathematical forms are shown below, where the CDF is denoted $F(z)$ and the density is $f(z)$. (Familiar synonyms are given as well.)

Probit

$$F(z) = \int_{-\infty}^{z} \frac{\exp(-t^2/2)}{\sqrt{2\pi}} dt = \Phi(z), \qquad f(z) = \phi(z)$$

Logit

$$F(z) = \frac{\exp(z)}{1+\exp(z)} = \Lambda(z), \qquad f(z) = \Lambda(z)[1 - \Lambda(z)]$$

Complementary log log or Weibull

$$F(z) = 1 - \exp(-\exp(z)) = C(z), \qquad f(z) = \exp(z)[1 - C(z)]$$

Gompertz or log log or extreme value

$$F(z) = \exp(-\exp(-z)) = G(z), \qquad f(z) = \exp(-z)G(z)$$

Arctangent

$$F(z) = 2/\pi \arctan(z), \qquad f(z) = 2/\pi \times 1/(1 + z^2)$$

The *ordered probit* model is an extension of the probit model for a binary outcome with normally distributed disturbances. The *ordered logit model* results from the assumption that ε has a standard logistic distribution instead of a standard normal. The *ordered Weibull*, *ordered Gompertz* and *ordered arctangent* models are based on asymmetric distributions with skews to the right and left, respectively. A variety of additional specifications and extensions are provided. Basic models are treated in this chapter. Extensions such as censoring and sample selection are given in Chapter E35. Panel data models for ordered choice are discussed in Chapter E36.

E34.2 Command for Ordered Probability Models

The essential command for estimating ordered probability models is

ORDERED ; Lhs = y or p0,p1,...pJ ; Rhs = regressors $

Note that the estimator accepts proportions data for a set of J proportions. The proportions would sum to one at each observation. The probit model is the default specification. To estimate an ordered logit, ordered Weibull, ordered Gompertz or ordered arctangent model instead, add

; Model = Logit

or **; Model = Weibull (this is the extreme value model)**

or **; Model = Gompertz**

or **; Model = Arctangent**

to the command. The standardized logistic distribution (mean zero, standard deviation approximately 1.81) is used as the basis of the model instead of the standard normal. The command builder for this model is found at Model:Discrete Choice/Ordered.

E34.2.1 Data Problems

If you are using individual data, the Lhs variable must be coded 0,1,...,*J*. All the values must be present in the data. *LIMDEP* will look for empty cells. If there are any, estimation is halted. (If value '*j*' is not represented in the data, then the threshold parameter, μ_j is not estimable.) In this circumstance, you will receive a diagnostic such as

```
ORDE,Panel,BIVA PROBIT:A cell has (almost) no observations.
Empty cell: Y        never takes value  2
```

This diagnostic means exactly what it says. The ordered probability model cannot be estimated unless all cells are represented in the data. Users frequently overlook the coding requirement, y = 0,1,... If you have a dependent variable that is coded 1,2,..., you will see the following diagnostic:

```
Models - Insufficient variation in dependent variable.
```

The reason this particular diagnostic shows up is that *LIMDEP* creates a new variable from your dependent variable, say *y*, which equals zero when *y* equals zero, and one when *y* is greater than zero. It then tries to obtain starting values for the model by fitting a regression model to this new variable. If you have miscoded the Lhs variable, the transformed variable always equals one, which explains the diagnostic. In fact, there is no variation in the transformed dependent variable. If this is the case, you can simply use **CREATE** to subtract 1.0 from your dependent variable to use this estimator.

E34.2.2 Other Standard Options

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps ancillary parameters μ_j with main parameter β vector in *b*.
; Margin	displays marginal effects.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec	requests computation of the cluster form of corrected covariance estimator.
; Stratum = spec	is used with **; Cluster** for stratified and clustered data sets.
; Robust	requests a sandwich estimator or robust VC for TSCS and some discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Prob = name	saves probabilities of outcome as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E34.3 Output from the Ordered Probability Estimators

All of the ordered probit/logit models begin with an initial set of least squares results of some sort. These are suppressed unless your command contains **; OLS**. The iterations are then followed by the maximum likelihood estimates in the usual tabular format. The final output includes a listing of the cell frequencies for the outcomes. When the data are stratified, this output will also include a table of the frequencies in the strata. The log likelihood function, and a log likelihood computed assuming all slopes are zero are computed. For the latter, the threshold parameters are still allowed to vary freely, so the model is simply one which assigns each cell a predicted probability equal to the sample proportion. This appropriately measures the contribution of the nonconstant regressors to the log likelihood function. As such, the chi squared statistic given is a valid test statistic for the hypothesis that all slopes on the nonconstant regressors are zero.

The sample below shows the standard output for a model with six outcomes. These are the German health care data used in several earlier examples. The dependent variable is the self reported health satisfaction rating. For the purpose of a convenient sample application, we have truncated the health satisfaction variable at five by discarding observations – in the original data set, it is coded 0,1,...,10.

HINT: The ordered logit model typically produces the same sort of scaling of the coefficient vector that arises in the binary choice models discussed in Chapter E27. As before, the difference becomes much less pronounced when the marginal effects are considered instead. We are unaware of a convenient specification test for distinguishing between the probit and logit models. A test of normality against the broader Pearson family of distributions is described in Glewwe (1997), but it is not especially convenient. A test for skewness based on the Vuong test seems like a possibility.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                   HSAT
Log likelihood function    -11284.68638
Restricted log likelihood  -11308.02002
Chi squared [   4 d.f.]        46.66728
Significance level               .00000
McFadden Pseudo R-squared      .0020635
Estimation based on N =   8140, K =   9
Inf.Cr.AIC  =22587.373 AIC/N =    2.775
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.32892***      .07276    18.27  .0000     1.18632   1.47152
  FEMALE|     .04526*        .02546     1.78  .0755     -.00465    .09517
  HHNINC|     .35590***      .07832     4.54  .0000      .20240    .50940
  HHKIDS|     .10604***      .02665     3.98  .0001      .05381    .15827
    EDUC|     .00928         .00630     1.47  .1407     -.00307    .02162
        |Threshold parameters for index
   Mu(1)|     .23635***      .01237    19.11  .0000      .21211    .26059
   Mu(2)|     .62954***      .01440    43.72  .0000      .60132    .65777
   Mu(3)|    1.10764***      .01406    78.78  .0000     1.08008   1.13519
   Mu(4)|    1.55676***      .01527   101.94  .0000     1.52683   1.58669
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+---------------------------------------------------------------------+
|                CELL FREQUENCIES FOR ORDERED CHOICES                  |
+---------------------------------------------------------------------+
|              Frequency        Cumulative  < =     Cumulative  > =   |
|Outcome      Count    Percent    Count    Percent    Count    Percent |
|----------- ------- ---------  ------- ---------  ------- --------- |
|HSAT=00        447     5.4914      447     5.4914     8140  100.0000 |
|HSAT=01        255     3.1327      702     8.6241     7693   94.5086 |
|HSAT=02        642     7.8870     1344    16.5111     7438   91.3759 |
|HSAT=03       1173    14.4103     2517    30.9214     6796   83.4889 |
|HSAT=04       1390    17.0762     3907    47.9975     5623   69.0786 |
|HSAT=05       4233    52.0025     8140   100.0000     4233   52.0025 |
+---------------------------------------------------------------------+
```

Cross tabulation of predictions and actual outcomes

y(i,j)	0	1	2	3	4	5	Total
0	0	0	0	0	0	447	447
1	0	0	0	0	0	255	255
2	0	0	0	0	0	642	642
3	0	0	0	0	0	1173	1173
4	0	0	0	0	0	1390	1390
5	0	0	0	0	0	4233	4233
Total	0	0	0	0	0	8140	8140

Row = actual, Column = Prediction, Model = Probit
Prediction is number of the most probable cell.

Cross tabulation of outcomes and predicted probabilities.

y(i,j)	0	1	2	3	4	5	Total
0	26	15	36	66	77	228	447
1	14	8	21	37	44	131	255
2	36	20	51	93	110	331	642
3	64	37	93	170	200	609	1173
4	75	43	109	200	237	725	1390
5	230	132	333	610	722	2206	4233
Total	445	255	644	1176	1389	4230	8140

Row = actual, Column = Prediction, Model = Probit
Value(j,m)=Sum(i=1,N)y(i,j)*p(i,m).
Column totals may not match cell sums because of rounding error.

The model output is followed by a $(J+1)\times(J+1)$ frequency table of predicted versus actual values. (This table is not given when data are grouped or when there are more than 10 outcomes.) The predicted outcome for this tabulation is the one with the largest predicted probability. Even though the model appears to be highly significant, the table of predictions has seems to suggest a lack of predictive power. Tables such as the one above are common with this model. The driver of the result is the sample configuration of the data. Note in the frequency table that the sample is quite unbalanced, and the highest outcome is quite likely to have the highest probability for every observation. The estimation criterion for the ordered probability model is unrelated to its ability to predict those cells, and you will rarely see a predictions table that closely matches the actual outcomes. It often happens that even in a set of results with highly significant coefficients, only one or a few of the outcomes are predicted by the model. The second table relates more closely to the aggregate predictions of the model. The table entries are the sample proportions that would be predicted for each outcome. For example, the first row of the table shows that 447 individuals in the sample chose outcome 0. For every individual, the model produces a full set of J+1 probabilities. For the 447 individuals, 8140 times the sum of the probabilities of outcome 0 equals 26, 8140 times the sum of the probabilities of outcome 1 equals 15, and so on.

E34.3.1 Robust Covariance Matrix Estimation

The Sandwich Estimator

The standard robust covariance matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log F_i}{\partial\hat{\boldsymbol{\gamma}}\partial\hat{\boldsymbol{\gamma}}'}\right)\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{\partial \log F_i}{\partial\hat{\boldsymbol{\gamma}}}\right)\left(\frac{\partial \log F_i}{\partial\boldsymbol{\gamma}}\right)'\right]\left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log F_i}{\partial\hat{\boldsymbol{\gamma}}\partial\hat{\boldsymbol{\gamma}}'}\right)\right]^{-1}$$

where $\hat{\boldsymbol{\gamma}}$ indicates the full set of parameters in the model. To obtain this matrix with any of the forms of the ordered choice models, use

; Robust

in the **ORDERED** command.

Clustering and Stratification

A related calculation is used when observations occur in groups which may be correlated. This is rather like a panel; one might use this approach in a random effects kind of setting in which observations have a common latent heterogeneity. The parameter estimator is unchanged in this case, but an adjustment is made to the estimated asymptotic covariance matrix. Full details on this estimator appear in Chapter R10. To specify this estimator, use

; Cluster = specification

where the specification is either a fixed number of observations or the name of a variable that provides an identifier for the cluster, such as an *id* number. Note that if there is exactly one observation per cluster, then this is $G/(G\text{-}1)$ times the sandwich estimator discussed above. Also, if you have fewer clusters than parameters, then this matrix is singular – it has rank equal to the minimum of G and K, the number of parameters.

The extension of this estimator to stratified data is described in detail in Section R10.3. To use this with the **; Cluster** specification, add

; Stratum = specification

E34.3.2 Saved Results

Computation of predictions and ancillary variables is as follows: For each observation, the predicted probabilities for all J+1 outcomes are computed. Then if you request **; List**, the listing will contain

Predicted Y: Y with the largest probability.

Residual: the largest of the J+1 probabilities (i.e., Prob[y = fitted Y]).

Var1: the estimate of $E[y_i] = \sum_{i=0}^{J} i \times \text{Prob}[Y_i = i]$.
(Note that since the outcomes are only ordinal, this is not a true expected value.)

Var2: the probability estimated for the observed Y.

Estimation results kept by the estimator are as follows:

Matrices: *b* = estimate of $\boldsymbol{\beta}$,
varb = estimated asymptotic covariance,
mu = J-1 estimated μs.

Scalars: *kreg*, *nreg*, and *logl*.

Last Model: The labels are *b_variables*, *mu*1, ...

Last Function: Prob(y = highest outcome | x)

The specification **; Par** adds $\boldsymbol{\mu}$ (the set of estimated threshold values) to *b* and *varb*. The additional matrix, *mu* is kept regardless, but the estimated asymptotic covariance matrix is lost unless the command contains **; Par**. The *Last Function* is used in the **SIMULATE** and **PARTIAL EFFECTS** routines. The default function is the probability of the highest outcome. You can specify a different outcome in the command with

; Outcome = j

where j is the desired outcome. For example, in our earlier application in which outcomes are 0,1,2,3,4,5, the command might specify

PARTIAL EFECTS ; Effects: hhninc ; Outcome = 3 $

and likewise for **SIMULATE**. A full examination of all outcomes is obtained by using

; Outcome = *

E34.4 Model Structure and Data

E34.4.1 Constant Term

This model must include a constant term, *one*, as the first Rhs variable. Since the equation does include a constant term, one of the μs is not identified. We normalize μ_0 to zero. (Consider the special case of the binary probit model with something other than zero as its threshold value. If it contains a constant, this cannot be estimated.) Data may be grouped or individual. (Survey data might logically come in grouped form.) If you provide individual data, the dependent variable is coded 0, 1, 2, ..., J. There must be at least three values. Otherwise, the binary probit model applies. If the data are grouped, a full set of proportions, p_0, p_1, ..., p_J, which sum to one at every observation must be provided. In the individual data case, the data are examined to determine the value of J, which will be the largest observed value of y which appears in the sample. In the grouped data case, J is one less than the number of Lhs variables you provide. Once again, we note that other programs sometimes use different normalizations of the model. For example, if the constant term is forced to equal zero, then one will instead, add a nonzero threshold parameter, μ_0, which equals zero in the presence of a nonzero constant term.

E34.4.2 Censored Data

Suppose that the dependent variable for the ordered probability model is censored for some observations. For example, suppose that Y takes values 0,1,2,...,10. But, for some observations, we observe only a five and an indicator that the dependent variable was actually at least five, though the actual value is unknown. Then, for this observation, the relevant probability is the sum of the probabilities from five to 10, not just the cell probability for $Y = 5$. These sorts of data are likely to occur in the context of the ordered extreme value model for duration described in Chapter E59.3. *LIMDEP* will accommodate this form of censoring, and modify the log likelihood function and all estimates accordingly. Censoring is indicated as in the other duration models. *That is, when data are censored, you can so indicate by including in your model command a second Lhs variable which is the censoring indicator.* Remember that the indicator takes values zero for the censored observations and one for the uncensored observations.

Mathematically, the censored data model is a simple extension of the familiar ordered probability model. Let $y = 0,1,...,J$. The probability that y equals j is

$$\text{Prob}[\text{observed } y = j] \;=\; F[\mu_j - \boldsymbol{\beta}'\mathbf{x}] - F[\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}].$$

The log likelihood and its derivatives are built up from this relationship. If, however, y is censored, then the observed value $y = j$ contributes a term

$$\text{Prob}[\text{observed } y = j] \;=\; \sum_{i=j}^{J} \{\text{F}[\mu_i - \boldsymbol{\beta}'\mathbf{x}] - \text{F}[\mu_{i-1} - \boldsymbol{\beta}'\mathbf{x}]\}.$$

The log likelihood and its derivatives are obtained just by summing all of the relevant cells.

> **NOTE:** Recall that *LIMDEP* deduces the value of J from the data – the highest value of y_i. Therefore, you must have some uncensored observations, and J is the largest value of y_i observed among these data points. By implication, if a censored y_i exceeds J, there is a problem in the data.

> **NOTE:** (On computation) This additional summation will not add any additional time to fit your model. The reason is that *LIMDEP* already obtains the log likelihood function by taking a weighted sum of all J+1 terms, where in the standard case, the weights are either [0,0,...,1,0,...] for the individual case or [p0,p1,...,pJ] in the grouped data case. For the censored data case, we merely change the weight vector to [0,0,...,1,1,1...], which is a trivial operation.

In the example below, we have randomly censored about 20% of the observations. The commands are

```
SAMPLE    ; All $
REJECT    ; _groupti < 7 $
CREATE    ; censor = Rnu(0,1) > .2 $
ORDERED   ; Lhs = newhsat,censor ; Rhs = one,female,hhninc,hhkids,educ
          ; Logit $
```

The results do reveal an impact of the censoring. For comparison, the same model estimated without censoring is presented with the results.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable               NEWHSAT
Log likelihood function    -11253.80999
Censoring indicator is CENSOR
Total observations =    6209.0
Uncensored =    5002.0, censored =    1207.0
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    3.32240***      .14571    22.80  .0000     3.03683   3.60798
  FEMALE|    -.32084***      .04839    -6.63  .0000     -.41569   -.22600
  HHNINC|     .21830         .15550     1.40  .1604     -.08647    .52307
  HHKIDS|     .42233***      .04748     8.89  .0000      .32926    .51539
    EDUC|     .09499***      .01233     7.70  .0000      .07082    .11916
        |Threshold parameters for index
   Mu(1)|     .51490***      .05948     8.66  .0000      .39831    .63149
   Mu(2)|    1.23940***      .05891    21.04  .0000     1.12393   1.35486
   Mu(3)|    1.91493***      .05111    37.47  .0000     1.81476   2.01509
   Mu(4)|    2.44072***      .04558    53.55  .0000     2.35138   2.53005
   Mu(5)|    3.44902***      .03893    88.60  .0000     3.37273   3.52532
   Mu(6)|    3.89066***      .03724   104.48  .0000     3.81767   3.96364
   Mu(7)|    4.51839***      .03603   125.39  .0000     4.44777   4.58902
   Mu(8)|    5.54920***      .03807   145.75  .0000     5.47458   5.62383
   Mu(9)|    6.26265***      .04367   143.40  .0000     6.17706   6.34825
--------+--------------------------------------------------------------------
(Uncensored)
-----------------------------------------------------------------------------
Log likelihood function    -12971.89392
```

```
--------+-------------------------------------------------------------
        |Index function for probability
Constant|    3.02189***      .13081    23.10  .0000     2.76551   3.27827
  FEMALE|    -.31859***      .04729    -6.74  .0000     -.41129   -.22590
  HHNINC|     .23133*        .13880     1.67  .0956     -.04072    .50338
  HHKIDS|     .47849***      .04529    10.56  .0000      .38972    .56726
    EDUC|     .10241***      .01122     9.12  .0000      .08041    .12441
        |Threshold parameters for index
   Mu(1)|     .49176***      .05264     9.34  .0000      .38859    .59493
   Mu(2)|    1.26288***      .05011    25.20  .0000     1.16468   1.36109
   Mu(3)|    1.94907***      .04093    47.62  .0000     1.86886   2.02929
   Mu(4)|    2.48180***      .03468    71.57  .0000     2.41383   2.54976
   Mu(5)|    3.48744***      .02747   126.94  .0000     3.43360   3.54129
   Mu(6)|    3.94860***      .02594   152.22  .0000     3.89776   3.99944
   Mu(7)|    4.61859***      .02627   175.79  .0000     4.56710   4.67009
   Mu(8)|    5.70197***      .03154   180.78  .0000     5.64015   5.76378
   Mu(9)|    6.48830***      .04110   157.86  .0000     6.40774   6.56886
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

E34.5 Partial Effects and Simulations

There is potentially a large amount of output for the ordered choice model, in addition to the basic model results. There is no single conditional mean because the outcomes are labels, not measures. There are J+1 probabilities to analyze,

$$\text{Prob[cell j]} = F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j\text{-}1} - \boldsymbol{\beta}'\mathbf{x}_i).$$

Typically, the highest or lowest cell is of interest. However, the **PARTIAL EFFECTS** (or just **PARTIALS**) and **SIMULATE** commands can be used to examine any or all of them.

Marginal effects in the ordered probability models are also quite involved. Since there is no meaningful conditional mean function to manipulate, we compute, instead, the effects of changes in the covariates on the cell probabilities. These are:

$$\partial\text{Prob[cell } j]/\partial\mathbf{x}_i \;=\; [f(\mu_{j\text{-}1} - \boldsymbol{\beta}'\mathbf{x}_i) - f(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i)] \times \boldsymbol{\beta},$$

where $f(.)$ is the appropriate density for the standard normal, $\phi(\bullet)$, logistic density, $\Lambda(\bullet)(1\text{-}\Lambda(\bullet))$, Weibull, Gompertz or arctangent. Each vector is a multiple of the coefficient vector. But it is worth noting that the magnitudes are likely to be very different. In at least one case, Prob[cell 0], and probably more if there are more than three outcomes, the partial effects have exactly the opposite signs from the estimated coefficients.

NOTE: This estimator segregates dummy variables for separate computation in the marginal effects. The marginal effect for a dummy variable is the difference of the two probabilities, with and without the variable.

Partial effects for the ordered probability models are obtained internally in the command by adding

; Partial Effects

in the command. This produces a table oriented to the outcomes, such as the one below. A second summary that is oriented to the variables rather than the outcomes is requested with

; Partial Effects ; Full

The internal results are computed at the means of the data. Partial effects can also be obtained with the **PARTIALS** command. The third set of results below is obtained with

PARTIALS ; Effects: hhninc ; Outcome = * $

This command produces average partial effects by default, but you can request that they be computed at the data means by adding **; Means** to the command. Probabilities for particular outcomes are obtained with the **SIMULATE** command. An example appears below.

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
    HSAT|      Effect    Elasticity      z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
 *FEMALE|    -.00498*        -.09207    -1.77  .0763     -.01049    .00053
  HHNINC|    -.03907***      -.23836    -4.53  .0000     -.05599   -.02216
 *HHKIDS|    -.01132***      -.20926    -4.08  .0000     -.01676   -.00588
    EDUC|    -.00102         -.20477    -1.47  .1409     -.00237    .00034
        |--------------[Partial effects on Prob[Y=01] at means]--------------
 *FEMALE|    -.00210*        -.06711    -1.78  .0758     -.00441    .00022
  HHNINC|    -.01647***      -.17397    -4.54  .0000     -.02358   -.00936
 *HHKIDS|    -.00483***      -.15473    -4.04  .0001     -.00718   -.00249
    EDUC|    -.00043         -.14945    -1.47  .1408     -.00100    .00014
        |--------------[Partial effects on Prob[Y=02] at means]--------------
 *FEMALE|    -.00414*        -.05244    -1.77  .0760     -.00872    .00043
  HHNINC|    -.03257***      -.13605    -4.50  .0000     -.04675   -.01838
 *HHKIDS|    -.00964***      -.12205    -3.98  .0001     -.01439   -.00489
    EDUC|    -.00085         -.11688    -1.47  .1412     -.00198    .00028
        |--------------[Partial effects on Prob[Y=03] at means]--------------
 *FEMALE|    -.00473*        -.03273    -1.77  .0764     -.00997    .00050
  HHNINC|    -.03727***      -.08501    -4.43  .0000     -.05375   -.02078
 *HHKIDS|    -.01121***      -.07751    -3.87  .0001     -.01689   -.00554
    EDUC|    -.00097         -.07303    -1.47  .1417     -.00227    .00032
        |--------------[Partial effects on Prob[Y=04] at means]--------------
 *FEMALE|    -.00208*        -.01214    -1.77  .0762     -.00438    .00022
  HHNINC|    -.01643***      -.03166    -4.34  .0000     -.02385   -.00901
 *HHKIDS|    -.00518***      -.03026    -3.66  .0002     -.00795   -.00241
    EDUC|    -.00043         -.02720    -1.47  .1427     -.00100    .00014
        |--------------[Partial effects on Prob[Y=05] at means]--------------
 *FEMALE|     .01803*         .03469     1.78  .0755     -.00185    .03792
  HHNINC|     .14181***       .09003     4.54  .0000      .08065    .20297
 *HHKIDS|     .04219***       .08116     3.99  .0001      .02145    .06292
    EDUC|     .00370          .07734     1.47  .1407     -.00122    .00861
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+--------------------------------------------------------------------+
| Summary of Marginal Effects for Ordered Probability Model (probit)  |
| Effects computed at means.  Effects for binary variables (*) are    |
| computed as differences of probabilities, other variables at means. |
| Binary variables change only by 1 unit so s.d. changes are not shown.|
| Elasticities for binary variables = partial effect/probability = %chgP |
+--------------------------------------------------------------------+
+--------------------------------------------------------------------+
|          Binary(0/1) Variable FEMALE      Changes in *FEMALE    % chg|
|         ------------------------------   ------------------------------
Outcome   Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00    -.00498    -.00498     .00000        -        -.00498    -.09207
Y = 01    -.00210    -.00708     .00498        -        -.00210    -.06711
Y = 02    -.00414    -.01122     .00708        -        -.00414    -.05244
Y = 03    -.00473    -.01595     .01122        -        -.00473    -.03273
Y = 04    -.00208    -.01803     .01595        -        -.00208    -.01214
Y = 05     .01803     .00000     .01803        -         .01803     .03469
+--------------------------------------------------------------------+
|           Continuous Variable HHNINC      Changes in HHNINC     % chg|
|         ------------------------------   ------------------------------
Outcome   Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00    -.03907    -.03907     .00000    -.00655    -.11703    -.23836
Y = 01    -.01647    -.05555     .03907    -.00276    -.04933    -.17397
Y = 02    -.03257    -.08811     .05555    -.00546    -.09753    -.13605
Y = 03    -.03727    -.12538     .08811    -.00625    -.11161    -.08501
Y = 04    -.01643    -.14181     .12538    -.00275    -.04921    -.03166
Y = 05     .14181     .00000     .14181     .02377     .42472     .09003
+--------------------------------------------------------------------+
|          Binary(0/1) Variable HHKIDS      Changes in *HHKIDS    % chg|
|         ------------------------------   ------------------------------
Outcome   Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00    -.01132    -.01132     .00000        -        -.01132    -.20926
Y = 01    -.00483    -.01615     .01132        -        -.00483    -.15473
Y = 02    -.00964    -.02579     .01615        -        -.00964    -.12205
Y = 03    -.01121    -.03701     .02579        -        -.01121    -.07751
Y = 04    -.00518    -.04219     .03701        -        -.00518    -.03026
Y = 05     .04219     .00000     .04219        -         .04219     .08116
+--------------------------------------------------------------------+
|           Continuous Variable EDUC        Changes in EDUC       % chg|
|         ------------------------------   ------------------------------
Outcome   Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00    -.00102    -.00102     .00000    -.00212    -.01120    -.20477
Y = 01    -.00043    -.00145     .00102    -.00089    -.00472    -.14945
Y = 02    -.00085    -.00230     .00145    -.00177    -.00934    -.11688
Y = 03    -.00097    -.00327     .00230    -.00202    -.01069    -.07303
Y = 04    -.00043    -.00370     .00327    -.00089    -.00471    -.02720
Y = 05     .00370     .00000     .00370     .00770     .04066     .07734
----------------------------------------------------------------------
```

PARTIALS ; Effects: hhninc ; Outcome = * $

```
-----------------------------------------------------------------------
Partial Effects  Analysis for Ordered Probit     Probability Y = 5
-----------------------------------------------------------------------
Effects on function with respect to HHNINC
Results are computed by average over sample observations
Partial effects for continuous HHNINC   computed by differentiation
Effect is computed as derivative     = df(.)/dx
-----------------------------------------------------------------------
df/dHHNINC          Partial    Standard
(Delta method)      Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
APE Prob(y= 0)      -.03930     .00872    4.51     -.05640     -.02220
APE Prob(y= 1)      -.01643     .00373    4.41     -.02374     -.00912
APE Prob(y= 2)      -.03238     .00734    4.41     -.04677     -.01800
APE Prob(y= 3)      -.03694     .00827    4.47     -.05315     -.02072
APE Prob(y= 4)      -.01624     .00382    4.26     -.02372     -.00876
APE Prob(y= 5)       .14129     .03099    4.56      .08055      .20204
```

SIMULATE ; Scenario: & hhninc = 0(.05)1 ; Plot(ci) ; Outcome = 4 $

```
-----------------------------------------------------------------------
Model Simulation Analysis for Ordered Probit     Probability Y = 4
-----------------------------------------------------------------------
Simulations are computed by average over sample observations
-----------------------------------------------------------------------
User Function       Function   Standard
(Delta method)       Value      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
Avrg. Function       .17068     .00988   17.27      .15131      .19005
HHNINC  =   .00      .17528     .01026   17.09      .15517      .19538
HHNINC  =   .05      .17477     .01021   17.11      .15476      .19479
HHNINC  =   .10      .17421     .01016   17.14      .15429      .19413
HHNINC  =   .15      .17360     .01011   17.17      .15379      .19342
HHNINC  =   .20      .17294     .01005   17.20      .15324      .19265
HHNINC  =   .25      .17223     .00999   17.23      .15264      .19182
HHNINC  =   .30      .17147     .00993   17.26      .15199      .19094
HHNINC  =   .35      .17065     .00987   17.28      .15130      .19001
HHNINC  =   .40      .16979     .00982   17.30      .15055      .18903
HHNINC  =   .45      .16888     .00976   17.30      .14975      .18801
HHNINC  =   .50      .16793     .00971   17.30      .14890      .18695
HHNINC  =   .55      .16692     .00966   17.28      .14799      .18586
HHNINC  =   .60      .16587     .00962   17.24      .14701      .18473
HHNINC  =   .65      .16478     .00959   17.18      .14598      .18358
HHNINC  =   .70      .16364     .00957   17.09      .14488      .18241
HHNINC  =   .75      .16246     .00957   16.98      .14371      .18122
HHNINC  =   .80      .16124     .00958   16.84      .14247      .18001
HHNINC  =   .85      .15998     .00960   16.66      .14116      .17880
HHNINC  =   .90      .15868     .00965   16.45      .13978      .17758
HHNINC  =   .95      .15734     .00971   16.21      .13832      .17637
HHNINC  =  1.00      .15596     .00979   15.93      .13678      .17515
```

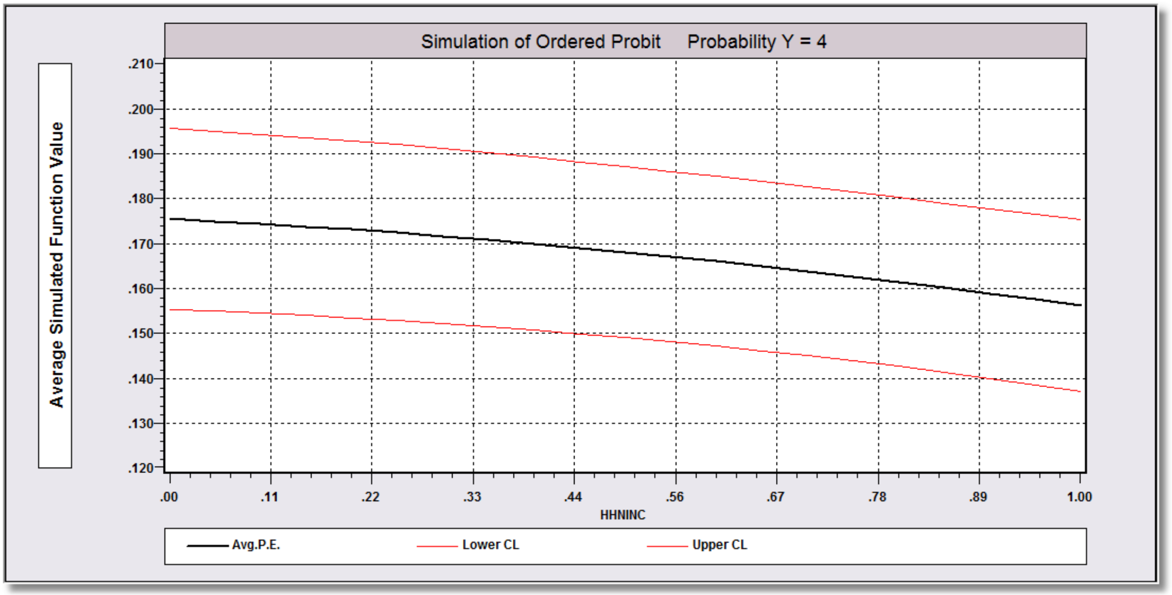


Figure E34.1 Simulated Probabilities

E34.6 Technical Details for Ordered Choice Models

For brevity, we generalize the basic model at this point by integrating both the heteroscedasticity and stratification that are presented in the next two chapters. Either or both can be assumed away. Define the augmented vector of threshold parameters

$$\boldsymbol{\mu} = \mu_{-1}\ \mu_0\ \mu_1\ \ldots\ \mu_{J-1}\ \mu_J, \text{ in which } \mu_{-1} = -\infty,\ \ \mu_0 = 0, \text{ and } \mu_J = +\infty.$$

Then, $\text{Prob}[y_{i,s} = j] = F[(\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i] - F[(\mu_{j-1,s} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i],\ j = 0,1,\ldots,J$

where s denotes the stratum, which may be one for all observations, 'w_i' is the individual specific standard deviation, which is 1.0 for all i, or an observed variable, w_i, or $\exp(\boldsymbol{\gamma}'\mathbf{z}_i)$ with unknown parameters $\boldsymbol{\gamma}$ and observed variables $\mathbf{z}_i$ which does not include a constant. Then, let

$F(.)$ = the CDF of the distribution of ε, normal, logistic; Weibull, arctangent or Gompertz.

The log likelihood function is

$$\log L = \Sigma_i \log L_i$$

$$= \Sigma_i \log \text{Prob}[Y_{i,s} = y_{i,s}],$$

where $Y_{i,s}$ = the theoretical random variable

and $y_{i,s}$ = the observed value of $Y_{i,s}$ for observation i in stratum s.

The first derivatives are

$$\frac{\partial \log L_i}{\partial \boldsymbol{\beta}} = \left[\frac{f\left(\dfrac{\mu_{j-1,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right) - f\left(\dfrac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right)}{F\left(\dfrac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right) - F\left(\dfrac{\mu_{j-1,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right)} \right] \frac{1}{w_i}\mathbf{x}_i$$

where $f(.)$ denotes the appropriate density, $\phi(.)$ or $\Lambda(1-\Lambda)$ for normal or logistic, etc. For convenience, denote

$f_{j,s} = f[(\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i]$

and $F_{j,s} = F[(\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i]$,

and likewise for 'j-1.' By convention,

$f_{-1,s} = F_{-1,s} = f_{J,s} = 0$, and $F_{J,s} = 1$.

Then, $\partial \log L_i/\partial\mu_{j,s} = [f_{j,s} / (F_{j,s} - F_{j-1,s})] / w_i$

and $\partial \log L_i/\partial\mu_{j-1,s} = -[f_{j-1,s} / (F_{j,s} - F_{j-1,s})] / w_i$.

These imply that $\partial \log L_i/\partial\mu_m = 0$ if $m = -1, 0,$ or J.

For the model with multiplicative heteroscedasticity,

$$\frac{\partial \log L_i}{\partial \boldsymbol{\gamma}} = \left[\frac{f\left(a_{j-1,s}\right)a_{j-1,s} - f\left(a_{j,s}\right)a_{j,s}}{F\left(a_{j,s}\right) - F\left(a_{j-1.s}\right)} \right]\mathbf{z}_i,\ a_{j,s} = \frac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}$$

For estimation with grouped data and observed proportions $p_0,...,p_J$,

$\log L_j = \Sigma_j\, p_j \log \text{Prob}[Y_{i,s} = j]$.

The preceding expressions are summed over all outcomes. Second derivatives are extremely tedious, but use common expressions and are in principle straightforward. The analytic Hessian is used for computing asymptotic standard errors.

The algorithm used to obtain the maximum likelihood estimates is BFGS. Starting values are obtained by least squares, either ordinary or generalized depending on the type of data. In either case, this initial regression is based on the dichotomy formed by using the binary indicator **1**[y > 0] as if a univariate probit model applied. For grouped data, p_+ and $p_0 = 1-p_+$ provide the dichotomy, and minimum chi squared estimates are obtained. The constant term and the values of the thresholds are then estimated by using the cell frequencies under the assumption that all of the slopes are zero. We segment the real line in such a way that the normal (or other distribution) probabilities corresponding to this partition match the sample cell frequencies. You may provide your own starting values with

; Start = start values for β, start values for $\mu_1,...,\mu_{J-1}$.

The first threshold parameter, μ_0 equals 0.0. If the model contains a constant term, this is not estimable. Note, also, that there is no μ_J. For example, if J = 2, so y = 0,1,2, then only μ_1 is to be estimated. The full parameter vector is

$$\boldsymbol{\Theta} = [\beta_1,...,\beta_K,\mu_1,...,\mu_{J-1}].$$

NOTE: It is necessary for the threshold parameters to be strictly ordered. That is, $\mu_j > \mu_{j-1}$. Occasionally, during the line search, this requirement will be violated by a trial value. A diagnostic will be issued,

```
ORDERED PROBIT - Current estimated thresholds not ordered.
```

but estimation will continue. This is merely a warning, and the line search will continue with a smaller step. But, if your data are such that there are many cells, and some of them are nearly empty, this condition may be persistent, and it is possible that the estimation process will break down.

The partial effects are obtained by a manipulation of the likelihood equations.

$$\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{z}_i)}{\partial \mathbf{x}_i} = \left[f\left(\frac{\mu_{j-1,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right) - f\left(\frac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{w_i}\right)\right]\frac{1}{w_i}\boldsymbol{\beta}$$

$$\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{z}_i)}{\partial \mathbf{z}_i} = \left[\frac{f\left(a_{j-1,s}\right)a_{j-1,s} - f\left(a_{j,s}\right)a_{j,s}}{F\left(a_{j,s}\right) - F\left(a_{j-1.s}\right)}\right]\mathbf{z}_i,\; a_{j,s} = \frac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{z}_i)}.$$

E35: Extended Ordered Choice Models

E35.1 Introduction

The basic ordered choice model is based on the following specification: There is a latent regression,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ \mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0,\ \mathrm{Var}[\varepsilon_i|\mathbf{x}_i] = 1,$$

The observation mechanism results from a complete censoring of the latent dependent variable as follows:

$$\begin{aligned} y_i &= 0 \text{ if } y_i \le \mu_0, \\ &= 1 \text{ if } \mu_0 < y_i \le \mu_1, \\ &= 2 \text{ if } \mu_1 < y_i \le \mu_2, \\ &\ldots \\ &= J \text{ if } y_i > \mu_{J-1}. \end{aligned}$$

The latent 'preference' variable, y_i^* is not observed. The observed counterpart to y_i^* is y_i. The probabilities which enter the log likelihood function are

$$\mathrm{Prob}[y_i = j] = \mathrm{Prob}[y_i^* \text{ is in the } j\text{th range}].$$

Estimation and analysis of the basic model are presented in Chapter E34. A variety of additional specifications and extensions are supported. The extensions shown in this chapter are:

- heteroscedasticity,
- sample selection and treatment effects,
- generalized ordered, proportional odds and the parallel regressions assumption,
- hierarchical ordered probit models,
- zero inflated ordered probit models,
- bivariate ordered probit and polychoric correlation.

E35.2 Weighting and Heteroscedasticity

An ordered probit model with simple heteroscedasticity,

$$\mathrm{Var}[\varepsilon_i] = w_i^2,$$

may be estimated with

ORDERED ; Rhs = ... ; Lhs = ...
; Wts = your weighting variable, wi
; Heteroscedastic $

Your command gives the name of the variable which carries the *observed* individual specific *standard deviations*. This formulation does not add new parameters to the model, and only instructs the estimator how the weighting variable is to be handled.

This approach is different from estimating the model with weights. Without **; Het**, this model is treated as any other weighted log likelihood, and the estimator maximizes

$$\log L \quad = \quad \sum_{i=1}^{n} w_i \log \mathrm{Prob}(\textit{observed outcome}_i)$$

where $\quad \mathrm{Prob}[\text{cell } j] \; = \; F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i).$

With **; Het**, the probabilities are built up from the heteroscedastic random variable, but the terms in the log likelihood are unweighted. With this form of the command, using **; Het**, the model is

$$\mathrm{Prob}[\text{cell } j] \; = \; F[(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i)/w_i] - F[(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i]$$

and $\quad \log L \quad = \quad \sum_{i=1}^{n} \log \mathrm{Prob}(\textit{observed outcome}_i)$

E35.3 Multiplicative Heteroscedasticity

The model with multiplicative heteroscedasticity,

$$\mathrm{Var}[\varepsilon_i] \; = \; [\exp(\boldsymbol{\gamma}'\mathbf{z}_i)]^2,$$

is requested with

```
ORDERED     ; Rhs = ... ; Lhs = ...
            ; Het
            ; Rh2 = list of variables in z $
```

NOTE: Do not include a constant (*one*) in **z**. A variable in **z** which has no variation, such as *one*, will lead to a singular Hessian, and the estimator will fail to converge.

This formulation adds a vector of new parameters to the model. For purposes of starting values, restrictions, and hypothesis tests, the full parameter vector becomes

$$\boldsymbol{\Theta} \; = \; [\beta_1,...,\beta_K,\gamma_1,...,\gamma_L,\mu_1,...,\mu_{J-1}].$$

You can use **; Rst** and **; CML:** for imposing restrictions as usual. As always, restrictions that force ancillary variance parameters (γ_h) to equal parameters in the conditional mean function (β_k) will rarely produce satisfactory results. In the saved results, the estimator of $\boldsymbol{\gamma}$ will always be included in *b* and *varb*. Thus, if you want to extract parts of the parameter vector after estimation, you might use

```
NAMELIST    ; x = ...
            ; z = ... $
ORDERED     ; Lhs  = y ; Rhs = x
            ; Rh2 = z ; Het $
CALC        ; k = Col(x) ; k1 = k+1 ; kt = k + Col(z) $
MATRIX      ; beta = b(1:k)
            ; gamma = b(k1:kt) $
```

The μ threshold parameters are still the ancillary parameters. Marginal effects, fitted values, and so on are requested exactly as before with this extension of the ordered probit model. In the *Last Model* labels list, the variance parameters will be denoted *c_variable*, so with this model, the complete list of labels is

Last Model = [B_...,C_...,MU1,...].

The *Last Function* for the model is the probability including the exponential heteroscedasticity model

$$\text{Prob}(y=1\,|\,\mathbf{x},\mathbf{z}) = F\left(\frac{\mu_j - \boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{z})}\right) - F\left(\frac{\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{z})}\right)$$

E35.3.1 Testing for Heteroscedasticity

The model with homoscedastic disturbances is nested in this model ($\boldsymbol{\gamma} = \mathbf{0}$) so the standard tests, i.e., LM, likelihood ratio, and Wald, are available for testing the specification. The first two of these will be very convenient. To carry out an LM test, you could use the following: First define the two variable lists.

```
NAMELIST    ; x = ...
            ; z = ... $
```

Fit the model without heteroscedasticity. This command saves *b* and *mu* needed later.

```
ORDERED     ; Lhs = y  ; Rhs = x $
```

Define the zero vector for the variance parameters.

```
MATRIX      ; {h = Col(z)} ; gamma = Init (h,1,0) $
```

Now, fit the heteroscedastic model, but do not iterate. This displays the LM statistic.

```
ORDERED     ; Lhs = y ; Rhs = x ; Rh2 = z ; Het
            ; Start = b,gamma,mu ; Maxit = 0 $
```

To use a likelihood ratio test, instead, the preceding is modified as follows:

1. Add **CALC ; lr = logl $** after the first **ORDERED** command.

2. Omit **; Maxit = 0** from the second **ORDERED** command.

3. Add the command

   ```
   CALC     ; List ; chi = 2*(logl - lr) $
   ```

   after the second **ORDERED** command; *chi* is the chi squared statistic. This can be referred to the table with

   ```
   CALC     ; cstar = Ctb(.95,L) $
   ```

   which provides the necessary critical value.

The following experiment illustrates these computations. We test for heteroscedasticity in the health satisfaction model, using the three standard tests in an ordered logit model as the platform. To simplify it a bit, we use a restricted sample of only those individuals observed in all seven periods.

```
SAMPLE     ; All $
REJECT     ; _groupti < 7 $
ORDERED    ; Lhs = newhsat
           ; Rhs = one,female,hhninc,hhkids,educ
           ; Logit $
CALC       ; lr = logl $
```

This command carries out the LM test. The starting values are from the previous model for β and μ and zeros for the elements of γ. The test is requested with **; Maxit = 0**.

```
ORDERED    ; Lhs = newhsat
           ; Rhs = one,female,hhninc,hhkids,educ
           ; Logit ; Het ; Rh2 = married,univ,working,female,hhninc
           ; Start = b,0,0,0,mu ; Maxit = 0 $
```

This command estimates the full heteroscedastic model. Based on these results, we then carry out the likelihood ratio and Wald tests.

```
ORDERED    ; Lhs = newhsat
           ; Rhs = one,female,hhninc,hhkids,educ
           ; Logit ; Het ; Rh2 = married,univ,working,female,hhninc $
CALC       ; lu = logl $
CALC       ; List ; lrtest = 2*(lu - lr) $
MATRIX     ; gamma = b(6:10) ; vgamma = varb(6:10,6:10) $
MATRIX     ; List
           ; waldstat = gamma'<vgamma>gamma $
```

As might be expected in a sample this large, the three tests give the same answer. The LM, LR and Wald statistics obtained are 84.16200, 84.26808 and 83.90174, respectively.

The first set of results are for the restricted, homoscedastic model.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                NEWHSAT
Log likelihood function       -12971.89392
Restricted log likelihood     -13138.97978
Chi squared [    4 d.f.]         334.17171
Significance level                  .00000
McFadden Pseudo R-squared         .0127168
Estimation based on N =    6209, K =  14
Inf.Cr.AIC  =25971.788 AIC/N =     4.183
Underlying probabilities based on Logistic
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    3.02189***      .13081    23.10  .0000     2.76551   3.27827
  FEMALE|    -.31859***      .04729    -6.74  .0000     -.41129   -.22590
  HHNINC|     .23133*        .13880     1.67  .0956     -.04072    .50338
  HHKIDS|     .47849***      .04529    10.56  .0000      .38972    .56726
    EDUC|     .10241***      .01122     9.12  .0000      .08041    .12441
        |Threshold parameters for index
   Mu(1)|     .49176***      .05264     9.34  .0000      .38859    .59493
   Mu(2)|    1.26288***      .05011    25.20  .0000     1.16468   1.36109
   Mu(3)|    1.94907***      .04093    47.62  .0000     1.86886   2.02929
   Mu(4)|    2.48180***      .03468    71.57  .0000     2.41383   2.54976
   Mu(5)|    3.48744***      .02747   126.94  .0000     3.43360   3.54129
   Mu(6)|    3.94860***      .02594   152.22  .0000     3.89776   3.99944
   Mu(7)|    4.61859***      .02627   175.79  .0000     4.56710   4.67009
   Mu(8)|    5.70197***      .03154   180.78  .0000     5.64015   5.76378
   Mu(9)|    6.48830***      .04110   157.86  .0000     6.40774   6.56886
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The next set of results is the computation of the Lagrange multiplier statistic. This next command does not reestimate the model. Note that the coefficient estimates are identical, save for the parameters in the variance function. The estimated standard errors do change, however, because in the restricted model above, the Hessian is computed and inverted just for the parameters estimated. In the results below, the Hessian is computed as if the inserted zeros for $\boldsymbol{\gamma}$ were actually the parameter estimates. These standard errors are not useful.

```
Maximum iterations reached. Exit iterations with status=1.
Maxit = 0. Computing LM statistic at starting values.
No iterations computed and no parameter update done.
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                NEWHSAT
LM Stat. at start values         92.77220
LM statistic kept as scalar      LMSTAT
Log likelihood function     -12971.89392
Restricted log likelihood   -13138.97978
Chi squared [   9 d.f.]        334.17171
Significance level                .00000
McFadden Pseudo R-squared       .0127168
Estimation based on N =   6209, K =  19
Inf.Cr.AIC  =25981.788 AIC/N =    4.185
Underlying probabilities based on Logistic
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|    3.02189***      .18716    16.15  .0000     2.65507   3.38871
  FEMALE|    -.31859***      .04747    -6.71  .0000     -.41164   -.22555
  HHNINC|     .23133         .15162     1.53  .1271     -.06584    .52849
  HHKIDS|     .47849***      .05058     9.46  .0000      .37936    .57762
    EDUC|     .10241***      .01246     8.22  .0000      .07798    .12683
```

```
        | Variance function
 MARRIED|        0.0          .02958      .00 1.0000 -.57975D-01  .57975D-01
    UNIV|        0.0          .06508      .00 1.0000 -.12755D+00  .12755D+00
 WORKING|        0.0          .02825      .00 1.0000 -.55371D-01  .55371D-01
  FEMALE|        0.0          .02483      .00 1.0000 -.48663D-01  .48663D-01
  HHNINC|        0.0          .07843      .00 1.0000 -.15372D+00  .15372D+00
        | Threshold parameters for index
   Mu(1)|      .49176***      .06836     7.19  .0000      .35778   .62574
   Mu(2)|     1.26288***      .09719    12.99  .0000     1.07240  1.45336
   Mu(3)|     1.94907***      .11474    16.99  .0000     1.72420  2.17395
   Mu(4)|     2.48180***      .12755    19.46  .0000     2.23181  2.73178
   Mu(5)|     3.48744***      .15442    22.58  .0000     3.18479  3.79010
   Mu(6)|     3.94860***      .16835    23.45  .0000     3.61864  4.27856
   Mu(7)|     4.61859***      .18971    24.35  .0000     4.24677  4.99041
   Mu(8)|     5.70197***      .22651    25.17  .0000     5.25801  6.14592
   Mu(9)|     6.48830***      .25426    25.52  .0000     5.98996  6.98664
-----------------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the estimates for the full heteroscedastic model. The test statistics appear after the estimated parameters.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable              NEWHSAT
Log likelihood function    -12924.94799
Restricted log likelihood  -13138.97978
Chi squared [   9 d.f.]       428.06357
Significance level               .00000
McFadden Pseudo R-squared      .0162898
Estimation based on N =   6209, K =  19
Inf.Cr.AIC  =25887.896 AIC/N =    4.169
Underlying probabilities based on Logistic
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     2.38708***      .14152    16.87  .0000     2.10971  2.66445
  FEMALE|     -.22820***      .03379    -6.75  .0000     -.29442  -.16199
  HHNINC|      .13810         .09576     1.44  .1492     -.04958   .32579
  HHKIDS|      .33481***      .03573     9.37  .0000      .26478   .40485
    EDUC|      .06415***      .00763     8.40  .0000      .04919   .07911
        |Variance function
 MARRIED|     -.13333***      .03198    -4.17  .0000     -.19601  -.07066
    UNIV|     -.19916***      .05658    -3.52  .0004     -.31007  -.08826
 WORKING|     -.18323***      .02928    -6.26  .0000     -.24062  -.12584
  FEMALE|     -.03756         .02478    -1.52  .1296     -.08613   .01101
  HHNINC|     -.19768***      .07590    -2.60  .0092     -.34643  -.04893
        |Threshold parameters for index
   Mu(1)|      .38333***      .05379     7.13  .0000      .27790   .48875
   Mu(2)|      .97539***      .07759    12.57  .0000      .82333  1.12746
   Mu(3)|     1.48986***      .09299    16.02  .0000     1.30761  1.67211
   Mu(4)|     1.88162***      .10423    18.05  .0000     1.67733  2.08590
   Mu(5)|     2.60926***      .12681    20.58  .0000     2.36072  2.85779
   Mu(6)|     2.93848***      .13795    21.30  .0000     2.66810  3.20885
   Mu(7)|     3.41196***      .15468    22.06  .0000     3.10880  3.71512
   Mu(8)|     4.16905***      .18272    22.82  .0000     3.81092  4.52718
   Mu(9)|     4.72049***      .20380    23.16  .0000     4.32105  5.11992
```

```
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

The final results are the test statistics for the hypothesis of homoscedasticity. The three results are, as expected, essentially the same.

```
LM Stat. at start values       92.77220  (from the earlier results)

[CALC] LRTEST  =     93.8918620

WALDSTAT|              1
--------+--------------
       1|       94.6903
```

E35.3.2 Partial Effects in the Heteroscedasticity Model

Partial effects in the ordered choice models with heteroscedasticity appear from two sources, in the latent utility and in the variance function. When variables appear in both places, the total effect is the sum of the two terms.

$$\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{z}_i)}{\partial \mathbf{x}_i} = \left[f\left(a_{j-1,s}\right) - f\left(a_{j,s}\right)\right]\frac{1}{w_i}\boldsymbol{\beta}, \; a_{j,s} = \frac{\mu_{j,s} - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{z}_i)}$$

$$\frac{\partial \text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{z}_i)}{\partial \mathbf{z}_i} = \left[\frac{f\left(a_{j-1,s}\right)a_{j-1,s} - f\left(a_{j,s}\right)a_{j,s}}{F\left(a_{j,s}\right) - F\left(a_{j-1.s}\right)}\right]\mathbf{z}_i.$$

Request the partial effects within the command with

; Partial Effects

(In previous versions, the command was **; Marginal Effects**. This form is still supported.)

The following results show the computation for the full model fit earlier. (Effects for outcomes 0 to 7 are omitted below.)

```
+-------------------------------------------+
| Marginal Effects for OrdLogit             |
| * Total effect = sum of terms             |
+----------+----------+----------+----------+
| Variable | NEWHSA=8 | NEWHS=9  | NEWHS=10 |
+----------+----------+----------+----------+
| FEMALE   |  -.02676 |  -.02181 |  -.02998 |
| HHNINC   |   .01619 |   .01320 |   .01814 |
| HHKIDS   |   .03925 |   .03200 |   .04399 |
| EDUC     |   .00752 |   .00613 |   .00843 |
| MARRIED  |   .01949 |  -.00278 |  -.02676 |
| UNIV     |   .02911 |  -.00415 |  -.03997 |
| WORKING  |   .02678 |  -.00382 |  -.03677 |
| HHNINC   |   .02889 |  -.00412 |  -.03967 |
| FEMALE   |   .00549 |  -.00078 |  -.00754 |
| FEMALE  *|  -.02127 |  -.02260 |  -.03752 |
| HHNINC  *|   .04508 |   .00908 |  -.02153 |
+----------+----------+----------+----------+
```

The **PARTIAL EFFECTS** (or just **PARTIALS**) and **SIMULATE** commands receive the estimates form the heteroscedastic ordered choice model, so you can use them to analyze the probabilities or partial effects. For example, to replace the preceding results, use

PARTIALS ; Effects: female / hhninc ; Outcome = * $

Three differences are first, this estimator uses average partial effects by default (or means if you request them), second, it uses partial differences for dummy variables while the built in computation uses scaled coefficients and, third, as seen below, the **PARTIAL EFFECTS** command produces standard errors and confidence intervals for the partial effects.

```
---------------------------------------------------------------------
Partial Effects  Analysis for Ordered Logit      (Het) Prob[Y = 10]
---------------------------------------------------------------------
Effects on function with respect to FEMALE
Results are computed by average over sample observations
Partial effects for binary var FEMALE   computed by first difference
---------------------------------------------------------------------
df/dFEMALE          Partial    Standard
(Delta method)      Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
APE Prob(y= 0)       .00195     .00148    1.32     -.00096      .00485
APE Prob(y= 1)       .00166     .00075    2.23      .00020      .00312
APE Prob(y= 2)       .00534     .00170    3.14      .00201      .00867
APE Prob(y= 3)       .00959     .00218    4.40      .00532      .01387
APE Prob(y= 4)       .01189     .00210    5.66      .00778      .01601
APE Prob(y= 5)       .03070     .00447    6.87      .02194      .03946
APE Prob(y= 6)       .01222     .00255    4.79      .00721      .01722
APE Prob(y= 7)       .00646     .00381    1.70     -.00100      .01393
APE Prob(y= 8)      -.02026     .00510    3.97     -.03025     -.01027
APE Prob(y= 9)      -.02224     .00323    6.89     -.02857     -.01591
APE Prob(y=10)      -.03732     .00645    5.79     -.04996     -.02468

---------------------------------------------------------------------
Partial Effects  Analysis for Ordered Logit      (Het) Prob[Y = 10]
---------------------------------------------------------------------
Effects on function with respect to HHNINC
Results are computed by average over sample observations
Partial effects for continuous HHNINC   computed by differentiation
Effect is computed as derivative     = df(.)/dx
---------------------------------------------------------------------
df/dHHNINC          Partial    Standard
(Delta method)      Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
APE Prob(y= 0)      -.01302     .00449    2.90     -.02183     -.00421
APE Prob(y= 1)      -.00620     .00215    2.89     -.01041     -.00199
APE Prob(y= 2)      -.01426     .00473    3.01     -.02354     -.00498
APE Prob(y= 3)      -.01675     .00575    2.91     -.02803     -.00547
APE Prob(y= 4)      -.01297     .00544    2.39     -.02362     -.00231
APE Prob(y= 5)      -.00775     .01253     .62     -.03231      .01681
APE Prob(y= 6)       .01008     .00739    1.36     -.00440      .02456
APE Prob(y= 7)       .02766     .01108    2.50      .00593      .04938
APE Prob(y= 8)       .04272     .01395    3.06      .01538      .07006
APE Prob(y= 9)       .01063     .00909    1.17     -.00718      .02845
APE Prob(y=10)      -.02014     .02072     .97     -.06076      .02047
```

E35.4 Sample Selection and Treatment Effects

The following describes an ordered probit counterpart to the standard sample selection model. This is only available for the ordered probit specification. The structural equations are, first, the main equation, the ordered choice model,

$$
\begin{aligned}
y_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ E[\varepsilon_i] = 0,\ \text{Var}[\varepsilon_i] = 1,\\
y_i &= 0 \text{ if } y_i \le \mu_0,\\
&= 1 \text{ if } \mu_0 < y_i \le \mu_1,\\
&= 2 \text{ if } \mu_1 < y_i \le \mu_2,\\
&\ldots\\
&= J \text{ if } y_i > \mu_{J-1}.
\end{aligned}
$$

Second is the selection equation, a univariate probit model,

$$
\begin{aligned}
d_i^* &= \boldsymbol{\alpha}'\mathbf{z}_i + u_i,\\
d_i &= 1 \text{ if } d_i^* > 0 \text{ and } 0 \text{ otherwise},
\end{aligned}
$$

The observation mechanism is

$$[y_i,\mathbf{x}_i] \text{ is observed if and only if } d_i = 1.$$

$$\varepsilon_i,u_i \sim N_2[0,0,1,1,\rho]; \text{ there is 'selectivity' if } \rho \text{ is not equal to zero.}$$

This model is a straightforward generalization of the bivariate probit model with sample selection in Section E33.4.

The treatment effects model includes d_i as an endogenous binary variable in the ordered probit equation;

$$
\begin{aligned}
y_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \gamma d_i + \varepsilon_i,\ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ E[\varepsilon_i] = 0,\ \text{Var}[\varepsilon_i] = 1,\\
y_i &= j \text{ if } \mu_{j-1} < y_i^* \le \mu_j,\ j = 0,1,\ldots,J\\
d_i^* &= \boldsymbol{\alpha}'\mathbf{z}_i + u_i,\\
d_i &= 1 \text{ if } d_i^* > 0 \text{ and } 0 \text{ otherwise},\\
\varepsilon_i,u_i &\sim N_2[0,0,1,1,\rho];\ d_i \text{ is endogenous if } \rho \text{ is not equal to zero.}
\end{aligned}
$$

This model is a generalization of the recursive bivariate probit model in Section E33.6.

E35.4.1 Command

These models require two passes to estimate. In the first, you fit a probit model for the selection (or treatment) variable, *d*. You then pass these values to the ordered probit model using a standard command for this operation, the **; Hold** parameter in the probit command. The two commands would be as follows: (This model is requested in the same fashion as *LIMDEP*'s other sample selectivity models.) Estimate first stage probit model and hold results for next step in the estimation.

PROBIT ; Lhs = d ; Rhs = Z list ; Hold $

Second, estimate the ordered probit model with selectivity

ORDERED ; Lhs = y ; Rhs = X ; ... as usual ; Selection $

You need not make any other changes in the ordered probit command. For the treatment effects case, the probit model is unchanged while the **ORDERED** command becomes

ORDERED ; Lhs = y ; Rhs = X,d ; ... as usual ; Selection ; All $

Note that the treatment variable now appears on the right hand side of the ordered choice model.

The **; Rst = ...** and **; CML:** options for imposing restrictions can be used freely with this model to constrain $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The parameter vector is

$$\boldsymbol{\Theta} = [\beta_1,\ldots,\beta_K,\alpha_1,\ldots,\alpha_L,\mu_1,\ldots,\mu_{J-1},\rho].$$

The usual warning about cross equation restrictions apply. You may also give your own starting values with **; Start = list ...**, though the internal values will usually be preferable.

E35.4.2 Saved Results

All results kept for the basic model are also kept; *b* and *varb* still include only $\boldsymbol{\beta}$, but **; Par** adds all of $[\boldsymbol{\mu},\boldsymbol{\alpha},\rho]$ to the parameter vector. This model adds two additional scalars:

rho = estimate of ρ,
varrho = estimate of asymptotic variance of estimated ρ.

NOTE: The estimates of $\boldsymbol{\alpha}$ update the estimates you stored with **; Hold** when you fit the probit model. Thus, for example, if you were to follow your **ORDERED** command immediately with the identical command, the starting values used for $\boldsymbol{\alpha}$ would be the MLEs from the prior ordered probit command, not the ones from the original probit model that you fit earlier. Also, if you were to follow this model command with a **SELECTION** model command, this estimate of $\boldsymbol{\alpha}$ would be used there, as well.

With the corrected estimates of [β,μ] in hand, predictions for this model are computed in the same manner as for the basic model without selection. The only difference is that no prediction for *y* is computed in the selection model if $d = 0$.

The **PARTIAL EFFECTS** and **SIMULATE** commands are not available for these two specifications (because they only operate on single equation models). An internal program for partial effects is provided. An application below illustrates.

E35.4.3 Applications

To illustrate the computations of this model, we have fit an equation for insurance purchase, then followed with an equation for health satisfaction in which insurance is taken to be a selection mechanism. The treatment effects formulation is shown later.

```
PROBIT      ; Lhs = public ; Rhs = one,age,hhninc,hhkids ; Hold $
ORDERED     ; Lhs = newhsat ; Rhs = one,age,educ,hhninc,female
            ; Selection
            ; Partial Effects $
```

This is the initial probit equation.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                PUBLIC
Log likelihood function      -1868.84461
Restricted log likelihood    -1976.59009
Chi squared [   3 d.f.]        215.49097
Significance level                .00000
McFadden Pseudo R-squared       .0545108
Estimation based on N =   6209, K =   4
Inf.Cr.AIC  = 3745.689 AIC/N =      .603
Results retained for SELECTION model.
Hosmer-Lemeshow chi-squared =  46.95244
P-value=  .00000 with deg.fr. =        8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  PUBLIC|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.24898***      .13551     9.22  .0000      .98339   1.51458
     AGE|     .01695***      .00285     5.96  .0000      .01137    .02253
  HHNINC|   -1.73406***      .12491   -13.88  .0000    -1.97889  -1.48923
  HHKIDS|    -.07027         .04906    -1.43  .1521     -.16643    .02589
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This ordered probit model is fit using the selected observations to obtain starting values for the full model.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable               NEWHSAT
Log likelihood function    -13609.65952
Estimation based on N =   6209, K =  14
Inf.Cr.AIC  =27247.319 AIC/N =    4.388
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.80968***      .11725    23.96  .0000     2.57986   3.03949
     AGE|    -.02310***      .00153   -15.13  .0000     -.02609   -.02011
    EDUC|     .04028***      .00808     4.99  .0000      .02445    .05611
  HHNINC|     .24424***      .08883     2.75  .0060      .07015    .41833
  FEMALE|    -.16710***      .02850    -5.86  .0000     -.22295   -.11124
        |Threshold parameters for index
   Mu(1)|     .20275***      .02260     8.97  .0000      .15846    .24703
   Mu(2)|     .55416***      .02389    23.20  .0000      .50735    .60098
   Mu(3)|     .88530***      .02158    41.03  .0000      .84301    .92759
   Mu(4)|    1.16592***      .01973    59.10  .0000     1.12726   1.20459
   Mu(5)|    1.75777***      .01743   100.82  .0000     1.72360   1.79194
   Mu(6)|    2.04344***      .01695   120.56  .0000     2.01022   2.07667
   Mu(7)|    2.45759***      .01729   142.18  .0000     2.42371   2.49147
   Mu(8)|    3.11320***      .01946   160.01  .0000     3.07507   3.15133
   Mu(9)|    3.53306***      .02325   151.96  .0000     3.48749   3.57863
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the full information maximum likelihood estimate of the full model

```
-----------------------------------------------------------------------------
Ordered Probit Model with Selection.
Dependent variable               NEWHSAT
Log likelihood function    -13607.57507
Restricted log likelihood  -13609.65952
Chi squared [   1 d.f.]          4.16889
Significance level                .04117
McFadden Pseudo R-squared       .0001532
Estimation based on N =   6209, K =  19
Inf.Cr.AIC  =27253.150 AIC/N =    4.389
--------+--------------------------------------------------------------------
  PUBLIC|                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.57206***      .16019    16.06  .0000     2.25809   2.88604
     AGE|    -.01972***      .00194   -10.15  .0000     -.02353   -.01591
    EDUC|     .04014***      .00784     5.12  .0000      .02478    .05550
  HHNINC|    -.06053         .12872     -.47  .6382     -.31282    .19176
  FEMALE|    -.16256***      .02716    -5.99  .0000     -.21579   -.10933
```

```
        |Threshold parameters for index
   Mu(1)|     .19073***      .02687      7.10  .0000       .13807     .24340
   Mu(2)|     .52241***      .04182     12.49  .0000       .44044     .60437
   Mu(3)|     .83633***      .05229     15.99  .0000       .73385     .93881
   Mu(4)|    1.10353***      .06012     18.35  .0000       .98569    1.22137
   Mu(5)|    1.67048***      .07410     22.54  .0000      1.52524    1.81572
   Mu(6)|    1.94557***      .07952     24.47  .0000      1.78972    2.10142
   Mu(7)|    2.34576***      .08663     27.08  .0000      2.17597    2.51554
   Mu(8)|    2.98257***      .09539     31.27  .0000      2.79561    3.16953
   Mu(9)|    3.39287***      .09921     34.20  .0000      3.19843    3.58731
        |Selection equation
Constant|    1.33407***      .13228     10.09  .0000      1.07481    1.59333
     AGE|     .01525***      .00287      5.32  .0000       .00963     .02087
  HHNINC|   -1.72207***      .09850    -17.48  .0000     -1.91514   -1.52901
  HHKIDS|    -.10648**       .04594     -2.32  .0205      -.19653    -.01643
        |Cor[u(probit),e(ordered probit)]
Rho(u,e)|     .50973***      .14253      3.58  .0003       .23038     .78908
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The FIML results provide two test statistics for 'selectivity.' The z statistic on the estimate of ρ is 3.58, which is well over the critical value of 1.96. The likelihood ratio test can be carried out using the initial results for the full model. The restricted value in

```
Log likelihood function    -13607.57507
Restricted log likelihood  -13609.65952
```

is based on the separate probit and ordered probit equations, which corresponds to the model with $\rho = 0$. The LR statistic would be 2(-13607.57507 - (-13609.65952) = 4.169. The critical chi squared with one degree of freedom would be 3.84, so the null hypothesis is rejected again.

A table of partial effects for the conditional model is produced for each outcome. Only the last one is shown here.

```
-----------------------------------------------------------------------------
Partial effects of variables on P[NEWHSAT  = 10|PUBLIC   = 1]
--------+--------------------------------------------------------------------
  PUBLIC|     Partial      Standard            Prob.      95% Confidence
 NEWHSAT|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Direct partial effect in ordered choice equation
     AGE|    -.00245***      .00033     -7.45  .0000      -.00310    -.00181
    EDUC|     .00499***      .00104      4.82  .0000       .00296     .00702
  HHNINC|    -.00753         .01591      -.47  .6360      -.03872     .02365
  FEMALE|    -.02022***      .00367     -5.52  .0000      -.02741    -.01304
        |Indirect partial effect in sample selection equation
     AGE|     .00052***      .00016      3.19  .0014       .00020     .00084
  HHNINC|    -.05896***      .01285     -4.59  .0000      -.08414    -.03378
  HHKIDS|    -.00365**       .00169     -2.16  .0307      -.00695    -.00034
        |Full partial effect = direct effect + indirect effect
     AGE|    -.00193***      .00046     -4.17  .0000      -.00284    -.00102
  HHNINC|    -.06649**       .02627     -2.53  .0114      -.11799    -.01499
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The treatment effects model is obtained by adding *public* to the **; Rhs** specification in the **ORDERED** command and **; All** to the command.

```
-----------------------------------------------------------------------------
Treatment Effects Model: Treatment=PUBLIC
Dependent variable              NEWHSAT
Log likelihood function    -14765.42035
Restricted log likelihood  -14770.39033
Chi squared [   1 d.f.]         9.93996
Significance level               .00162
McFadden Pseudo R-squared      .0003365
Estimation based on N =   6209, K =  20
Inf.Cr.AIC  =29570.841 AIC/N =    4.763
Model estimated: Jun 18, 2011, 15:38:04
--------+--------------------------------------------------------------------
  PUBLIC|                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.27014***      .22312    10.17  .0000     1.83283   2.70746
     AGE|    -.02027***      .00154   -13.13  .0000     -.02330   -.01724
    EDUC|     .03917***      .00692     5.66  .0000      .02561    .05273
  HHNINC|     .06610         .09022      .73  .4638     -.11072    .24292
  FEMALE|    -.14568***      .02612    -5.58  .0000     -.19687   -.09450
  PUBLIC|     .34172**       .13586     2.52  .0119      .07544    .60801
        |Threshold parameters for index
   Mu(1)|     .19408***      .02587     7.50  .0000      .14337    .24479
   Mu(2)|     .52700***      .03637    14.49  .0000      .45572    .59828
   Mu(3)|     .85528***      .04110    20.81  .0000      .77471    .93584
   Mu(4)|    1.13190***      .04397    25.74  .0000     1.04573   1.21808
   Mu(5)|    1.70234***      .04863    35.01  .0000     1.60703   1.79766
   Mu(6)|    1.97911***      .05078    38.98  .0000     1.87959   2.07864
   Mu(7)|    2.38797***      .05406    44.17  .0000     2.28201   2.49393
   Mu(8)|    3.02974***      .05925    51.13  .0000     2.91361   3.14587
   Mu(9)|    3.45667***      .06272    55.12  .0000     3.33375   3.57959
        |Index function for probit equation
Constant|    1.26527***      .13081     9.67  .0000     1.00889   1.52164
     AGE|     .01641***      .00282     5.83  .0000      .01090    .02193
  HHNINC|   -1.68223***      .10083   -16.68  .0000    -1.87986  -1.48459
  HHKIDS|    -.09807**       .04589    -2.14  .0326     -.18802   -.00812
        |Cor[u(probit),e(ordered probit)]
Rho(1,2)|     .41059***      .08110     5.06  .0000      .25164    .56955
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E35.4.4 Technical Details for the Selection Model

In the sample selection model, $[\varepsilon,u]$ are assumed to have a bivariate standard normal distribution with correlation ρ. Then, the probabilities in the log likelihood are:

For observations with $d_i = 0$, Prob = Prob$[d = 0]$ = univariate normal CDF.

For observations with $d_i = 1$, Prob = Prob$[y_i^*$ in particular range and $d = 1 \mid \rho]$

= bivariate normal probability.

The log likelihood for the model with sample selection is

$$\log L = \Sigma_{d=0} \log \Phi(-\boldsymbol{\alpha}'\mathbf{x}_2) + \Sigma_{d=1} \log \{\Phi_2[a_j,\boldsymbol{\alpha}'\mathbf{z},\rho] - \Phi_2[a_{j-1},\boldsymbol{\alpha}'\mathbf{z},\rho]\}$$

where $\Phi(\bullet)$ = standard normal CDF,

$\Phi_2(\bullet,\bullet,\bullet)$ = bivariate standard normal CDF,

$a_j = \mu_j - \boldsymbol{\beta}'\mathbf{x}$,

$a_{j-1} = \mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}$,

and j = the value taken by y_i for that observation.

The same convention used above is maintained for the μs. The first derivatives are tedious but straightforward. They can be derived by applying the formulas given in Chapter E33 for the bivariate probit model. The derivation is a bit simpler here because for the differentiation of the bivariate CDF, q_1 and q_2 are both +1.

The second step reestimates $\boldsymbol{\alpha}$ from the probit model along with $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$, obtaining a FIML set of estimates for all parameters including ρ. The ordered probit command results in two full rounds of estimation. In the first round, the model is estimated as if there were no selection. This provides the remaining starting values. The starting value for ρ is zero. Then, in the second round, the FIML estimates are computed. This model is rather difficult to estimate, and it is best to allow *LIMDEP* to use its own starting values. (In spite of this, nonconvergence can be a problem. When problems arise, be sure first to check the scaling of the independent variables.)

NOTE: This model is *not* fit by computing a 'lambda' variable, $\lambda_i = \phi(\boldsymbol{\alpha}'\mathbf{z}_i)/\ \Phi(\boldsymbol{\alpha}'\mathbf{z}_i)$ from the results of the first step probit and including it in the ordered probit at the second. It is estimated by maximizing the likelihood function above with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and ρ. There will be no coefficient shown for such a variable in the estimation results, though the estimated ρ is shown.

NOTE: (This is another frequently asked question.) All observations in the sample are used in fitting this model, not just the ones for which $d = 1$. The observations for which $d = 0$ contribute to the probit part of the log likelihood. The remainder contribute both to the probit and the ordered probit.

The treatment effects model is developed in exactly the same steps as the recursive bivariate probit model in Section E33.6. The relevant probabilities that enter the log likelihood are

$$\log L_i \quad = \log [\text{Prob}(d_i = 1 \text{ or } 0) \times \text{Prob}(y_i = j \mid d_i = 1 \text{ or } 0)$$

which, once again, is simply the joint probability. Thus, the log likelihood function has terms

$$\Sigma_{\textit{all observations}} \log \{\Phi_2[(a_j - \gamma d_i), q_i\boldsymbol{\alpha}'\mathbf{z}, q_i\rho] - \Phi_2[a_{j-1} - \gamma d_{i,}\ q_i\boldsymbol{\alpha}'\mathbf{z}, q_i\rho]\}$$

where $q_i = 2d_i - 1 = -1\ (+1)$ when $d_i = 0\ (1)$.

E35.5 Generalized Ordered Choice and Parallel Regressions

Two specification questions that bear directly on the model discussed to this point are the 'proportional odds' assumption and the 'parallel regressions' assumption. We consider them in turn.

E35.5.1 The Proportional Odds Assumption

The proportional odds assumption is imposed (only) by the ordered logit model. If

$$\text{Prob}[y_i = j] = \Lambda(\mu_j - \boldsymbol{\beta}'\mathbf{x}) - \Lambda(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x})$$

Then,

$$\text{Prob}[y_i \le j] = \Lambda(\mu_j - \boldsymbol{\beta}'\mathbf{x})$$

Using the simple algebra of the logit model, it follows that

$$\log\{\text{Prob}[y_i \le j] / \text{Prob}[y_i > j]) = \mu_j - \boldsymbol{\beta}'\mathbf{x}_i.$$

This is known as the proportional odds assumption, and it is viewed as a restrictive assumption of the model. The implication is that the log-odds for any outcome, μ_j, differs from that for any other only by a constant, j. A number of alternative tests and assumptions have been proposed to relax the 'restriction.' In point of fact, the researcher bound by this restriction is a prisoner of the logistic assumption to begin with. It does not apply to any other model that we have considered, so the simple expedient to pursue if this assumption is viewed as problematic is to switch to an ordered probit model. But, this is merely a question of functional form, and the probit model may be no less 'restrictive' in this regard than the logit model. There is, however, a more substantive issue to consider. Whether the proportional odds assumption imposes a restriction on behavior is at least conceivable. The question is not unrelated to that of the 'independence from irrelevant alternatives' implication of the logit model in the discrete choice framework. That question seems at least ambiguous here – whether an assumption about the log odds translates backwards into an assumption about behavior seems at least uncertain.

We note, some authors have advocated abandoning the latent regression model altogether, in some cases, in favor of a multinomial logit model for the J+1 outcomes. By this prescription, one loses the ordered nature of the data, which could be argued to be at higher cost than the initial assumption that the same parameter vector, $\boldsymbol{\beta}$, appears in the probabilities of all J+1 outcomes to begin with. This issue is not settled in the literature, and can't be resolved here. We conclude at this point only that the alternatives that have been suggested can all be fit with *LIMDEP*, using other modeling frameworks.

E35.5.2 Brant Test of the Parallel Regressions Assumption

The 'parallel regressions' assumption, such as it is, also characterizes the ordered logit model, but no other functional form. The assumption, itself, is a curious one. The term appears to have gotten some impetus from a frequently cited 'result' for the ordered probit model in Long (1997) which states, for, say, a five outcome model, that

$$\frac{\partial \Pr[y \le 1 | x]}{\partial x} = \frac{\partial \Pr[y \le 2 | x]}{\partial x} = \frac{\partial \Pr[y \le 3 | x]}{\partial x}$$

at any trio of points, x_1, x_2 and x_3, at which all three probabilities are equal. 'It is this sense that the regression curves are parallel.' (Long (1997), page 141) The implication that the 'regression curves' implied by the model are parallel is cited as a significant restriction of the ordered choice model. In fact, these are not 'regressions' in any sense – they are not conditional mean functions. But, that is merely terminology. In this model, *some things* are parallel. The so called parallel regressions restriction has attracted some attention. A familiar 'test' for the assumption is the Brant test. Arguably, the Brant test is a test of preference homogeneity. It does provide an interesting specification test for the model – the preference heterogeneity implication of the model is testable. What rejection of the hypothesis then suggests is less than obvious, however.

The Brant test of homogeneity is constructed as follows: According to the ordered logit model,

$$\text{Prob}(y_i > j \mid \mathbf{x}_i) = \Lambda[\boldsymbol{\beta}'\mathbf{x}_i - \mu_j],\ j = 0, 1, ..., J\text{-}1.$$

If we define $z_{ij} = 1[y_i > j]$, then this defines a simple binary logit model for the J-1 binary outcomes, z_{ij}. The force of the restrictions of the model is that each such probability 'model' has the same coefficient vector, $\boldsymbol{\beta}$, though each has its own constant term. Define these as $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$,... The test is carried out by constructing a Wald test of the null hypothesis that $\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\beta}_0 - \boldsymbol{\beta}_2 = \mathbf{0}$, etc. Note that the model does not imply that the constant terms are the same. We will return to this detail later. To carry out the test, we compute the J-1 binary logit models, and obtain $\mathbf{b}_0,...,\mathbf{b}_{J\text{-}1}$. With each coefficient vector, we compute the predicted probabilities, $p_{ij} = \exp(\mathbf{b}_j'\mathbf{x}_i)/[1+\exp(\mathbf{b}_j'\mathbf{x}_i)]$ for the sample and the quantities $w_{imj} = p_{ij} = p_{im}p_{ij}$. The moment matrix $\mathbf{V}_{mj} = \Sigma_i\, w_{imj}\mathbf{x}_i\mathbf{x}_i'$ is computed, where $\mathbf{x}_i$ includes the constant term. The matrix,

$$\mathbf{A}_{mj} = \mathbf{V}_{mm}^{-1}\, \mathbf{V}_{mj}\, \mathbf{V}_{jj}^{-1},$$

estimates the asymptotic covariance of $\mathbf{b}_m$ and $\mathbf{b}_j$. The row and column corresponding to the constant term are removed. Then, the covariance matrix, $\mathbf{V}$, for the set of estimates $\mathbf{b} = [\mathbf{b}_0',\mathbf{b}_1',...,\mathbf{b}_{J\text{-}1}']'$ is assembled in partitioned form using the blocks defined above (and their transposes). The Wald test of the homogeneity restriction is carried out using the chi squared statistic,

$$Wald = (\mathbf{Db})'\,[\mathbf{DVD}']^{-1}(\mathbf{Db}),$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} & \mathbf{-I} & \mathbf{...} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{-I} & \mathbf{...} \\ \mathbf{...} & \mathbf{...} & \mathbf{...} & \mathbf{...} \\ \mathbf{I} & \mathbf{0} & \mathbf{...} & \mathbf{-I} \end{bmatrix}.$$

The statistic has a limiting chi squared distribution with degrees of freedom equal to $(J\text{-}1)K_1$ where K_1 is the number of independent variables in the model, not including the constant term. (E.g., if the original problem has y taking values 0,1,2,3,4, then we will compute $J = 4$ logit coefficient vectors, $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$, and (**J** - 1) contrasts, $\mathbf{b}_0 - \mathbf{b}_1$, $\mathbf{b}_0 - \mathbf{b}_2$ and $\mathbf{b}_0 - \mathbf{b}_3$.).

The Wald statistic is an omnibus test for homogeneity of the entire coefficient vector. It is possible that some of the coefficients are (or appear to be) homogeneous while others are not. The test can be carried out coefficient by coefficient by isolating just one of the contrasts in the set of K_1. Define the matrix **C** with J-1 rows and $(J\text{-}1)K_1$ columns. The row is a set of J-1 row vectors. There is a single one in the kth position of the jth row vector in the jth row of **C**, and zeros elsewhere Then, the chi squared for the kth coefficient is

$$Wald_k = [\mathbf{C}(\mathbf{Db})]'[\mathbf{C}(\mathbf{DVD}')\mathbf{C}]^{-1}[\mathbf{C}(\mathbf{Db})].$$

This statistic has a limiting chi squared distribution with J-1 degrees of freedom.

Application of the Brant Test

The Brant test is automated in *LIMDEP*. You need only add

; Brant test

to your **ORDERED ; Logit** command. The full set of results is computed and reported.

We have applied this test to the treatment effects model fit earlier, while treating *public* as exogenous. The following results are obtained.

```
+------------------------------------------------+
| Brant specification test for equal coefficient |
| vectors in the ordered prob. model. The model  |
| implies that logit[Prob(y>j|x)]=beta(j)*x - mj |
| for all j = 0,..., 9. The chi squared test is  |
| H0:beta(0) = beta(1) = ... beta( 9)            |
| Chi squared test statistic =    236.72126      |
| Degrees of freedom         =     45            |
| P value                    =       .00000      |
+------------------------------------------------+

===========================================================================
Specification Tests for Individual Coefficients in Ordered Logit Model
(Note, Coefficients for values beyond y = 5 are not reported.)
Degrees of freedom for each of these tests is  9
===========================================================================
        |   Brant Test    | Coefficients in implied model Prob(y > j).    |
Variable| Chi-sq  P value |   0   |   1   |   2   |   3   |   4   |   5   |
AGE     |  30.71   .00033 | -.0324| -.0297| -.0369| -.0379| -.0406| -.0315|
EDUC    |  48.26   .00000 |  .4105|  .3437|  .1978|  .0776|  .0540|  .0804|
HHNINC  |  66.60   .00000 | 1.3013| 1.2476|  .6096|  .7653|  .7257|  .8924|
FEMALE  |  41.78   .00000 |  .2032|  .1441| -.0715| -.1607| -.1800| -.3864|
PUBLIC  |  63.03   .00000 | -.7669| -.9268|-1.8901| -.9694| -.6624| -.5484|
```

Based on these results, the null hypothesis is rejected. The results for the individual coefficients suggest that the hypothesis is rejected for all the individual coefficients as well.

The Brant test can be adapted to the normal distribution (ordered probit model). Two changes are required. First, in the procedure, itself, we would fit probit models rather than logit models. Second, we must change the computation of the parts of the asymptotic covariance matrix to conform to the probit model. To do this, we will use the BHHH estimator. The MLE of the parameter vector in each regression is written $(\mathbf{b}_j - \boldsymbol{\beta}) = \mathbf{H}_j^{-1}\mathbf{g}_j$ where $\mathbf{H}_j$ is the Hessian and $\mathbf{g}_j$ is the first derivatives vector. We then use the information matrix equality to invoke the BHHH estimator for the asymptotic variance of $\mathbf{b}_j$ which we write $(\mathbf{G}_j'\mathbf{G}_j)^{-1}$. For the asymptotic covariances, we once again invoke the information matrix equality, and the estimator of $\mathbf{H}_j$, which produces a sandwich style estimator,

$$\text{Est.Asy.Cov}[\mathbf{b}_j,\mathbf{b}_m] = (\mathbf{G}_j'\mathbf{G}_j)^{-1} (\mathbf{G}_j'\mathbf{G}_m) (\mathbf{G}_m'\mathbf{G}_m)^{-1}.$$

The remaining detail is how to compute the rows of $\mathbf{G}_j$. For the probit model, the relevant derivative is

$$\partial \log P_{ji} / \partial \boldsymbol{\beta}_j = (2z_{ji}-1)\, \phi[\boldsymbol{\beta}_j'\mathrm{x}_i] / \Phi[(2z_{ji}-1)\, \boldsymbol{\beta}_j'\mathbf{x}_i]\, \mathbf{x}_i.$$

LIMDEP detects this internally and adjusts the computations. For the earlier example, the automatically generated results are as follows:

```
+------------------------------------------------+
| Brant specification test for equal coefficient |
| vectors in the ordered probit model. The model |
| implies that normit[Prb(y>j|x)]=beta(j)*x - mj |
| for all j = 0,..., 9. The chi squared test is  |
| H0:beta(0) = beta(1) = ... beta( 9)            |
| Chi squared test statistic =    200.97546      |
| Degrees of freedom         =     45            |
| P value                    =      .00000       |
+------------------------------------------------+

===========================================================================
Specification Tests for Individual Coefficients in Ordered Logit Model
(Note, Coefficients for values beyond y = 5 are not reported.)
Degrees of freedom for each of these tests is  9
===========================================================================
        |   Brant Test    | Coefficients in implied model Prob(y > j).    |
Variable| Chi-sq  P value |   0   |   1   |   2   |   3   |   4   |   5   |
AGE     |  36.28   .00004 | -.0126| -.0123| -.0170| -.0191| -.0219| -.0191|
EDUC    |  34.75   .00007 |  .1508|  .1360|  .0869|  .0350|  .0267|  .0457|
HHNINC  |  40.90   .00001 |  .4913|  .5027|  .2817|  .3933|  .4139|  .5594|
FEMALE  |  38.25   .00002 |  .0576|  .0362| -.0485| -.0945| -.1093| -.2371|
PUBLIC  |  42.63   .00000 | -.2887| -.3505| -.7556| -.4342| -.3220| -.3090|
```

The results are consistent with those for the ordered logit model, which might be expected.

E35.6 Generalized Ordered Choice Models

The preceding notwithstanding, researchers have devoted considerable attention to restructuring the ordered choice model to redeem it from the objectionable result noted above. We note, first, the latent regressions, ordered logit model already analyzed here implies the preference structure,

$$\text{Prob}[y_i \leq j \mid \mathbf{x}_i] = \Lambda[\mu_j - \boldsymbol{\beta}' \mathbf{x}_i].$$

The parallel regressions assumption is that the same $\boldsymbol{\beta}$ appears in every equation. The generalized ordered logit model suggested, e.g., by Williams (2006) is (using our normalizations)

$$\text{Prob}[\, y_i \leq j \mid \mathbf{x}_i \,] = \Lambda[\, \mu_j - \boldsymbol{\beta}_j' \mathbf{x}_i], j = 1,2,...,J.$$

Versions of this model appear in a number of publications. This specification has two major flaws. First, there is no parametric restriction, other than the one we seek to avoid to begin with ($\boldsymbol{\beta}_j = \boldsymbol{\beta}$ for all j) that can be used to make the probabilities of the $J + 1$ outcomes sum to one. The model is internally inconsistent unless each outcome is viewed as a model in its own right – a peculiar assumption about the distribution of preferences across individuals. Worse, for the interior outcomes of the dependent variable (i.e., not zero and not J), the probability is

$$\text{Prob}[y_i = j \mid \mathbf{x}_i] = \Lambda[\, \mu_j - \boldsymbol{\beta}_j'\mathbf{x}_i] - \Lambda[\, \mu_{j-1} - \boldsymbol{\beta}_{j-1}' \mathbf{x}_i]$$

a difference which cannot be forced even to be positive. For any $\boldsymbol{\beta}_j$ and $\boldsymbol{\beta}_{j-1}$, whether or not this difference is positive will be data dependent, and if there is more than one variable in $\mathbf{x}_i$, would be pure luck as much as anything else. The model is not an internally consistent probability model defined over an outcome space.

The difficulty being dealt with here ultimately arises from an assumption that the coding of the dependent variable in the model is structural. Why the observed respondent should have preferences that are structurally defined in terms of the coding of the survey is difficult to fathom. For example, in the simplest imaginable cases, it is difficult to see why the preference orderings of respondents should be functions of whether the surveyor presents them with a three point or a five point scale. In more general terms, in the generalized ordered logit model, the parameter vector seems to be a function of the dependent variable. This is unlike the multinomial logit model, in which the multiplicity of parameter vectors is merely the parameterization of J distinct utility functions defined across J alternatives. Here, each parameter vector is identified with a different response to the same question. It is unclear how one should interpret such a structure.

There is another aspect of this construction that suggests the ambiguity of the model. It is not possible to simulate data that correspond to the assumptions of the model. In the ordered probit model, with $\mathbf{x}$, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ known, in order to simulate a draw on y, we would compute $\boldsymbol{\beta}'\mathbf{x}$, draw a random normal value ε, compute $y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$, then see which interval, $(-\infty,0)$, $(0,\mu_1)$, (μ_1,μ_2) etc. contains y^* to produce $y = 0$ or 1 or 2, etc. This is not possible for the 'generalized model' suggested here, because one needs to know y in order to compute $\boldsymbol{\beta}_j'\mathbf{x}$ (you need to know which $\boldsymbol{\beta}$ to use) so the outcome has to be known before ε is even drawn. This is the implication of the internal incoherency of the model.

E35.7 Hierarchical Ordered Probit Models

The hierarchical ordered probit model (or generalized ordered probit model) is a univariate ordered probit model in which the threshold parameters depend on variables. (We opt for the acronym HOPIT model as slightly more melodious than GOPIT. In the original proposal of this model (Pudney and Shields (2000)), the thresholds were modeled as linear functions of the data, producing the model

$$\begin{aligned} y^* &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon \\ y &= 0 \text{ if } y^* \leq 0, \\ &= 1 \text{ if } 0 < y^* \leq \mu_1, \\ &= 2 \text{ if } \mu_1 < y^* \leq \mu_2, \\ &\ldots \\ \mu_j &= \boldsymbol{\delta}_j'\mathbf{z}. \end{aligned}$$

(There is no disturbance on the equation for the threshold variables.) The model has an inherent identification problem, because in

$$\text{Prob}[y = j] = \Phi(\mu_j - \boldsymbol{\beta}'\mathbf{x}) - \Phi(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}),$$

if **x** and **z** have variables in common, then (with a sign change) the same model is produced whether the common variable appears in μ_j or $\boldsymbol{\beta}'\mathbf{x}$. (Pudney and Shields note and discuss this.) The *LIMDEP* implementation avoids this indeterminacy by using a different functional form. (That does imply that we achieve identification through functional form.)

Two forms of the model are provided.

$$\text{Form 1:}\ \mu_j = \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z})$$

$$\text{Form 2:}\ \mu_j = \exp(\theta_j + \boldsymbol{\delta}_\mathbf{j}'\mathbf{z})$$

Note that in form 1, each μ_j has a different constant term, but the same coefficient vector, while in form 2, each threshold parameter has its own parameter vector. (We note, for purposes of estimation, it is always necessary for μ_j to be greater than μ_{j-1}. We are able to impose that on form 1 fairly easily by parameterizing θ_j in a way that does so. However, for form 2, this is much more difficult to obtain, and users should expect to see diagnostics about unordered thresholds when they use form 2.) The threshold coefficients will be difficult to compare between the original ordered probit model and form 2 of the HOPIT model. For form 1, the model reverts to the unmodified ordered probit model if the single vector $\boldsymbol{\delta}$ equals **0**.

The command for this model augments the usual ordered probit command with the specification for the thresholds,

```
ORDERED     ; Lhs = ... ; Rhs = ...
            ; HO1 = list of variables or ; HO2 = list of variables $
```

The list of variables in the HO1 or HO2 part must not contain a constant term (*one*). All other options for the ordered probit model are exactly as described previously, including fitted values, restrictions, marginal effects, and so on, unchanged. *This form of the ordered probit model can also be combined with the sample selection corrected ordered probit model described in Section E35.4.*

In the example below, the model is first fit to the health satisfaction variable with no modification to the thresholds. In the HOPIT model fit next, the thresholds vary with whether or not the family has kids in the household and with the number of types of insurance they have. For purpose of a limited example, we use a subset of the sample.

```
SAMPLE      ; All $
CREATE      ; insuranc = public + addon $
ORDERED     ; Lhs = hsat ; Rhs = one,age,educ,female,hhninc
            ; Partial Effects $
ORDERED     ; Lhs = hsat ; Rhs = one,age,educ,female,hhninc
            ; HO1 = hhkids,insuranc
            ; Partial Effects $
```

These are the estimates for the base case. (We have omitted the partial effects.)

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                   HSAT
Log likelihood function    -56876.85183
Restricted log likelihood  -57836.42214
Chi squared [   4 d.f.]      1919.14061
Significance level               .00000
McFadden Pseudo R-squared      .0165911
Estimation based on N =  27326, K =  14
Inf.Cr.AIC  =********* AIC/N =    4.164
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.68410***      .04392    61.12  .0000     2.59802   2.77018
     AGE|    -.02096***      .00056   -37.71  .0000     -.02205   -.01987
    EDUC|     .03341***      .00284    11.76  .0000      .02784    .03898
  FEMALE|    -.05800***      .01259    -4.61  .0000     -.08268   -.03332
  HHNINC|     .26478***      .03631     7.29  .0000      .19362    .33594
        |Threshold parameters for index
   Mu(1)|     .19340***      .01002    19.30  .0000      .17376    .21305
   Mu(2)|     .49929***      .01087    45.93  .0000      .47799    .52060
   Mu(3)|     .83548***      .00990    84.39  .0000      .81608    .85489
   Mu(4)|    1.10462***      .00908   121.63  .0000     1.08682   1.12242
   Mu(5)|    1.66162***      .00801   207.44  .0000     1.64592   1.67732
   Mu(6)|    1.93021***      .00774   249.46  .0000     1.91504   1.94537
   Mu(7)|    2.33753***      .00777   300.92  .0000     2.32230   2.35275
   Mu(8)|    2.99283***      .00851   351.70  .0000     2.97615   3.00951
   Mu(9)|    3.45210***      .01017   339.31  .0000     3.43216   3.47204
--------+--------------------------------------------------------------------
```

These are the estimates for the HO1 hierarchical model.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                   HSAT
Log likelihood function    -56868.23498
Restricted log likelihood  -57836.42214
Chi squared [   4 d.f.]      1936.37431
Underlying probabilities based on Normal
HOPIT (covariates in thresholds) model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.66036***      .04828    55.10  .0000     2.56573   2.75499
     AGE|    -.02035***      .00058   -35.09  .0000     -.02149   -.01921
    EDUC|     .03313***      .00293    11.30  .0000      .02738    .03887
  FEMALE|    -.06072***      .01259    -4.83  .0000     -.08539   -.03606
  HHNINC|     .26373***      .03648     7.23  .0000      .19222    .33523
        |Estimates of t(j) in mu(j)=exp[t(j)+d*z]
Theta(1)|   -1.62461***      .06134   -26.49  .0000    -1.74484  -1.50439
Theta(2)|    -.67653***      .03254   -20.79  .0000     -.74029   -.61276
Theta(3)|    -.16186***      .02193    -7.38  .0000     -.20485   -.11888
Theta(4)|     .11739***      .01750     6.71  .0000      .08309    .15170
Theta(5)|     .52583***      .01258    41.79  .0000      .50117    .55049
Theta(6)|     .67578***      .01122    60.25  .0000      .65379    .69776
Theta(7)|     .86747***      .00979    88.62  .0000      .84828    .88665
Theta(8)|    1.11497***      .00843   132.20  .0000     1.09844   1.13150
Theta(9)|    1.25794***      .00787   159.74  .0000     1.24250   1.27337
        |Threshold covariates mu(j)=exp[t(j)+d*z]
  HHKIDS|    -.01830***      .00526    -3.48  .0005     -.02862   -.00799
INSURANC| .15082D-04**    .5872D-05     2.57  .0102  .35726D-05  .26592D-04
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
(Partial effects for outcomes 0-9 are omitted.)
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
    HSAT|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=10] at means]--------------
     AGE|    -.00377***    -1.52276   -11.54  .0000     -.00441   -.00313
    EDUC|     .00614***      .64474     9.12  .0000      .00482    .00746
 *FEMALE|    -.01123       -.10424     -.50  .6182     -.05541    .03294
  HHNINC|     .04887***      .15964     3.51  .0004      .02161    .07613
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E35.8 Zero Inflated Ordered Probit (ZIOP, ZIHOP) Models

Harris and Zhao (2007) have developed a zero inflated ordered probit (ZIOP) counterpart to the zero inflated Poisson model. The ZIOP formulation would appear

$$
\begin{aligned}
d^* &= \boldsymbol{\alpha}'\mathbf{w} + u, & d &= 1\,(d^* > 0)\\
y^* &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon, & y &= 0 \text{ if } y^* \le 0 \text{ or } d = 0 \leftarrow\\
& & &\ 1 \text{ if } 0 < y^* \le \mu_1 \text{ and } d = 1,\\
& & &\ 2 \text{ if } \mu_1 < y^* \le \mu_2 \text{ and } d = 1,\\
& & &\ \text{and so on.}
\end{aligned}
$$

The first equation is assumed to be a probit model (based on the normal distribution) – this estimator does not support a logit formulation. The correlation between u and ε is ρ, which by default equals zero, but may be estimated instead. The latent class nature of the formulation has the effect of inflating the number of observed zeros, even if u and ε are uncorrelated. The model with correlation between u and ε is an optional specification that analysts might want to test. The zero inflation model may also be combined with the hierarchical (generalized) model discussed in the previous section. Thus, it might also be specified as part of the model that

$$
\begin{aligned}
\text{Form 1:}\ \mu_j &= \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z})\\
\text{Form 2:}\ \mu_j &= \exp(\theta_j + \boldsymbol{\delta}_j'\mathbf{z})
\end{aligned}
$$

The command structure for ZIOP and ZIHOP models are

```
PROBIT      ; Lhs = d ; Rhs = variables in w ; Hold $
ORDERED     ; Lhs = y ; Rhs = variables in x
            ; ZIOP $
```

This form of the model imposes $\rho = 0$. To allow the correlation to be a free parameter, add

```
            ; Correlation
```

to the command.

NOTE: The **; HO1** and **; HO2** specifications discussed in the preceding section may also be used with this model.

In the example below, we continue the analysis of the health care data. The (artificial) model has the zero inflation probability based on the presence of ‘public’ insurance while the ordered outcome continues to be the self reported health satisfaction. Here, we have used the entire sample of 27,236 observations.

The commands are:

```
SAMPLE      ; All $
PROBIT      ; Lhs = public
            ; Rhs = one,age,hhninc,hhkids,married ; Hold $
ORDERED     ; Lhs = hsat
            ; Rhs = one,age,educ,female
            ; ZIO ; Correlated $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                PUBLIC
Log likelihood function      -9229.32605
Restricted log likelihood    -9711.25153
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  PUBLIC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.51862***      .05021    30.25  .0000     1.42022   1.61702
     AGE|     .00553***      .00105     5.26  .0000      .00347    .00759
  HHNINC|   -1.55524***      .05120   -30.37  .0000    -1.65560  -1.45489
  HHKIDS|    -.08320***      .02370    -3.51  .0004     -.12966   -.03675
 MARRIED|     .10035***      .02694     3.72  .0002      .04754    .15316
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------


-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                  HSAT
Log likelihood function     -56903.42663
Restricted log likelihood   -57836.42214
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.70343***      .04379    61.73  .0000     2.61760   2.78926
     AGE|    -.02078***      .00056   -37.41  .0000     -.02186   -.01969
    EDUC|     .03881***      .00274    14.16  .0000      .03344    .04419
  FEMALE|    -.05742***      .01259    -4.56  .0000     -.08210   -.03274
        |Threshold parameters for index
   Mu(1)|     .19279***      .00999    19.29  .0000      .17320    .21238
   Mu(2)|     .49771***      .01085    45.88  .0000      .47645    .51896
   Mu(3)|     .83298***      .00989    84.26  .0000      .81361    .85236
   Mu(4)|    1.10156***      .00907   121.43  .0000     1.08378   1.11934
   Mu(5)|    1.65744***      .00800   207.07  .0000     1.64175   1.67313
   Mu(6)|    1.92551***      .00773   249.00  .0000     1.91036   1.94067
   Mu(7)|    2.33231***      .00776   300.37  .0000     2.31709   2.34753
   Mu(8)|    2.98735***      .00851   351.12  .0000     2.97067   3.00402
   Mu(9)|    3.44694***      .01018   338.75  .0000     3.42700   3.46688
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------
Zero Inflated Ordered Probit Model.
Dependent variable                  HSAT
Log likelihood function    -56895.22719
Restricted log likelihood  -56903.42663
--------+-----------------------------------------------------------------
  PUBLIC|                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------
        |Index function for probability
Constant|    2.77007***      .04944    56.03  .0000     2.67317   2.86697
     AGE|    -.02150***      .00057   -37.68  .0000     -.02262   -.02038
    EDUC|     .03769***      .00284    13.27  .0000      .03212    .04325
  FEMALE|    -.05844***      .01255    -4.66  .0000     -.08304   -.03384
        |Threshold parameters for index
   Mu(1)|     .19868***      .01235    16.08  .0000      .17447    .22289
   Mu(2)|     .50918***      .01694    30.05  .0000      .47597    .54239
   Mu(3)|     .84768***      .01897    44.70  .0000      .81051    .88486
   Mu(4)|    1.11767***      .01978    56.50  .0000     1.07890   1.15644
   Mu(5)|    1.67504***      .02062    81.25  .0000     1.63463   1.71545
   Mu(6)|    1.94359***      .02087    93.15  .0000     1.90269   1.98449
   Mu(7)|    2.35098***      .02119   110.97  .0000     2.30946   2.39251
   Mu(8)|    3.00678***      .02174   138.30  .0000     2.96417   3.04939
   Mu(9)|    3.46677***      .02222   156.00  .0000     3.42322   3.51033
        |Zero inflation probit probability
Constant|    -.30749        1.71064     -.18  .8573    -3.66028   3.04530
     AGE|     .10718         .06555     1.63  .1021     -.02131    .23566
  HHNINC|    -.19155         .62143     -.31  .7579    -1.40954   1.02644
  HHKIDS|    -.59894**       .24410    -2.45  .0141    -1.07737   -.12051
 MARRIED|    1.06982         .94393     1.13  .2571     -.78024   2.91988
        |Cor[u(probit),e(ordered probit)]
Rho(u,e)|    -.90968        1.40561     -.65  .5175    -3.66462   1.84525
--------+-----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------
```

E35.9 Bivariate Ordered Probit and Polychoric Correlation

The bivariate ordered probit model is analogous to the SUR model for the ordered probit case:

$$
\begin{aligned}
y_{ji}^* &= \boldsymbol{\beta}_j'\mathbf{x}_{ji} + \varepsilon_{ji}\\
y_{ji} &= 0 \text{ if } y_{ji}^* \le 0,\\
&\quad 1 \text{ if } 0 < y_{ji}^* < \mu_1,\\
&\quad 2, \ldots \text{ and so on, } j = 1,2,
\end{aligned}
$$

for a pair of ordered probit models that are linked by Cor(ε_{1i},ε_{2i}) = ρ. The model can be estimated one equation at a time using the results described earlier. Full efficiency in estimation and an estimate of ρ are achieved by full information maximum likelihood estimation. *LIMDEP*'s implementation of the model uses FIML, rather than GMM. Either variable (but not both) may be binary. If both are binary, the bivariate probit model should be used. (The development here draws on Butler and Chatterjee (1997) who analyzed maximum likelihood and GMM estimators for the bivariate extension of the ordered probit model.)

The command structure requires prior estimation of the two univariate models to provide starting values for the iterations. The third command then fits the bivariate model. We assume that the first variable is multinomial.

```
ORDERED      ; Lhs = y1 ; Rhs = ... $
MATRIX       ; b1 = b ; mu1 = mu $
```

Use one of the following. If the second variable has more than two outcomes, use

```
ORDERED      ; Lhs = y2 ; Rhs = ... $
MATRIX       ; b2 = b ; mu2 = mu $
```

If the second variable is binary, use

```
PROBIT       ; Lhs = y2 ; Rhs = ... $
MATRIX       ; b2 = b $
```

Then, estimate the bivariate model with

```
ORDERED      ; Lhs = y1,y2 ; Rh1 = ... ; Rh2 = ...
             ; Start = b1,mu1,b2,mu2, 0 $
```

The variable *mu*2 is omitted if *y*2 is binary. The final zero in the list of starting values is for ρ. You may use some other value if you have one.

The standard options for estimation are available (iteration controls, technical output, cluster corrections, etc.). You may also retain fitted values with **; Keep = yf1,yf2** (note that both names are provided). Probabilities for the joint observed outcome are retained with **; Prob = name**. Listings of probabilities for outcomes are obtained with **; List** as usual.

To illustrate the estimator, we use the health care utilization data analyzed earlier. The two outcomes are *y*1 = health care satisfaction, taking values 0 to 5 (we reduced the sample) and *y*2 = the number of types of health care insurance. Results for a bivariate ordered probit model appear below. The initial univariate models are omitted.

```
SAMPLE       ; All $
REJECT       ; newhsat > 5  | _groupti < 7  $
ORDERED      ; Lhs = newhsat ; Rhs = one,age,educ,female,hhninc $
MATRIX       ; b1 = b ; mu1 = mu $
CREATE       ; insuranc = public + addon $
CROSSTAB     ; Lhs = newhsat  ; Rhs = insuranc $
ORDERED      ; Lhs = insuranc ; Rhs = one,age,educ,hhninc,hhkids $
MATRIX       ; b2 = b ; mu2 = mu $
ORDERED      ; Lhs = newhsat,insuranc
             ; Rh1 = one,age,educ,female,hhninc
             ; Rh2 = one,age,educ,hhninc,hhkids
             ; Start = b1,mu1,b2,mu2,0 $
```

```
-----------------------------------------------------------------------------
Bivariate Ordered Probit Model
Dependent variable              BivOrdPr
Log likelihood function      -3099.59435
Restricted log likelihood    -3100.36600
--------+--------------------------------------------------------------------
 NEWHSAT|                  Standard            Prob.      95% Confidence
INSURANC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for Probability Model for NEWHSAT
Constant|     1.98379***       .23742     8.36  .0000     1.51846   2.44913
     AGE|     -.01233***       .00288    -4.28  .0000     -.01797   -.00668
    EDUC|      .01815          .01667     1.09  .2762     -.01452    .05082
  FEMALE|      .09626*         .05301     1.82  .0694     -.00764    .20016
  HHNINC|      .13547          .17765      .76  .4457     -.21271    .48365
        |Index function for Probability Model for INSURANC
Constant|     2.57737***       .38142     6.76  .0000     1.82980   3.32493
     AGE|      .01847***       .00609     3.03  .0024      .00654    .03040
    EDUC|     -.13925***       .02090    -6.66  .0000     -.18022   -.09828
  HHNINC|     -.63131*         .33803    -1.87  .0618    -1.29383    .03121
  HHKIDS|     -.01720          .10527     -.16  .8702     -.22353    .18912
        |Threshold Parameters for Probability Model for NEWHSAT
  MU(01)|      .24263***       .03171     7.65  .0000      .18048    .30479
  MU(02)|      .67851***       .04404    15.41  .0000      .59220    .76483
  MU(03)|     1.15093***       .04917    23.41  .0000     1.05456   1.24730
  MU(04)|     1.61433***       .05193    31.09  .0000     1.51255   1.71611
        |Threshold Parameters for Probability Model for INSURANC
LMDA(01)|     4.07012***       .09615    42.33  .0000     3.88168   4.25856
        |Disturbance Correlation = RHO(1,2)
RHO(1,2)|     -.06225          .06013    -1.04  .3005     -.18010    .05560
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+-------------------------------------------------------------------+
|Cross Tabulation                                                   |
|Row variable is NEWHSAT  (Out of range 0-49:      0)               |
|Number of Rows =  6      (NEWHSAT  =  0 to  5)                     |
|Col variable is INSURANC (Out of range 0-49:      0)               |
|Number of Cols =  3      (INSURANC =  0 to  2)                     |
|Chi-squared independence tests:                                    |
|Chi-squared[  10] =   17.61732   Prob value =  .06177              |
|G-squared  [  10] =   27.62274   Prob value =  .00207              |
+-------------------------------------------------------------------+
|               INSURANC                                            |
+---------+---------------------+------+                            |
| NEWHSAT|      0      1      2| Total|                             |
+---------+---------------------+------+                            |
|       0|      2     87      0|    89|                             |
|       1|      1     54      0|    55|                             |
|       2|      0    156      2|   158|                             |
|       3|     14    250      3|   267|                             |
|       4|     22    307      7|   336|                             |
|       5|     59    963     12|  1034|                             |
+---------+---------------------+------+                            |
|   Total|     98   1817     24|  1939|                             |
+-------------------------------------------------------------------+
```

Polychoric Correlation

The polychoric correlation coefficient is used to quantify the correlation between discrete variables that are qualitative measures. The standard interpretation is that the discrete variables are discretized counterparts to underlying quantitative measures. We typically use ordered probit models to analyze such data. The polychoric correlation measures the correlation between $y_1 = 0,1,...,J_1$ and $y_2 = 0,1,...,J_2$. (Note, J_1 need not equal J_2.) One of the two variables may be binary as well. (If both variables are binary, we use the tetrachoric correlation coefficient described in Section E33.3.)

By this description, the polychoric correlation is simply the correlation coefficient in the bivariate ordered probit model when the two equations contain only constant terms. Thus, to compute the polychoric correlation for a pair of qualitative variables, you can use *LIMDEP*'s bivariate ordered probit model. The commands are as follows: The first two model commands compute the starting values, and the final one computes the correlation.

```
      ORDERED   ; Lhs = y1 ; Rhs = one $
      MATRIX    ; b1 = b ; mu1 = mu $
      ORDERED   ; Lhs = y2 ; Rhs = one $
      MATRIX    ; b2 = b ; mu2 = mu $

or    PROBIT    ; Lhs = y2 ; Rhs = one $
      MATRIX    ; b2 = b $

Then, ORDERED   ; Lhs = y1,y2 ; Rh1 = one ; Rh2 = one
                ; Start = b1,mu1,b2,mu2,0 $
```

For a simple example, we compute the polychoric correlation between self reported health status and sex in the health care usage data examined earlier. Results appear below. Note that the 'model' for sex is simply a computational device.

```
ORDERED   ; Lhs = newhsat ; Rhs = one $
MATRIX    ; b1 = b ; mu1 = mu $
PROBIT    ; Lhs = female ; Rhs = one $
MATRIX    ; b2 = b $
ORDERED   ; Lhs = newhsat,female
          ; Rh1 = one ; Rh2 = one ; Start = b1,mu1,b2,0 $
```

```
-----------------------------------------------------------------------------
Bivariate Ordered Probit Model
Dependent variable             BivOrdPr
Log likelihood function     -3976.40233
Restricted log likelihood   -3977.17511
--------+--------------------------------------------------------------------
 NEWHSAT|                  Standard            Prob.      95% Confidence
  FEMALE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Mean inverse probability for NEWHSAT
Constant|    1.68575***      .04935    34.16  .0000     1.58903   1.78248
        |Mean inverse probability for FEMALE
Constant|     .05109*        .02849     1.79  .0729     -.00475    .10693
        |Threshold Parameters for Probability Model for NEWHSAT
  MU(01)|     .24123***      .03150     7.66  .0000      .17950    .30296
  MU(02)|     .67373***      .04341    15.52  .0000      .58864    .75882
  MU(03)|    1.14226***      .04824    23.68  .0000     1.04770   1.23681
  MU(04)|    1.60213***      .05087    31.49  .0000     1.50242   1.70184
        |Polychoric Correlation for NEWHSAT  and FEMALE
RHO(1,2)|     .03998         .03216     1.24  .2138     -.02305    .10302
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E36: Panel Data Models for Ordered Choice

E36.1 Introduction

The basic ordered choice model is based on the following specification: There is a latent regression,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ \mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0,\ \mathrm{Var}[\varepsilon_i|\mathbf{x}_i] = 1,$$

The observation mechanism results from a complete censoring of the latent dependent variable as follows:

$$\begin{aligned} y_i &= 0 \text{ if } y_i \le \mu_0, \\ &= 1 \text{ if } \mu_0 < y_i \le \mu_1, \\ &= 2 \text{ if } \mu_1 < y_i \le \mu_2, \\ &\ldots \\ &= J \text{ if } \ y_i > \mu_{J\text{-}1}. \end{aligned}$$

The latent 'preference' variable, y_i^* is not observed. The observed counterpart to y_i^* is y_i. Four stochastic specifications are provided for the basic model shown above. The *ordered probit* model based on the normal distribution was developed by Zavoina and McElvey (1975). It applies in applications such as surveys, in which the respondent expresses a preference with the above sort of ordinal ranking. The variance of ε_i is assumed to be one, since as long as y_i^*, $\boldsymbol{\beta}$, and ε_i are unobserved, no scaling of the underlying model can be deduced from the observed data. Estimates are obtained by maximum likelihood. The probabilities which enter the log likelihood function are

$$\mathrm{Prob}[y_i = j] = \mathrm{Prob}[y_i^* \text{ is in the } j\text{th range}].$$

The model may be estimated either with individual data, with $y_i = 0, 1, 2, \ldots$ or with grouped data, in which case each observation consists of a full set of J+1 proportions, $p_{0i},\ldots,p_{Ji}$. This chapter gives the panel data extensions of the ordered choice model.

NOTE: The panel data versions of the ordered choice models require individual data.

There are four classes of panel data models in *LIMDEP*, fixed effects, random effects, random parameters, and latent class. All four are supported for all five of the functional forms presented in Chapter E34.

E36.2 Fixed Effects Ordered Choice Models

The fixed effects models are estimated by maximum likelihood. The command for requesting the model is in two parts. You must fit the model without fixed effects first, to provide the starting values, then the command for the fixed effects estimator follows. The first command and the second must be identical, save for the panel specification in the second command and the constant term in the first, as noted below.

```
ORDERED   ; Lhs = dependent variable
          ; Rhs = independent variables
          [ ; Model = Weibull, Logit , Arctangent or Gompertz ] $
ORDERED   ; Lhs = dependent variable
          ; Rhs = independent variables
          ; Pds = fixed number of periods or count variable
          ; Fixed Effects
          [ ; Model = Weibull, Logit, Arctangent or Gompertz ] $
```

NOTE: The Rhs in your first command must contain a constant term, *one* as the first variable. Your Rhs list for a fixed effects model generally should not include a constant term as the fixed effects model fits a complete set of constants for the set of groups. But, for the ordered probit model, you must provide the identical Rhs list as in the first command, so for this model, do include *one*. It will be removed prior to beginning estimation. When you set up your commands, leaving *one* in the Rhs list will help insure that your model specification is correct. It will look correct. Note, it is crucial that you fit the pooled model first so that *LIMDEP* can find the right starting values for the second estimation step.

The fixed effects model assumes a group specific effect:

$$\text{Prob}[y_{it} = j] = F(j, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$$

where α_i is the parameter to be estimated. You may also fit a two way fixed effects model

$$\text{Prob}[y_{it} = j] = F(j, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \gamma_t)$$

where γ_t is an additional, time (period) specific effect. The time specific effect is requested by adding

; Time

to the command if the panel is balanced, and

; Time = variable name

if the panel is unbalanced. For the unbalanced panel, we assume that overall, the sample observation period is $t = 1,2,..., T$ and that the '*Time*' variable gives for the specific group, the particular values of t that apply to the observations. Thus, suppose your overall sample is five periods. The first group is three observations, periods 1, 2, 4, while the second group is four observations, 2, 3, 4, 5. Then, your panel specification would be

	; Pds = Ti	for example, where *Ti* = 3, 3, 3, 4, 4, 4, 4
and	**; Time = Pd**	for example, where *Pd* = 1, 2, 4, 2, 3, 4, 5.

NOTE: See the discussion below on how this model is estimated. It places an important restriction on the two way fixed effects model.

You must provide the starting values for the iterations by fitting the basic model without fixed effects. You will have a constant term in these results even though it is dropped from the fixed effects model. This is used to get the starting value for the fixed effects. Iterations begin with the restricted model that forces all the fixed effects to equal the constant term in the restricted model.

Results that are kept for this model are

Matrices: *b* = estimate of $\boldsymbol{\beta}$
varb = asymptotic covariance matrix for estimate of $\boldsymbol{\beta}$.
alphafe = estimated fixed effects

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

The upper limit on the number of groups is 100,000. Technical details on the method of estimation for this model are given below and in Chapter R23. Full estimation of the fixed effects model in this fashion encounters the 'incidental parameters' problem.

NOTE: In the ordered probit model with fixed effects α_i, the individual effect coefficient cannot be estimated if the dependent variable within the group takes the same value in every period. The results will indicate how many such groups had to be removed from the sample.

E36.2.1 Standard Model Specifications for Panel Data Ordered Choice Models

This is the full list of general specifications that are applicable to this model estimator. See Chapter E1 and references noted there for further details on these specifications.

Controlling Output from Model Commands

; Par keeps ancillary parameters μ_j with main parameter β vector in *b*.
; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Asymptotic Covariance Matrices

Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Hpt = n sets the number of points to use for Hermite quadrature
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Prob = name saves probabilities as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E36.2.2 Application

We have fit a fixed effects ordered probit model with the German health care data used in the previous examples. This is an unbalanced panel with 7,293 individuals. The health status variable is coded 0 to 10. The model is fit using the commands below. We first fit the pooled model, then the fixed effects model.

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
ORDERED     ; Lhs = newhsat
            ; Rhs = one,hhninc,hhkids,educ ; Partial Effects $
ORDERED     ; Lhs = newhsat
            ; Rhs = one,hhninc,hhkids,educ ; Partial Effects
            ; Fixed Effects ; Pds = _groupti $
```

```
-----------------------------------------------------------------------------
FIXED EFFECTS OrdPrb Model
Dependent variable              NEWHSAT
Log likelihood function    -42217.91813
Estimation based on N =  27326, K =5679
Inf.Cr.AIC  =95793.836 AIC/N =    3.506
Model estimated: Jun 19, 2011, 16:33:13
Probability model based on Normal
Unbalanced panel has   7293 individuals
Skipped 1626 groups with inestimable ai
Ordered probability model
Ordered probit (normal) model
LHS variable = values 0,1,...,10
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Index function for probability
  HHNINC|    -.38858***      .06374    -6.10  .0000     -.51351   -.26365
  HHKIDS|     .07337***      .02718     2.70  .0069      .02010    .12665
    EDUC|    -.04469*        .02635    -1.70  .0898     -.09633    .00695
   MU(1)|     .32638***      .02045    15.96  .0000      .28630    .36646
   MU(2)|     .84692***      .02743    30.88  .0000      .79316    .90068
   MU(3)|    1.39245***      .03005    46.34  .0000     1.33355   1.45135
   MU(4)|    1.81634***      .03102    58.55  .0000     1.75554   1.87714
   MU(5)|    2.68396***      .03226    83.19  .0000     2.62072   2.74719
   MU(6)|    3.10845***      .03272    95.01  .0000     3.04432   3.17258
   MU(7)|    3.76428***      .03340   112.69  .0000     3.69880   3.82975
   MU(8)|    4.79590***      .03478   137.88  .0000     4.72773   4.86407
   MU(9)|    5.50760***      .03610   152.55  .0000     5.43684   5.57836
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The results below compare the estimated partial effects for the outcome y = 10 for the fixed effects model followed by the pooled model. The differences are large. Note that the *educ* coefficient is significantly negative in the fixed effects model and significantly positive in the pooled model. The log likelihood for the pooled model is -57420.08880, so the LR test statistic is about 30,000 with 7,293 degrees freedom. The critical chi squared for 7,292 degrees of freedom, given with the command

CALC ; List ; Ctb(.95,7292) $

is 7,491, which suggests that the fixed effects estimator, at least at this point is preferred. The remains some question, however, because of the incidental parameters problem. Based on received results, in the OP setting, the coefficient is biased away from zero, but not in sign, which still weighs in favor of the FEM result.

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
 NEWHSAT|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=10] at means]--------------
  HHNINC|     .00025         .52441       .93  .3532     -.00028    .00078
 *HHKIDS|     .00469         .17144      1.46  .1431     -.00159    .01097
    EDUC|    -.00282***    -1.16548    -10.59  .0000     -.00334   -.00230
        |--------------[Partial effects on Prob[Y=10] at means]--------------
  HHNINC|     .03739***      .11620      5.36  .0000      .02372    .05105
 *HHKIDS|     .04378***      .38649     16.73  .0000      .03865    .04891
    EDUC|     .00996***      .99545     18.30  .0000      .00889    .01103
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Technical Notes

The fixed effects model is fit essentially by 'brute force.' *LIMDEP* actually estimates the full $K + N$ up to 100,150 coefficients by Newton's method. It is possible to fit the huge number of coefficients because we take advantage of the properties of the sparse second derivatives matrix. One of the implications, however, is that there is no covariance matrix computed for the fixed effects. It is possible to test for the fixed effects model with a likelihood ratio test or a Lagrange multiplier test, but since the covariance matrix is not computed, it is not possible to do any kind of inference for individual fixed effects.

The two way fixed effects estimator is computed by actually creating the time specific dummy variables and adding them to the model. This means that the usual 150 parameter limit on model size applies to the number of variables in the model plus the number of periods (minus one).

Marginal effects in the fixed effects model are computed at the means of the data and with the sample average of the fixed effects estimates as the constant term.

The unconditional log likelihood is maximized by using Newton's method. A full discussion of the method is given in Chapter R23.

E36.3 Random Effects Ordered Choice Models

The random effects model is

$$y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$$

where $i = 1,...,N$ indexes groups and $t = 1,...,T_i$ indexes periods. (As always, the number of periods may vary by individual.) The unique term, ε_{it}, is distributed as $N[0,1]$, standard logistic, extreme value, or Gompertz as specified in the general model discussed earlier. The group specific term, u_i is distributed as $N[0,\sigma^2]$ for all cases. Note that the unobserved heterogeneity, u_i is the same in every period. The parameters of the model are fit by maximum likelihood. As in the binary choice models, the underlying variance, $\sigma^2 = \sigma_u^2 + \sigma_\varepsilon^2$ is not identified. The reduced form parameter, $\rho = \sigma_u^2/\left(\sigma_\varepsilon^2 + \sigma_u^2\right)$, is estimated directly. With the normalization that we used earlier, $\sigma_\varepsilon^2 = 1$, we can determine $\sigma_u = \sqrt{\rho/(1-\rho)}$. Further discussion of the estimation of the structural parameters appears in Chapter R24. The ordered probability model with random effects is estimated in the same fashion as the binary probability models with random effects. The heterogeneity is handled by using Hermite quadrature to integrate the effect out of the joint density of the T_i observations for the ith group. Technical details appear at the end of this section.

E36.3.1 Commands

The specification is for the ordered probability model. Use

```
ORDERED     ; Lhs = ... ; Rhs = ...
            ; Panel spec.
            [ ; Model = Logit, Comploglog, Arctangent or Gompertz] $
```

where the **; Pds** specification follows the standard convention, fixed T or variable name for variable T. The default is the ordered probit. Request the ordered logit just by adding **; Model = Logit** etc. to the command. The random effects model is the default panel data model for the ordered probability models, so you need only include the **; Pds** specification in the command.

NOTE: The random effect, u_i is assumed to be normally distributed in all models. Thus, the logit, arctangent, and other models contain a hybrid of distributions.

All other options are the same as were listed earlier for the pooled ordered probability models.

Marginal effects are computed by setting the heterogeneity term, u_i to its expected value of zero. In order to do the computations of the marginal effects, it is also necessary to scale the coefficients. The ordered probability model with the random effect in the equation is based on the index function $(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) / (1 + \sigma_u^2)$.

This estimator can accommodate restrictions, so

; Rst = list

and **; CML: specification**

are both available. Restrictions may be tested and imposed exactly as in the model with no heterogeneity. Since restrictions can be imposed on all parameters, including ρ, you can fix the value of ρ at any desired value. Do note that forcing the ancillary parameter, in this case, ρ, to equal a slope parameter will almost surely produce unsatisfactory results, and may impede or even prevent convergence of the iterations.

Starting values for the iterations are obtained by fitting the basic model without random effects. Thus, the initial results in the output for these models will be the ordered choice models discussed earlier. You may provide your own starting values for the parameters with

; Start = ... the list of values for β, values for μ, value for ρ

There is no natural moment based estimator for ρ, so a relatively low guess is used as the starting value instead. The starting value for ρ is approximately .2 ($\theta = [2\rho/(1-\rho)]^{1/2} \approx .29$ – see the technical details below. Maximum likelihood estimates are then computed and reported, along with the usual diagnostic statistics. (An example appears below.)

E36.3.2 Output and Results

Your data may not be consistent with the random effects model. That is, there may be no discernible evidence of random effects in your data. In this case, the estimate of ρ will turn out to be negligible. If so, the estimation program issues a diagnostic and reverts back to the original, uncorrelated formulation and reports (again) the results for the basic model.

Results that are kept for this model are

Matrices: *b* = estimate of **β**
varb = asymptotic covariance matrix for estimate of **β**.

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function
rho = estimated value of ρ
varrho = estimated asymptotic variance of estimator of ρ.

Last Model: *b_variables*

Last Function: Prob(y = outcome | **x**)

The additional specification

; Par

in the command requests that μ and σ_u be included in *b* and the additional rows and columns be included in *varb*. The **PARTIAL EFFECTS** and **SIMULATE** commands use the same probability function as the pooled model. The default outcome is the highest one, but you may use **; Outcome = j** to specify a specific one, or **; Outcome = *** for all.

NOTE: The hypothesis of no group effects can be tested with a Wald test (simple t test) or with a likelihood ratio test. The LM approach, using **; Maxit = 0** with a zero starting value for ρ does not work in this setting because with ρ = 0, the last row of the covariance matrix turns out to contain zeros.

NOTE: This model is fit by approximating the necessary integrals in the log likelihood function by Hermite quadrature. An alternative approach to estimating the same model is by Monte Carlo simulation. You can do exactly this by fitting the model as a random parameters model with only a random constant term. This model is described in Section E36.4.

E36.3.3 Application

In the following example, we fit random effects ordered probit models for the health status data. The pooled estimator is fit with and without the clustered data correction. Then, the random effects model is fit, first using the Butler and Moffitt method, then as a random parameters model with a random constant term.

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
ORDERED     ; Lhs = newhsat ; Rhs = one,hhninc,hhkids,educ $
ORDERED     ; Lhs = newhsat ; Rhs = one,hhninc,hhkids,educ
            ; Cluster = id $
ORDERED     ; Lhs = newhsat ; Rhs = one,hhninc,hhkids,educ
            ; Panel  $
ORDERED     ; Lhs = newhsat ; Rhs = one,hhninc,hhkids,educ $
ORDERED     ; Lhs = newhsat ; Rhs = one,hhninc,hhkids,educ
            ; Panel ; RPM ; Fcn = one(n) ; Halton ; Pts = 25 $
```

The first pair of estimation results shown below compares the cluster estimator of the covariance matrix to the pooled estimator which ignores the panel data structure. As can be seen in the results, the robust standard errors are somewhat higher. The second set of results compares two estimators of the random effects model. The first results are based on the quadrature estimator. The second uses maximum simulated likelihood. These two estimators give almost the same results. They would be closer still had we used a larger number of Halton draws. We set this to 25 to speed up the computation. With, say, 250, the results of the two estimators would be extremely close.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable               NEWHSAT
Log likelihood function    -57420.08880
Restricted log likelihood  -57816.35761
Chi squared [   3 d.f.]       792.53762
Significance level               .00000
McFadden Pseudo R-squared      .0068539
Estimation based on N =  27326, K =  13
Inf.Cr.AIC  =********* AIC/N =    4.204
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.42634***      .03136    45.48  .0000     1.36487   1.48781
  HHNINC|     .19469***      .03624     5.37  .0000      .12366    .26571
  HHKIDS|     .22199***      .01261    17.61  .0000      .19728    .24669
    EDUC|     .05187***      .00276    18.81  .0000      .04647    .05728
        |Threshold parameters for index
   Mu(1)|     .19061***      .00988    19.29  .0000      .17123    .20998
   Mu(2)|     .49125***      .01073    45.80  .0000      .47023    .51228
   Mu(3)|     .82152***      .00979    83.95  .0000      .80233    .84070
   Mu(4)|    1.08609***      .00898   120.91  .0000     1.06849   1.10370
   Mu(5)|    1.63179***      .00793   205.69  .0000     1.61624   1.64734
   Mu(6)|    1.88965***      .00767   246.35  .0000     1.87462   1.90469
   Mu(7)|    2.28993***      .00770   297.40  .0000     2.27484   2.30503
   Mu(8)|    2.92948***      .00843   347.32  .0000     2.91295   2.94601
   Mu(9)|    3.38076***      .01008   335.50  .0000     3.36101   3.40051
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.42634***      .05039    28.30  .0000     1.32757   1.52511
  HHNINC|     .19469***      .05008     3.89  .0001      .09653    .29284
  HHKIDS|     .22199***      .01886    11.77  .0000      .18503    .25894
    EDUC|     .05187***      .00432    12.00  .0000      .04340    .06035
        |Threshold parameters for index
   Mu(1)|     .19061***      .02054     9.28  .0000      .15035    .23086
   Mu(2)|     .49125***      .03180    15.45  .0000      .42892    .55358
   Mu(3)|     .82152***      .03548    23.16  .0000      .75198    .89105
   Mu(4)|    1.08609***      .03432    31.64  .0000     1.01882   1.15337
   Mu(5)|    1.63179***      .03334    48.95  .0000     1.56644   1.69713
   Mu(6)|    1.88965***      .03261    57.95  .0000     1.82574   1.95357
   Mu(7)|    2.28993***      .02965    77.24  .0000     2.23183   2.34804
   Mu(8)|    2.92948***      .02827   103.62  .0000     2.87407   2.98489
   Mu(9)|    3.38076***      .02920   115.77  .0000     3.32353   3.43800
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Effects Ordered Probability Model
Dependent variable              NEWHSAT
Log likelihood function    -53631.92165
Underlying probabilities based on Normal
Unbalanced panel has   7293 individuals
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    2.19480***      .07252    30.27  .0000     2.05267   2.33692
  HHNINC|    -.03764         .04636     -.81  .4169     -.12850    .05323
  HHKIDS|     .18979***      .01866    10.17  .0000      .15322    .22635
    EDUC|     .07474***      .00609    12.27  .0000      .06280    .08668
        |Threshold parameters for index model
  Mu(01)|     .27725***      .01553    17.85  .0000      .24680    .30769
  Mu(02)|     .71390***      .02041    34.98  .0000      .67391    .75390
  Mu(03)|    1.18482***      .02235    53.01  .0000     1.14101   1.22863
  Mu(04)|    1.55571***      .02305    67.49  .0000     1.51053   1.60089
  Mu(05)|    2.32085***      .02394    96.95  .0000     2.27393   2.36777
  Mu(06)|    2.68712***      .02427   110.74  .0000     2.63956   2.73469
  Mu(07)|    3.25778***      .02467   132.08  .0000     3.20944   3.30612
  Mu(08)|    4.16499***      .02560   162.70  .0000     4.11482   4.21517
  Mu(09)|    4.79284***      .02605   183.99  .0000     4.74178   4.84390
        |Std. Deviation of random effect
   Sigma|    1.01361***      .01233    82.23  .0000      .98945   1.03778
--------+--------------------------------------------------------------------
Random Coefficients  OrdProbs Model
Dependent variable              NEWHSAT
Log likelihood function    -53699.77298
Ordered probit (normal) model
Simulation based on  25 Halton draws
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
 NEWHSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
  HHNINC|    -.02668         .03421     -.78  .4354     -.09373    .04037
  HHKIDS|     .18456***      .01227    15.05  .0000      .16052    .20860
    EDUC|     .07680***      .00278    27.58  .0000      .07134    .08226
        |Means for random parameters
Constant|    2.13724***      .03627    58.93  .0000     2.06615   2.20832
        |Scale parameters for dists. of random parameters
Constant|    1.04507***      .00729   143.43  .0000     1.03079   1.05935
        |Threshold parameters for probabilities
   MU(1)|     .26755***      .01479    18.09  .0000      .23856    .29653
   MU(2)|     .69343***      .01916    36.20  .0000      .65588    .73097
   MU(3)|    1.15786***      .02068    55.98  .0000     1.11732   1.19840
   MU(4)|    1.52579***      .02116    72.09  .0000     1.48431   1.56728
   MU(5)|    2.28879***      .02177   105.11  .0000     2.24612   2.33147
   MU(6)|    2.65507***      .02203   120.53  .0000     2.61189   2.69824
   MU(7)|    3.22614***      .02239   144.06  .0000     3.18225   3.27003
   MU(8)|    4.13325***      .02334   177.07  .0000     4.08750   4.17900
   MU(9)|    4.75862***      .02385   199.56  .0000     4.71188   4.80535
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E36.3.4 Technical Details for the Random Effects Models

The structure of the random effects model is

$$z_{it} \mid u_i \;=\; \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \varepsilon_{it} \;+\; \sigma_u u_i$$

where u_i ~ N[0,1], and ε_{it} is the stochastic term in the model that provides the *conditional* distribution,

$$\text{Prob}[y_{it} = j \mid \mathbf{x}_{it}, u_i] \;=\; F[j, \boldsymbol{\mu},(\boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \sigma_u u_i)]\; i = 1,...,N,\; t = 1,...,T_i.$$

where $F(.)$ is based on the distribution discussed earlier (normal, logistic, extreme value, arctangent, Gompertz). The parameter vector for the random effects model is

$$\boldsymbol{\theta} \;=\; [\beta_1,...,\beta_K,\, \mu_1,\ldots,\mu_{J\text{-}1},\, \rho]'.$$

With the usual normalization, $\sigma_\varepsilon = 1$,

$$\sigma_u \;=\; \sqrt{\frac{\rho}{1-\rho}}\,.$$

The log likelihood function is

$$\log L \;=\; \Sigma_i \;\log L_i$$

where $\log L_i$ is the contribution of the ith individual (group) to the total. Conditioned on u_i, the joint probability for the ith group is

$$\text{Prob}[Y_{i1} = y_{i1},...,Y_{iTi} = y_{iTi} \mid \mathbf{x}_{i1,...,}u_i] \;=\; \prod_{t=1}^{T_i} F\left[y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i\right]$$

where now, u_i is normalized to unit variance. Since u_i is unobserved, it is necessary to obtain the unconditional log likelihood by taking the expectation of this over the distribution of u_i. For convenience, write the tth term in the probability above as $G(y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma u_i)$, where $\gamma = \sigma_u$, so that

$$L_i \mid u_i \;=\; \prod_{t=1}^{T_i} G(\,y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma} u_i\,)\,.$$

Then,

$$L_i \;=\; E_{ui}\,[L_i \mid u_i]$$

$$= \int_{-\infty}^{\infty} \frac{\exp(-u_i^2/2)}{\sqrt{2\boldsymbol{\pi}}} \prod_{t=1}^{T_i} F(y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma} u_i)\; d\;\mu_i$$

NOTE: It can be seen in the likelihood function that it is necessary actually to compute the product of the probabilities for the group, not the sum of the logs. For this reason, the number of observations in a group cannot be extremely large. Since the probability is likely to be on the order of .25 or so, the product of 100 probabilities is on the order of 10^{-100}. This means that the end result is more rounding error than result. In worse cases, the computation will 'overflow' – that is, exceed the computer's capacity to compute the value. For example, the correct result for the product of 100 probabilities on the order of .01 cannot be computed in the accuracy of the computer, which is about $10^{+/-380}$. The diagnostic that this estimator produces mentions a 'Bad counter...' When the counter for group size exceeds 100, the estimator assumes that you have made some kind of error.

Then, finally,

$$\log L \quad = \sum\nolimits_{i=1}^{N} \log L_i$$

The function is maximized by solving the likelihood equations:

$$\frac{\partial \log L}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} = \sum\nolimits_{i=1}^{N} \frac{\partial \log L_i}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}} = \mathbf{0}.$$

For convenience below, let $\boldsymbol{\theta}$ denote the full parameter vector, $[\boldsymbol{\beta},\gamma]'$.

The integration is done with Hermite quadrature. Make the change of variable to $v_i = u_i/\sqrt{2}$. Then,

$$\log L_i = \ log \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v_i^2) \prod_{t=1}^{T_i} F(\, y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta v_i \,)\, dv_i$$

where $\delta = \gamma \times \sqrt{2}$ [so $\rho = \delta^2/(2 + \delta^2)$] and $\sigma_u = [\rho/(1-\rho)]^{1/2}$]. The integral of the form

$$\int_{-\infty}^{\infty} \exp(-v^2) g(v) dv$$

is approximated by the Hermite quadrature,

$$\int_{-\infty}^{\infty} \exp(-v^2) g(v) dv \approx \sum\nolimits_{h=1}^{H} w_h g(z_h)$$

where w_h are the weights and z_h are the abscissas for the approximation. (See Section R23.3.1, Butler and Moffitt (1982) and Abramovitz and Stegun (1972) for further details.) Collecting terms, then, the log likelihood is computed with

$$\text{logL} \ \approx \sum\nolimits_{i=1}^{N} \ log \left\{ \frac{1}{\sqrt{\pi}} \sum\nolimits_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} F(\, y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta z_h \,) \right] \right\}$$

The derivatives of the log likelihood function are approximated as well (the derivation appears in Chapter R23),

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} \approx \sum_{i=1}^{N} \frac{\left\{\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} F(y_{it},\boldsymbol{\mu},\boldsymbol{\beta}'\mathbf{x}_{it}+\delta z_h)\right]\left[\sum_{t=1}^{T_i}\frac{\partial \log F(y_{it},\boldsymbol{\mu},\boldsymbol{\beta}'\mathbf{x}_{it}+\delta z_h)}{\partial \boldsymbol{\theta}}\right]\right\}}{\left\{\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} w_h \left[\prod_{t=1}^{T_i} F(y_{it},\boldsymbol{\mu},\boldsymbol{\beta}'\mathbf{x}_{it}+\delta z_h)\right]\right\}}$$

Note that L_i and its derivatives are approximated separately. The summation involves two separate integrals. We use a 20 point quadrature by default, but you can change the number of quadrature points by including **; Hpt = p** in the command, where 'p' is the desired number of points, from 4 to 96 (even). In some cases, the accuracy of the computations will improve with the number of quadrature points. However, the amount of computation will increase as well at the same rate.

The variance, δ, appears linearly in the function along with $\boldsymbol{\beta}$, so no complication is added by this additional parameter as the summation is done over the nodes. In each case, the term is

$$F(y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma z_h) = F\left[\mu_y - (\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma z_h)\right] - F\left[\mu_{y-1} - (\boldsymbol{\beta}'\mathbf{x}_{it} + \gamma z_h)\right]$$

The forms of the particular distribution functions, $F_{it}(.)$, differ among the five models. The functional forms appear in Section E34.1. The asymptotic covariance matrix is estimated by the BHHH estimator,

$$\mathbf{H} = \left[\sum_{i=1}^{N}\left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)\left(\frac{1}{L_i}\frac{\partial L_i}{\partial \boldsymbol{\theta}}\right)'\right]^{-1}$$

It is necessary to account for the presence of the random effect when computing probabilities or marginal effects from this model. The CDF is computed from

$$y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i.$$

Thus,

$$\text{Prob}[y_{it}^* \le \mu] = \text{Prob}[\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i \le \mu]$$

or

$$= \text{Prob}[\varepsilon_{it} + u_i \le \mu - \boldsymbol{\beta}'\mathbf{x}_{it}]$$

$$= \text{Prob}\left[\frac{\varepsilon_{it}+u_i}{\sqrt{1+\sigma_u^2}} \le \frac{\mu-\boldsymbol{\beta}'\mathbf{x}_{it}}{\sqrt{1+\sigma_u^2}}\right]$$

$$= \text{Prob}\,[v_{it} \le \mu^* - \boldsymbol{\beta}^{*\prime}\mathbf{x}_{it}]$$

where $v_{it} \sim N[0,1]$. These are the probabilities that enter the calculation of marginal effects and fitted values.

E36.4 Random Parameters and Random Thresholds Ordered Choice Models

The structure of the random parameters model is based on the conditional probability

$$\text{Prob}[y_{it} = j |\, \mathbf{x}_{it}, \boldsymbol{\beta}_i] = F(\,j,\boldsymbol{\mu}, \boldsymbol{\beta}_i'\mathbf{x}_{it} + \alpha_i),\ i = 1,...,N,\ t = 1,...,T_i.$$

where $F(.)$ is the distribution discussed earlier (normal, logistic, extreme value, Gompertz). The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) parameters generated by

$$\text{E}[\boldsymbol{\beta}_i |\, \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

(the second term is optional – the mean may be constant),

$$\text{Var}[\boldsymbol{\beta}_i |\, \mathbf{z}_i] = \boldsymbol{\Sigma}.$$

The model is operationalized by writing

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i.$$

As noted earlier, the heterogeneity term is optional. In addition, it may be assumed that some of the parameters are nonrandom. It is convenient to analyze the model in this fully general form here. We accommodate nonrandom parameters just by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$.

NOTE: If there is no heterogeneity in the mean, and only the constant term is considered random – the model may specify that some parameters are nonrandom – then this model is functionally equivalent to the random effects model of the preceding section. The estimation technique is different, however. An application appears in the previous section.

Two major extensions of the RP-OC model are provided. The threshold parameters, μ_{ij} and disturbance variance of ε_i may also be random, in the form

$$\mu_{ij} = \mu_{i,j\text{-}1} + \exp(\alpha_j + \boldsymbol{\delta}'\mathbf{w}_i + \theta u_{ij}),\ \mu_0 = 0,\ u_{ij} \sim \text{N}[0,1]$$

$$\varepsilon_{it} \sim \text{N}[0,\sigma_i^2],\ \sigma_i = \exp(\boldsymbol{\gamma}'\mathbf{f}_i + \tau h_i),\ h_i \sim \text{N}[0,1]$$

This model is developed in Section E36.4.4.

E36.4.1 Model Commands

The basic model command for this form of the model is, as is the fixed effects estimator, given in two parts. The model is fit conventionally first to provide the starting values, then fully specified.

```
ORDERED     ; Lhs = dependent variable
            ; Rhs = independent variables
            [ ; Model = Gompertz, Logit or Weibull ] $
ORDERED     ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Pds = fixed periods or count variable
            ; RPM
            ; Fcn = random parameters specification
            [ ; Model = Gompertz, Logit or Weibull ] $
```

NOTE: For this model, your Rhs list should include a constant term.

Starting values for the iterations are provided by the user by fitting the basic model without random parameters first. Note in the applications below that the two random parameters ordered probit estimators are each preceded by an otherwise identical fixed parameters version.

NOTE: The command cannot reuse an earlier set of results. You must refit the basic model without random parameters each time. Thus,

```
ORDERED     ; ... $
ORDERED     ; RPM ; ... $
ORDERED     ; RPM ; ... $
```

will not work properly. Each random parameters model must be preceded by a set of starting values.

Correlated Random Parameters

The preceding defines an estimator for a model in which the covariance matrix of the random parameters is diagonal. To extend it to a model in which the parameters are freely correlated, add

; Correlation (or just ; Cor)

to the command. Note that this formulation of the model has an ambiguous interpretation if your parameters are not jointly normally distributed. A correlated mixture of several distributions is difficult to interpret.

Heterogeneity in the Means

The preceding examples have specified that the mean of the random variable is fixed over individuals. If there is measured heterogeneity in the means, in the form of

$$E[\beta_{ki}] = \beta_k + \Sigma_m \delta_{km} z_{mi}$$

where z_m is a variable that is measured for each individual, then the command may be modified to

; RPM = list of variables in z.

In the data set, these variables must be repeated for each observation in the group.

Autocorrelation

You may change the character of the heterogeneity from a time invariant effect to an AR(1) process, $v_{kit} = \rho_k v_{ki,t-1} + w_{kit}$. (See Section R24.7 for details.)

Controlling the Simulation

There are two parameters of the simulations that you can change. *R* is the number of points in the simulation. Authors differ in the appropriate value. Train (2009) recommends several hundred. Bhat suggests 1,000 is an appropriate value. The program default is 100. You can choose the value with

; Pts = number of draws, R

The value of 50 that we set in our experiments above was chosen purely to produce an example that you could replicate without spending an inordinate amount of waiting for the results.

The standard approach to simulation estimation is to use random draws from the specified distribution. As suggested immediately above, good performance in this connection requires very large numbers of draws. The drawback to this approach is that with large samples and large models, this entails a huge amount of computation and can be very time consuming. Some authors have documented dramatic speed gains with no degradation in simulation performance through the use of a small number of Halton draws instead of a large number of random draws. Halton sequences are discussed in Section R24.7. Some authors (e.g., Bhat (1999)) have found that a Halton sequence of draws with only one tenth the number of draws as a random sequence is equally effective. To use this approach, add

; Halton

to your model command.

In order to replicate an estimation, you must use the same random draws. One implication of this is that if you give the identical model command twice in sequence, you will not get the identical set of results because the random draws in the sequences will be different. To obtain the same results, you must reset the seed of the random number generator with a command such as

CALC ; Ran (seed value) $

(Note that we have used **Ran(12345)** before each of our examples above, precisely for this reason. The specific value you use for the seed is not of consequence; any odd number will do.

In this connection, we note a consideration which is crucial in this sort of estimation. The random sequence used for the model estimation must be the same in order to obtain replicability. In addition, during estimation of a particular model, the same set of random draws must be used for each person every time. That is, the sequence $\mathbf{v}_{i1}$, $\mathbf{v}_{i2}$, ..., $\mathbf{v}_{iR}$ used for each individual must be the same every time it is used to calculate a probability, derivative, or likelihood function. (If this is not the case, the likelihood function will be discontinuous in the parameters, and successful estimation becomes unlikely. This has been called simulation 'noise' or 'buzz' in the literature.) One way to achieve this which has been suggested in the literature is to store the random numbers in advance, and simply draw from this reservoir of values as needed. Because *LIMDEP* is able to use very large samples, this is not a practical solution, especially if the number of draws is large as well. We achieve the same result by assigning to each individual, i, in the sample, their own random generator seed which is a unique function of the global random number seed, S, and their group number, i;

$$\text{Seed}(S,i) = S + 123.0 \times i, \text{ then minus 1.0 if the result is even.}$$

Since the global seed, S, is a positive odd number, this seed value is unique, at least within the several million observation range of *LIMDEP*.

Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one, x1, x2, x3, x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be constant. For those that you wish to specify as random, use

; Fcn = variable name (distribution), variable name (distribution), ...

Numerous distributions may be specified. All random variables, v_{ik}, have mean zero. See Section R24.3 for details.

c for constant (zero variance), $v_i = 0$
n for normally distributed, v_i = a standard normally distributed variable
u for uniform, v_i= a standard uniform distributed variable in (-1,+1)
t for triangular (the 'tent' distribution)
h for negative half normal, $v = (2\pi)^{-1/2} - |u|$
e for centered lognormal, $v = \text{Exp}(u) - \text{Sqr(e)}$
s for Johnson S_b, $v = \text{Exp}(u) / [1 + \text{Exp}(u)]$
l for lognormal

Each of these is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. The latter two distributions are provided as one may wish to reduce the amount of variation in the tails of the distribution of the parameters across individuals and to limit the range of variation. (See Train (2009) for discussion.) To specify that the constant term and the coefficient on $x1$ are normally distributed with fixed mean and variance, use

; Fcn = one(n), x1(n)

This specifies that the first and second coefficients are random while the remainder are not. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

Standard Model Specifications for the Random Parameters Ordered Choice Models

This is the full list of general that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps individual specific parameter estimates.
; Partial Effects	displays marginal effects, same as **; Marginal Effects**.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E36.4.2 Results

Results saved by this estimator are:

Matrices:	*b*	= estimate of θ
	varb	= asymptotic covariance matrix for estimate of θ.
	beta_i	= individual specific parameters, if **; Par** is requested.
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function
Last Model:	*b_variables*	

Last Function: Prob($y_{it} = J|\mathbf{x}_{it}$) = Probability of the highest cell.
May be changed with **; Outcome = j** or **; Outcome = ***.

E36.4.3 Application

The following example illustrates the random parameters ordered probit model. The data are recoded to make a more compact example, and the sample is restricted to those groups that have seven observations, to speed up the simulations. The first two ordered probit models are the fixed parameters, pooled estimator followed by the random parameters case in which two of the five coefficients are random. After the random parameters model is estimated, the individual specific estimates of $E[\beta_{educ}|hs,x]$ are collected in a variable then a kernel estimator describes the distribution of the conditional means across the sample. The results are rearranged to compare the coefficient estimates then the partial effects across the specifications.

The results include estimates of the means and standard deviations of the distributions of the random parameters and the estimates of the nonrandom parameters. The log likelihood shown is conditioned on the random draws, so one might be cautious about using it to test hypotheses, for example, that the parameters are random at all by comparing it to the log likelihood from the basic model with all nonrandom coefficients.

The commands are:

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = one,age,educ,hhninc,handdum $
CREATE      ; hs = newhsat $
RECODE      ; hs ; 0/3 = 0 ; 4/6 = 1 ; 7/8 = 2 ; 9/10 = 3 $
HISTOGRAM ; Rhs = hs $
REJECT      ; ti < 7 $
ORDERED     ; Lhs = hs ; Rhs = x ; Partial Effects  $
ORDERED     ; Lhs = hs ; Rhs = x
            ; RPM ; Panel ; Fcn = age(n),educ(n) ; Halton ; Pts = 25
            ; Partial Effects ; Par $
SAMPLE      ; 1-887 $
MATRIX      ; mb_educ = beta_i(1:118,1:1) $
CREATE      ; be_educ = mb_educ $
KERNEL      ; Rhs = be_educ $
ORDERED     ; Lhs = hs ; Rhs = x ; Partial Effects  $
ORDERED     ; Lhs = hs ; Rhs = x
            ; RPM ; Panel ; Fcn = age(n),educ(n) ; Halton ; Pts = 25
            ; Correlated ; Partial Effects ; Par $
```

```
+----------------------------------------------------------------+
|                 CELL FREQUENCIES FOR ORDERED CHOICES           |
+----------------------------------------------------------------+
|                Frequency        Cumulative  < =    Cumulative  > =  |
|Outcome       Count     Percent    Count     Percent   Count     Percent |
|----------- ------- ---------  ------- ---------  ------- --------- |
|HS=00           569     9.1641      569     9.1641     6209  100.0000 |
|HS=01          2000    32.2113     2569    41.3754     5640   90.8359 |
|HS=02          2342    37.7194     4911    79.0949     3640   58.6246 |
|HS=03          1298    20.9051     6209   100.0000     1298   20.9051 |
+----------------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                    HS
Log likelihood function      -7679.52077
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      HS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    1.72050***      .10585    16.25  .0000     1.51304    1.92796
     AGE|    -.02354***      .00155   -15.19  .0000     -.02658    -.02051
    EDUC|     .06417***      .00687     9.34  .0000      .05069     .07764
  HHNINC|     .26574***      .08773     3.03  .0025      .09381     .43768
 HANDDUM|    -.34752***      .03370   -10.31  .0000     -.41358    -.28146
        |Threshold parameters for index
   Mu(1)|    1.17217***      .01623    72.20  .0000     1.14035    1.20399
   Mu(2)|    2.24966***      .01942   115.83  .0000     2.21160    2.28773
```

```
--------+-------------------------------------------------------------
Random Coefficients  OrdProbs Model
Dependent variable                    HS
Log likelihood function      -6724.01324
Estimation based on N =   6209, K =   9
Unbalanced panel has    887 individuals
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      HS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Nonrandom parameters
Constant|    2.56865***      .11016    23.32  .0000     2.35275   2.78455
  HHNINC|     .18922**       .08693     2.18  .0295      .01884    .35960
 HANDDUM|    -.18622***      .03508    -5.31  .0000     -.25497   -.11747
        |Means for random parameters
     AGE|    -.04128***      .00159   -26.01  .0000     -.04439   -.03817
    EDUC|     .10807***      .00748    14.45  .0000      .09341    .12273
        |Scale parameters for dists. of random parameters
     AGE|     .01357***      .00034    39.55  .0000      .01289    .01424
    EDUC|     .08208***      .00155    53.01  .0000      .07905    .08512
        |Threshold parameters for probabilities
   MU(1)|    1.64297***      .02744    59.87  .0000     1.58918   1.69676
   MU(2)|    3.17465***      .03234    98.16  .0000     3.11126   3.23804
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------

----------------------------------------------------------------------
Random Coefficients  OrdProbs Model
Dependent variable                    HS
Log likelihood function       -994.76038
--------+-------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      HS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Nonrandom parameters
Constant|    2.97520***      .25659    11.60  .0000     2.47230   3.47811
  HHNINC|     .23351         .22085     1.06  .2903     -.19934    .66637
 HANDDUM|    -.25589***      .09735    -2.63  .0086     -.44670   -.06508
        |Means for random parameters
     AGE|    -.04495***      .00386   -11.66  .0000     -.05250   -.03739
    EDUC|     .06925***      .01533     4.52  .0000      .03921    .09930
        |Diagonal elements of Cholesky matrix
     AGE|     .00860***      .00262     3.29  .0010      .00347    .01373
    EDUC|     .04047***      .00337    12.02  .0000      .03388    .04707
        |Below diagonal elements of Cholesky matrix
lEDU_AGE|     .03878***      .01003     3.87  .0001      .01912    .05844
        |Threshold parameters for probabilities
   MU(1)|    1.65758***      .08339    19.88  .0000     1.49414   1.82102
   MU(2)|    3.11571***      .09843    31.65  .0000     2.92279   3.30864
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

```
Implied covariance matrix of random parameters
Var_Beta|             1             2
--------+----------------------------
       1|   .739584E-04   .333495E-03
       2|   .333495E-03     .00314200
Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|     .00859991
       2|      .0560536
Implied correlation matrix of random parameters
Cor_Beta|             1             2
--------+----------------------------
       1|       1.00000       .691818
       2|       .691818       1.00000
```

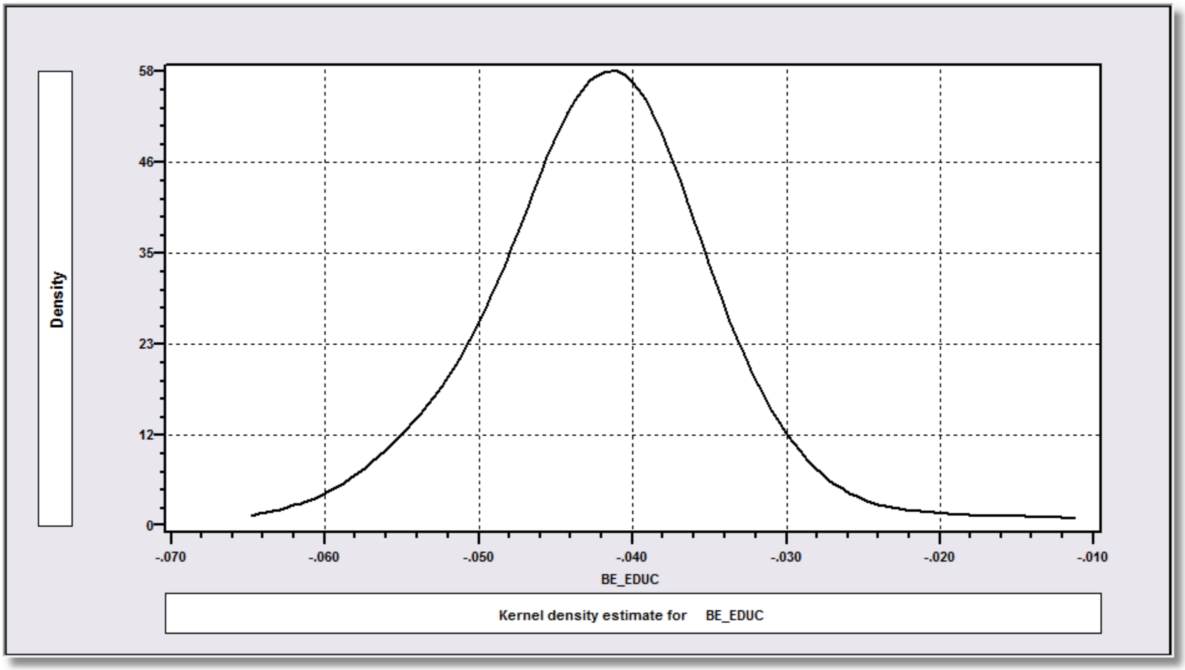


Figure E36.1 Estimators of E[β(educ)|y,x]

```
(Fixed parameters)
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
      HS|      Effect    Elasticity      z       |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
     AGE|     .00353***     1.93407    14.53  .0000       .00305    .00401
    EDUC|    -.00962***    -1.30082    -9.18  .0000     -.01168   -.00757
  HHNINC|    -.03986***     -.17200    -3.02  .0025     -.06570   -.01402
 HANDDUM|     .05213***      .13505    10.09  .0000      .04200    .06225
(outcomes 1 and 2 omitted)
        |--------------[Partial effects on Prob[Y=03] at means]--------------
     AGE|    -.00654***    -1.46872   -14.52  .0000     -.00742   -.00566
    EDUC|     .01782***      .98783     9.17  .0000      .01401    .02163
  HHNINC|     .07381***      .13061     3.02  .0025      .02598    .12164
 HANDDUM|    -.09653***     -.10255   -10.15  .0000     -.11517   -.07788
--------+--------------------------------------------------------------------
(Random parameters)
-----------------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
     AGE|     .00247***     4.25914    16.65  .0000      .00218    .00276
    EDUC|    -.00647***    -2.75143   -12.52  .0000     -.00748   -.00546
  HHNINC|    -.01133**      -.15380    -2.16  .0306     -.02159   -.00106
 HANDDUM|     .01115***      .09088     5.22  .0000      .00696    .01533
 (Outcomes 1 and 2 omitted, effects reordered)
        |--------------[Partial effects on Prob[Y=03] at means]--------------
     AGE|    -.00776***    -3.12921   -22.25  .0000     -.00844   -.00708
    EDUC|     .02031***     2.02149    13.54  .0000      .01737    .02325
  HHNINC|     .03557**       .11300     2.17  .0296      .00351    .06762
 HANDDUM|    -.03500***     -.06677    -5.27  .0000     -.04801   -.02199
--------+--------------------------------------------------------------------
(Correlated random parameters)
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
     AGE|     .00344***     4.40201     6.82  .0000      .00245    .00443
    EDUC|    -.00530***    -1.78538    -4.17  .0000     -.00779   -.00281
  HHNINC|    -.01786        -.19039    -1.05  .2927     -.05114    .01541
 HANDDUM|     .01958***      .13543     2.67  .0077      .00519    .03397
        |--------------[Partial effects on Prob[Y=03] at means]--------------
     AGE|    -.00772***    -3.51945    -9.49  .0000     -.00931   -.00612
    EDUC|     .01189***     1.42743     4.34  .0000      .00653    .01726
  HHNINC|     .04010         .15222     1.06  .2906     -.03427    .11448
 HANDDUM|    -.04395**      -.10827    -2.55  .0107     -.07768   -.01022
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E36.4.4 Random Parameters HOPIT Model

This model extends the hierarchical ordered probit model in several directions. The core model is an ordered probit specification:

$$
\begin{aligned}
y_{it}^* &= \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\\
y_{it} &= 0 \text{ if } y_{it}^* \le 0,\\
&= 1 \text{ if } 0 < y_{it}^* \le \mu_1,\\
&= 2 \text{ if } \mu_1 < y_{it}^* \le \mu_2,\\
&\ldots\\
&= J \text{ if } y_{it}^* > \mu_J
\end{aligned}
$$

as usual. The model is constructed to include random coefficients, $\boldsymbol{\beta}_i$, random variance heterogeneity, σ_i, and random thresholds, μ_{ij}. The random parameters form of the general model is

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{h}_i + \boldsymbol{\Gamma}\mathbf{w}_i$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix of standard deviations and $w_{ik} \sim N[0,1]$, $k = 1,\ldots,K$. The mean of the random parameter vector is $\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{h}_i$ where $\mathbf{h}_i$ may be a set of variables specified in the model. The disturbance in the model may be heteroscedastic and distributed with random standard deviation as well, with

$$\varepsilon_{it} \sim N[0,\sigma_i^2],\ \ \sigma_i = \exp[\boldsymbol{\gamma}'\mathbf{z}_i + \tau v_i] \text{ where } v_i \sim N[0,1].$$

Finally, the thresholds are formed as shown for the cross section variant of this model in Section E35.6;

$$
\begin{aligned}
\mu_{ij} &= \mu_{i,j-1} + \exp(\alpha_j + \boldsymbol{\delta}'\mathbf{w}_i + \theta_j u_{ij}], \text{ where } u_{ij} \sim N[0,1]\\
\mu_0 &= 0 \text{ and } \mathbf{x}_{it} \text{ contains a constant term.}
\end{aligned}
$$

The various parts are optional. In addition, the model may be fit with cross section or panel data. As usual, panel data are likely to be more effective. The command for this model is

ORDERED ; Lhs = … ; Rhs = …

; RPM for the random coefficients, β

or **; RPM = list of variables in $\mathbf{h_i}$**

; RTM for the random thresholds model
; Limits = list of variables for the $\mathbf{w_i}$ in the thresholds
; Random Effects to use a common $\mathbf{u_i}$ in the thresholds

; RVM for the random term i, $\mathbf{v_i}$ in $\boldsymbol{\sigma_i}$
; Het ; Hfn = list of variables in $\mathbf{z_i}$ for the heteroscedastic model $

When the model includes any of the three random components, the maximum simulated likelihood estimator is used. The default model is an ordered probit specification. You may specify an ordered logit model instead by adding

; Logit

to the command.

The simulation can be modified with

; Pts = the number of points or draws

and **; Halton**

to indicate that Halton sequences rather than random draws be used for the simulations. Halton sequences are recommended. The simulation is over the J elements in μ_{ij} plus the element v_i in σ_i plus the K variables in the Rhs specification. If you specify a 'random effects' model, then the same single random term appears in all of the threshold equations.

If you are using a panel data set, use either

SETPANEL ; Group = variable name
; Pds = variable name $

with **; Panel**

in the **ORDERED** command, or, if the Pds variable is already prepared, use

; Pds = the group count variable.

Partial effects for this model are computed internally and requested with

; Partial Effects.

This general form of the random parameters ordered probit model does not use the template random parameters form described in Chapter R24. (Note that there is no **; Fcn** = specification component in the command.) As formulated, all parameters on the variables in the Rhs list are assumed to be random. You can modify this by imposing a constraint that the corresponding diagonal element of $\mathbf{\Gamma}$, which is the standard deviation of the random part of that element of $\boldsymbol{\beta}_i$, be equal to zero. To do this, include in the command

; Rh2 = list of variables with nonrandom parameters.

Thus, the full list of variables in the model is those in the Rhs list plus those in the Rh2 list. There is no overlap – variables must appear in only one of these two lists.

Results saved by this estimator are:

Matrices:	*b*	= estimate of β
	varb	= asymptotic covariance matrix for estimate of θ.
	betartop	= full set of parameter estimates , if **; Par** is requested.
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function
Last Model:	*b_variables*	
Last Function:	None	

Standard Model Specifications for the Random Parameters Ordered Choice Models

This is the full list of general that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Partial Effects displays partial effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

The following application uses the subset of the GSOEP sample that have five observations in each group. The application is further speeded up by using only 10 Halton draws in the simulations. This is sufficient for a numerical example, but would be far too small for an actual application. The estimated model allows for unobserved heterogeneity in all three places, the parameters, thresholds and disturbance variance.

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
REJECT      ; ti # 5 $
ORDERED     ; Lhs = hsat ; Rhs = one,age,educ ; Rh2 = hhninc,married,hhkids
            ; RPM ; RTM ; RVM
            ; Limits = female ; Pts = 10
            ; Halton ; Panel ; Maxit = 25 $
```

```
-----------------------------------------------------------------------------
Random Thresholds Ordered Choice Model
Dependent variable                  HSAT
Log likelihood function     -10134.79176
Restricted log likelihood   -10899.81624
Chi squared [  17 d.f.]       1530.04896
Significance level                .00000
McFadden Pseudo R-squared       .0701869
Estimation based on N =   5255, K =  29
Inf.Cr.AIC  =20327.584 AIC/N =    3.868
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Latent Regression Equation
Constant|    4.17571***      .16744    24.94  .0000     3.84754   4.50388
     AGE|    -.04388***      .00218   -20.13  .0000     -.04815   -.03961
    EDUC|     .06261***      .00965     6.49  .0000      .04370    .08153
  HHNINC|     .35696***      .11753     3.04  .0024      .12662    .58731
 MARRIED|     .09078*        .04999     1.82  .0694     -.00719    .18876
  HHKIDS|    -.09768**       .04371    -2.23  .0254     -.18334   -.01201
        |Intercept Terms in Random Thresholds
Alpha-01|   -1.19538***      .13834    -8.64  .0000    -1.46653   -.92423
Alpha-02|    -.69311***      .08966    -7.73  .0000     -.86884   -.51739
Alpha-03|    -.70446***      .06420   -10.97  .0000     -.83029   -.57862
Alpha-04|   -1.14567***      .08731   -13.12  .0000    -1.31679   -.97455
Alpha-05|    -.19232***      .03307    -5.82  .0000     -.25713   -.12751
Alpha-06|   -1.03759***      .05273   -19.68  .0000    -1.14094   -.93424
Alpha-07|    -.58017***      .03466   -16.74  .0000     -.64810   -.51224
Alpha-08|    -.04815*        .02878    -1.67  .0943     -.10456    .00826
Alpha-09|    -.39987***      .04048    -9.88  .0000     -.47920   -.32054
        |Standard Deviations of Random Thresholds
Alpha-01|     .24187***      .07688     3.15  .0017      .09118    .39256
Alpha-02|     .34510***      .06721     5.14  .0000      .21338    .47682
Alpha-03|     .19508**       .08818     2.21  .0270      .02224    .36792
Alpha-04|     .26252***      .08332     3.15  .0016      .09922    .42582
Alpha-05|     .11536***      .03689     3.13  .0018      .04305    .18767
Alpha-06|     .17729***      .06490     2.73  .0063      .05009    .30448
Alpha-07|     .23047***      .03758     6.13  .0000      .15683    .30412
Alpha-08|     .15433***      .02927     5.27  .0000      .09697    .21170
Alpha-09|     .04443         .04045     1.10  .2721     -.03486    .12371
```

```
        |Variables in Random Thresholds
  FEMALE|    -.03079**       .01291    -2.38  .0171     -.05609   -.00549
        |Standard Deviations of Random Regression Parameters
Constant|     .06490         .05458     1.19  .2344     -.04208    .17187
     AGE|     .02166***      .00083    26.18  .0000      .02004    .02328
    EDUC|     .00519**       .00234     2.22  .0264      .00061    .00977
  HHNINC|        0.0    .....(Fixed Parameter).....
 MARRIED|        0.0    .....(Fixed Parameter).....
  HHKIDS|        0.0    .....(Fixed Parameter).....
        |Latent Heterogeneity in Variance of Epsilon
  Tau(v)|     .29096***      .01860    15.65  .0000      .25451    .32741
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.

+----------------------------------------------------------------+
| Summary of Marginal Effects for Ordered Probability Model (probit)   |
| Effects are computed by averaging  over observs. during simulations. |
| Binary variables change only by 1 unit so s.d. changes are not shown.|
| Elasticities for binary variables = partial effect/probability = %chgP |
+----------------------------------------------------------------+
+----------------------------------------------------------------+
|          Regression Variable AGE        Changes in AGE        % chg|
|        -----------------------------   -----------------------------
Outcome    Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  -----------------------------   -----------------------------
Y = 00     .00158     .00158     .00000     .01766     .06166    5.85945
Y = 01     .00057     .00215    -.00158     .00640     .02235    3.00925
Y = 02     .00128     .00343    -.00215     .01425     .04973    2.42584
Y = 03     .00168     .00511    -.00343     .01876     .06548    1.83159
Y = 04     .00130     .00641    -.00511     .01451     .05065    1.18846
Y = 05     .00336     .00977    -.00641     .03753     .13101     .94528
Y = 06     .00154     .01131    -.00977     .01720     .06003     .70612
Y = 07     .00046     .01176    -.01131     .00511     .01782     .12789
Y = 08    -.00304     .00872    -.01176    -.03401    -.11873    -.56476
Y = 09    -.00344     .00528    -.00872    -.03840    -.13403   -1.42223
Y = 10    -.00528     .00000    -.00528    -.05901    -.20598   -2.34240
+----------------------------------------------------------------+
|          Regression Variable EDUC       Changes in EDUC       % chg|
|        -----------------------------   -----------------------------
Outcome    Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  -----------------------------   -----------------------------
Y = 00    -.00226    -.00226     .00000    -.00540    -.02482   -2.13858
Y = 01    -.00082    -.00307     .00226    -.00196    -.00900   -1.09832
Y = 02    -.00182    -.00489     .00307    -.00435    -.02002    -.88538
Y = 03    -.00240    -.00729     .00489    -.00573    -.02636    -.66849
Y = 04    -.00185    -.00914     .00729    -.00443    -.02039    -.43376
Y = 05    -.00479    -.01394     .00914    -.01147    -.05273    -.34501
Y = 06    -.00220    -.01613     .01394    -.00525    -.02416    -.25772
Y = 07    -.00065    -.01679     .01613    -.00156    -.00717    -.04668
Y = 08     .00434    -.01244     .01679     .01039     .04779     .20613
Y = 09     .00490    -.00754     .01244     .01173     .05395     .51909
Y = 10     .00754     .00000     .00754     .01803     .08291     .85493
+----------------------------------------------------------------+
|          Regression Variable HHNINC     Changes in HHNINC     % chg|
|        -----------------------------   -----------------------------
Outcome    Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  -----------------------------   -----------------------------
```

```
Y = 00    -.01286    -.01286     .00000    -.00229    -.03857    -.37184
Y = 01    -.00466    -.01752     .01286    -.00083    -.01398    -.19097
Y = 02    -.01037    -.02790     .01752    -.00185    -.03111    -.15394
Y = 03    -.01366    -.04156     .02790    -.00244    -.04096    -.11623
Y = 04    -.01057    -.05213     .04156    -.00188    -.03168    -.07542
Y = 05    -.02733    -.07946     .05213    -.00487    -.08195    -.05999
Y = 06    -.01252    -.09198     .07946    -.00223    -.03755    -.04481
Y = 07    -.00372    -.09570     .09198    -.00066    -.01115    -.00812
Y = 08     .02477    -.07093     .09570     .00442     .07427     .03584
Y = 09     .02796    -.04297     .07093     .00499     .08384     .09025
Y = 10     .04297     .00000     .04297     .00766     .12884     .14865
+------------------------------------------------------------------------+
|            Regression Variable MARRIED     Changes in MARRIED    % chg|
|         ------------------------------   ------------------------------
Outcome    Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00    -.00327    -.00327     .00000    -.00138    -.00327    -.20824
Y = 01    -.00119    -.00446     .00327    -.00050    -.00119    -.10695
Y = 02    -.00264    -.00710     .00446    -.00111    -.00264    -.08621
Y = 03    -.00347    -.01057     .00710    -.00147    -.00347    -.06509
Y = 04    -.00269    -.01326     .01057    -.00113    -.00269    -.04224
Y = 05    -.00695    -.02021     .01326    -.00293    -.00695    -.03359
Y = 06    -.00318    -.02339     .02021    -.00134    -.00318    -.02509
Y = 07    -.00095    -.02434     .02339    -.00040    -.00095    -.00455
Y = 08     .00630    -.01804     .02434     .00266     .00630     .02007
Y = 09     .00711    -.01093     .01804     .00300     .00711     .05054
Y = 10     .01093     .00000     .01093     .00461     .01093     .08325
+------------------------------------------------------------------------+
|            Regression Variable HHKIDS      Changes in HHKIDS     % chg|
|         ------------------------------   ------------------------------
Outcome    Effect  dPy<=nn/dX dPy>=nn/dX   1 StdDev   Low to High   Elast
-------  ------------------------------   ------------------------------
Y = 00     .00352     .00352     .00000     .00173     .00352     .11752
Y = 01     .00128     .00480    -.00352     .00063     .00128     .06036
Y = 02     .00284     .00763    -.00480     .00139     .00284     .04865
Y = 03     .00374     .01137    -.00763     .00183     .00374     .03674
Y = 04     .00289     .01426    -.01137     .00142     .00289     .02384
Y = 05     .00748     .02174    -.01426     .00367     .00748     .01896
Y = 06     .00343     .02517    -.02174     .00168     .00343     .01416
Y = 07     .00102     .02619    -.02517     .00050     .00102     .00257
Y = 08    -.00678     .01941    -.02619    -.00332    -.00678    -.01133
Y = 09    -.00765     .01176    -.01941    -.00375    -.00765    -.02853
Y = 10    -.01176     .00000    -.01176    -.00577    -.01176    -.04698
-------------------------------------------------------------------------
Indirect Partial Effects for Ordered Choice Model
Variables in thresholds
Outcome  FEMALE
Y = 00    .000000
Y = 01   -.000468
Y = 02   -.001603
Y = 03   -.002728
Y = 04   -.002883
Y = 05   -.009219
Y = 06   -.005379
Y = 07   -.005158
Y = 08    .002091
Y = 09    .007557
Y = 10    .017791
```

E36.5 Latent Class Ordered Choice Models

The ordered choice model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is

$$\text{Prob}[Y_{it} = y_{it}|\ \mathbf{x}_{it}] = F(y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it}) = P(i,t), y_{it} = 0, 1,...,.$$

Henceforth, we use the term 'group' to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. Unobserved heterogeneity in the distribution of Y_{it} is assumed to impact the density in the form of a random effect. The continuous distribution of the heterogeneity is approximated by using a finite number of 'points of support.' The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J 'classes' with a model which allows for heterogeneity as follows: The probability of observing y_{it} given that regime j applies is

$$P(i,t|j) = \text{Prob}[Y_{it} = y_{it}|\ \mathbf{x}_{it}, j]$$

where the density is now specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$P(i,t|j) = F[y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta_j], \text{Prob}[\text{class} = j] = F_j.$$

We formulate this approximation more generally as,

$$P(i,t|j) = F[y_{it}, \boldsymbol{\mu}, \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}_j'\mathbf{x}_{it}], F_j = \exp(\theta_j) / \Sigma_{\text{j}} \exp(\theta_j), \text{ with } \theta_J = 0.$$

In this formulation, each group has its own parameter vector, $\boldsymbol{\beta}_j' = \boldsymbol{\beta} + \boldsymbol{\delta}_j$, though the variables that enter the mean are assumed to be the same. (This can be changed by imposing restrictions on the full parameter vector, as described below.) This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters – each $\boldsymbol{\delta}_j$ has only one nonzero element in the location of the constant term. You may also specify that the latent class probabilities depend on person specific characteristics, so that

$$\theta_{ij} = \boldsymbol{\theta}_j'\mathbf{z}_i, \boldsymbol{\theta}_J = \mathbf{0}.$$

Technical details for estimation of latent class models are given in Section R25.9.

E36.5.1 Command

The estimation command for this model is

ORDERED **; Lhs = ...**
; Rhs = independent variables
[; Model = Weibull, Logit or Gompertz]
; LCM (for latent class model)
[; LCM = list of variables in z_i for multinomial logit class probabilities]
; Pds = panel data specification $

The default number of support points is five. You may set *J* to 2, 3, ..., 10 with

; Pts = the value you wish.

Some particular values computed for the latent class model are

; Group = the index of the most likely latent class
; Cprob = estimated posterior probability for the most likely latent class

You can obtain a listing of these two results by using

; List.

Standard Model Specifications for the Latent Class Ordered Choice Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

You can use the **; Rst = list** option to structure the latent class model so that different variables appear in different classes. Alternatively, you can use this to force the Heckman and Singer form of the model as follows, where we use a three class model as an example:

```
NAMELIST   ; x = ... one, list of variables $
CALC       ; k1 = Col(x) - 1 ; kmu = Max(y) - 1 $
ORDERED    ; Lhs = ... ; Rhs = x ; LCM ; Pts = 3
           ; Rst = d1, k1_b, kmu_mu,
                   d2, k1_b, kmu_mu,
                   d3, k1_b, kmu_mu,  t1,t2,t3 $
```

E36.5.2 Results

Results saved by this estimator are

Matrices: *b* = full parameter vector, $[\boldsymbol{\beta}_1', \boldsymbol{\beta}_2',\ldots F_1,\ldots,F_J]$
varb = full covariance matrix

(Note that *b* and *varb* involve $J\times(K$+#outcomes - 1) estimates.)

*beta_* = individual specific parameters, if **; Par** is requested
b_class = a $J\times K$ matrix with each row equal to the corresponding $\boldsymbol{\beta}_j$
class_pr = a $J\times 1$ vector containing the estimated class probabilities

Scalars: *kreg* = number of variables in Rhs list
nreg = total number of observations used for estimation
logl = maximized value of the log likelihood function
exitcode = exit status of the estimation procedure

Last Function: None

Application

To illustrate the model, we will fit an ordered probit model with three latent classes. We have modified the health care data set to set up a compact example. (The latent class estimator is actually unable to resolve more than one class with nine threshold parameters.) We have censored the health satisfaction measure to three classes for purpose of this exercise. The ordered probit model is the same one specified earlier. Some of the numerical results are omitted to simplify comparison of the estimated models. The first set of commands creates the data set.

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
CREATE      ; health = newhsat $
RECODE      ; health ; 0/4 = 0 ; 5/8 = 1 ; 9/10 = 2 $
NAMELIST    ; x = one,hhninc,hhkids,educ $
```

We now fit the base case pooled model.

```
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects $
```

This is a three class latent class model.

```
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects
            ; LCM ; Pts = 3 ; Panel $
```

This fits two random effects models, the continuous, normally distributed effects model and Heckman and Singer’s discrete approximation.

```
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects ; Panel $
ORDERED     ; Quietly ; Lhs = health ; Rhs = x $
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects
            ; LCM  ; Pts = 3 ; Panel
            ; Rst =  alpha0,3_b,cmu,alpha1,3_b,cmu,
                     alpha2,3_b,cmu,theta0,theta1,theta2 $
```

This model specifies that the class probabilities depend on age and sex.

```
SAMPLE      ; All $
ORDERED     ; Quietly ; Lhs = health ; Rhs = x $
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects
            ; LCM = one,age,female ; Pts = 3 ; Panel $
```

Finally, we use a small subsample to show the listing of the posterior class probabilities.

```
REJECT      ; ti # 3 $
ORDERED     ; Quietly ; Lhs = health ; Rhs = x $
ORDERED     ; Lhs = health ; Rhs = x ; Partial Effects
            ; LCM = one,age,female ; Pts = 3 ; Panel ; List $
```

This is the base case, pooled ordered probit model, with no group effects followed by the estimates of the parameters of the three class latent class model.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                HEALTH
Log likelihood function    -24522.47670
Restricted log likelihood  -24801.77601
Chi squared [   3 d.f.]       558.59861
Significance level               .00000
McFadden Pseudo R-squared      .0112613
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =49054.953 AIC/N =    1.795
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HEALTH|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|     .38694***      .03538    10.94  .0000      .31761    .45628
  HHNINC|     .15134***      .04069     3.72  .0002      .07160    .23109
  HHKIDS|     .21408***      .01419    15.09  .0000      .18627    .24188
    EDUC|     .04904***      .00311    15.77  .0000      .04294    .05513
        | Threshold parameters for index
   Mu(1)|    1.83426***      .01130   162.26  .0000     1.81210   1.85641
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Latent Class / Panel OrdProbs Model
Dependent variable                HEALTH
Log likelihood function    -21956.55643
Estimation based on N =  27326, K =  17
Inf.Cr.AIC  =43947.113 AIC/N =    1.608
Model estimated: Jul 19, 2011, 18:58:26
Unbalanced panel has   7293 individuals
Ordered probability model
Ordered probit (normal) model
LHS variable = values 0,1,..., 2
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HEALTH|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Model parameters for latent class 1
Constant|    1.16608***      .10831    10.77  .0000      .95379   1.37838
  HHNINC|    -.22927**       .08945    -2.56  .0104     -.40458   -.05395
  HHKIDS|     .10979***      .03316     3.31  .0009      .04480    .17479
    EDUC|     .08077***      .00937     8.62  .0000      .06241    .09913
   MU(1)|    1.73212***      .04607    37.60  .0000     1.64184   1.82241
        | Model parameters for latent class 2
Constant|     .62012***      .07038     8.81  .0000      .48218    .75805
  HHNINC|    -.06265         .07865     -.80  .4257     -.21681    .09151
  HHKIDS|     .24254***      .02664     9.11  .0000      .19034    .29475
    EDUC|     .06115***      .00621     9.85  .0000      .04899    .07332
   MU(1)|    2.68221***      .02902    92.43  .0000     2.62533   2.73909
        | Model parameters for latent class 3
Constant|   -1.00572***      .11321    -8.88  .0000    -1.22762   -.78383
  HHNINC|     .52603***      .12473     4.22  .0000      .28157    .77050
  HHKIDS|     .24566***      .04766     5.15  .0000      .15225    .33908
    EDUC|     .05198***      .01000     5.20  .0000      .03239    .07157
   MU(1)|    1.88097***      .06379    29.49  .0000     1.75595   2.00600
```

```
        | Estimated prior probabilities for class membership
Class1Pr|     .27635***      .00916    30.17  .0000      .25839    .29430
Class2Pr|     .56896***      .01168    48.69  .0000      .54605    .59186
Class3Pr|     .15470***      .00823    18.80  .0000      .13857    .17083
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

These are the estimated marginal effects for the two models presented above, with the pooled estimates first followed by those derived from the latent class model.

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                         Prob.      95% Confidence
  HEALTH|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
  HHNINC|    -.03364***     -.08477    -3.72  .0002     -.05137   -.01591
 *HHKIDS|    -.04653***     -.33304   -15.36  .0000     -.05247   -.04060
    EDUC|    -.01090***     -.88316   -15.70  .0000     -.01226   -.00954
        |--------------[Partial effects on Prob[Y=01] at means]--------------
  HHNINC|    -.01184***     -.00657    -3.63  .0003     -.01824   -.00545
 *HHKIDS|    -.01875***     -.02955   -11.05  .0000     -.02208   -.01542
    EDUC|    -.00384***     -.06848   -11.47  .0000     -.00449   -.00318
        |--------------[Partial effects on Prob[Y=02] at means]--------------
  HHNINC|     .04548***      .07091     3.72  .0002      .02150    .06947
 *HHKIDS|     .06528***      .28908    14.74  .0000      .05660    .07396
    EDUC|     .01474***      .73880    15.58  .0000      .01288    .01659
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                         Prob.      95% Confidence
  HEALTH|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
  HHNINC|     .00289         .01116      .34  .7345     -.01381    .01959
 *HHKIDS|    -.03296***     -.36179   -10.53  .0000     -.03910   -.02683
    EDUC|    -.01068***    -1.32670   -12.47  .0000     -.01236   -.00900
        |--------------[Partial effects on Prob[Y=01] at means]--------------
  HHNINC|     .00154         .00073      .34  .7350     -.00738    .01046
 *HHKIDS|    -.01987***     -.02682    -7.68  .0000     -.02494   -.01479
    EDUC|    -.00569***     -.08698    -8.07  .0000     -.00707   -.00431
        |--------------[Partial effects on Prob[Y=02] at means]--------------
  HHNINC|    -.00443        -.00928     -.34  .7347     -.03004    .02118
 *HHKIDS|     .05283***      .31427    10.18  .0000      .04265    .06300
    EDUC|     .01637***     1.10240    12.05  .0000      .01371    .01903
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the random effects model. It is comparable to the Heckman and Singer form that follows. The first model with continuously distributed effects suggests a random constant term with mean 2.33642 and standard deviation 0.99095. From the Heckman and Singer model, using the three estimated constants and the three estimated prior probabilities, we find a mean of 2.19016 and standard deviation 0.90994. Since the remaining coefficients in the latent class model do not differ across classes, they are directly comparable to the random effects model. The overall similarity is to be expected, but there are some substantive differences. For example, the latent class model predicts a much smaller influence of marital status than does the random effects model.

```
-----------------------------------------------------------------------------
Random Effects Ordered Probability Model
Dependent variable                HEALTH
Log likelihood function    -22042.38298
Restricted log likelihood  -24522.47670
Chi squared [   1 d.f.]      4960.18744
Significance level               .00000
McFadden Pseudo R-squared      .1011355
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =44096.766 AIC/N =    1.614
Underlying probabilities based on Normal
Unbalanced panel has   7293 individuals
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  HEALTH|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|     .64927***      .07239     8.97  .0000      .50739    .79115
  HHNINC|    -.03500         .05665     -.62  .5367    -.14603    .07603
  HHKIDS|     .20576***      .02188     9.40  .0000      .16288    .24865
    EDUC|     .07118***      .00625    11.40  .0000      .05894    .08343
        | Threshold parameters for index model
  Mu(01)|    2.56175***      .01686   151.90  .0000     2.52870   2.59480
        | Std. Deviation of random effect
   Sigma|    1.00299***      .01483    67.63  .0000      .97392   1.03206
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

--------+--------------------------------------------------------------------
Latent Class / Panel OrdProbs Model
Dependent variable                HEALTH
Log likelihood function    -22048.67454
Estimation based on N =  27326, K =   9
Inf.Cr.AIC  =44115.349 AIC/N =    1.614
Unbalanced panel has   7293 individuals
Ordered probability model
Ordered probit (normal) model
LHS variable = values 0,1,..., 2
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  HEALTH|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Model parameters for latent class 1
Constant|    2.12385***      .06069    35.00  .0000     2.00490   2.24279
  HHNINC|    -.07289         .05188    -1.40  .1601    -.17458    .02880
  HHKIDS|     .20014***      .01936    10.34  .0000     .16220    .23808
    EDUC|     .05987***      .00507    11.81  .0000     .04994    .06981
   MU(1)|    2.46535***      .01693   145.63  .0000     2.43217   2.49853
```

```
        | Model parameters for latent class 2
Constant|    -.95230***      .06385   -14.92  .0000    -1.07743   -.82717
  HHNINC|    -.07289         .05188    -1.40  .1601     -.17458    .02880
  HHKIDS|     .20014***      .01936    10.34  .0000      .16220    .23808
    EDUC|     .05987***      .00507    11.81  .0000      .04994    .06981
   MU(1)|    2.46535***      .01693   145.63  .0000     2.43217   2.49853
        | Model parameters for latent class 3
Constant|     .56180***      .05806     9.68  .0000      .44801    .67560
  HHNINC|    -.07289         .05188    -1.40  .1601     -.17458    .02880
  HHKIDS|     .20014***      .01936    10.34  .0000      .16220    .23808
    EDUC|     .05987***      .00507    11.81  .0000      .04994    .06981
   MU(1)|    2.46535***      .01693   145.63  .0000     2.43217   2.49853
        | Estimated prior probabilities for class membership
Class1Pr|     .23642***      .00833    28.38  .0000      .22009    .25275
Class2Pr|     .13069***      .00723    18.07  .0000      .11652    .14487
Class3Pr|     .63289***      .00995    63.60  .0000      .61338    .65239
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The following takes a closer look at the distributions of heterogeneity implied by the continuous random effects model and the discrete distribution implied by the Heckman and Singer model. The program below plots the two distributions. The densities are evaluated at 500 points ranging from the mean of the continuous distribution plus and minus three standard deviations. (The program could be made generic based on the model results. We have used the actual values in a few commands.)

```
MATRIX      ; ah = [2.12385/-.95230/.56180] $
MATRIX      ; ph = [.23642/.13069/.63289] $
SAMPLE      ; 1-500 $
CALC        ; min = .64927 - 3*1.00299
            ; max = .64927 + 3*1.00929
            ; delta = .002 * (max-min) $
CREATE      ; alpha = Trn(min,delta) $
CREATE      ; Normal = 1/1.00929  * N01((alpha - .64927)/1.00929) $
CALC        ; ahs1 = ah(2) ; ahs2 = ah(3) ; ahs3 = ah(1) $
CALC        ; mid12 = .5*(ahs2+ahs1) ; mid23 = .5*(ahs2+ahs3) $
CALC        ; dhs1 = ph(2)/(mid12-min) $
CALC        ; dhs2 = ph(3)/(mid23-mid12) $
CALC        ; dhs3 = ph(1)/(max-mid23) $
CREATE      ; hecksing =dhs1*(alpha < mid12) +
                        dhs2*(alpha >= mid12) * (alpha < mid23) +
                        dhs3*(alpha >= mid23) $
PLOT        ; Lhs = alpha ; Rhs = normal,hecksing
            ; Fill ; Limits = 0,.45 ; Endpoints = min,max
            ; Title = Discrete & Continuous Distributions of Heterogeneity
            ; Yaxis = RndmEfct $
```

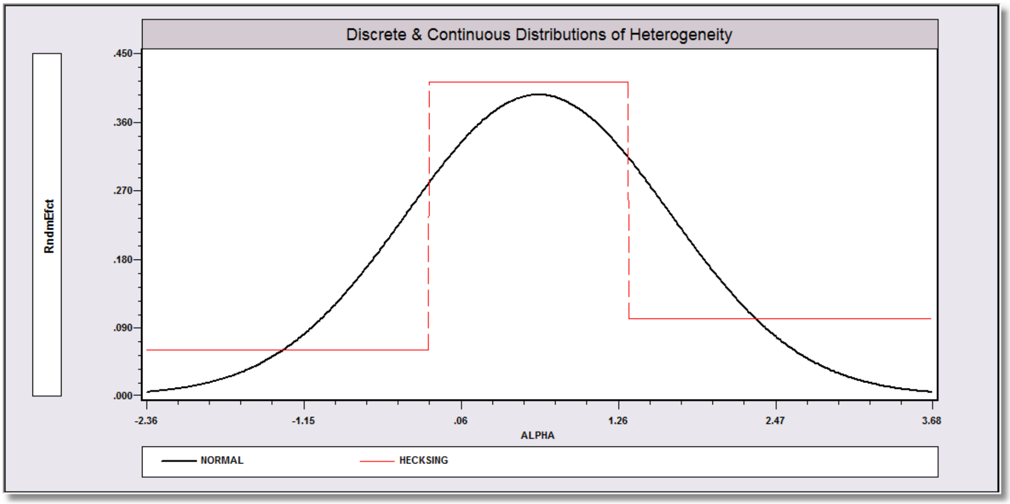


Figure E36.2 Discrete and Continuous Distributions of Heterogeneity

These are the estimated marginal effects for the two models. Once again, they are quite similar, as might be expected.

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                               Prob.      95% Confidence
  HEALTH|      Effect    Elasticity      z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
  HHNINC|     .00552         .01381      .62   .5368      -.01199     .02303
 *HHKIDS|   -.03196***     -.22713    -9.53   .0000      -.03853    -.02539
    EDUC|   -.01122***     -.90314   -11.26   .0000      -.01318    -.00927
        |--------------[Partial effects on Prob[Y=01] at means]--------------
  HHNINC|     .00203         .00114      .62   .5350      -.00437     .00842
 *HHKIDS|   -.01283***     -.02046    -6.92   .0000      -.01646    -.00920
    EDUC|   -.00412***     -.07437    -8.19   .0000      -.00511    -.00313
        |--------------[Partial effects on Prob[Y=02] at means]--------------
  HHNINC|   -.00754        -.01144     -.62   .5362      -.03145     .01636
 *HHKIDS|    .04479***      .19287     9.10   .0000       .03514     .05444
    EDUC|    .01534***      .74797    11.24   .0000       .01267     .01802
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the Heckman and Singer form of the model.

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                               Prob.      95% Confidence
  HEALTH|      Effect    Elasticity      z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
```

```
        |--------------[Partial effects on Prob[Y=00] at means]--------------
  HHNINC|      .00993          .04901     1.40  .1606      -.00394    .02380
 *HHKIDS|     -.02655***      -.37215   -10.42  .0000      -.03154   -.02155
    EDUC|     -.00816***     -1.29445   -11.47  .0000      -.00955   -.00676
        |--------------[Partial effects on Prob[Y=01] at means]--------------
  HHNINC|      .00772          .00353     1.40  .1614      -.00308    .01852
 *HHKIDS|     -.02285***      -.02968    -7.96  .0000      -.02848   -.01723
    EDUC|     -.00634***      -.09323    -8.90  .0000      -.00774   -.00494
        |--------------[Partial effects on Prob[Y=02] at means]--------------
  HHNINC|     -.01765         -.03913    -1.41  .1600      -.04227    .00697
 *HHKIDS|      .04940***       .31106     9.90  .0000       .03962    .05917
    EDUC|      .01450***      1.03341    11.49  .0000       .01202    .01697
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

In the model below, the class probabilities depend on age and sex. These are averaged over the data in the table at the end of the results. The constant probabilities from the model estimated earlier are shown with them. An important feature to note here is that there is no natural ordering of classes in the latent class model. The ordering of the second and third classes has changed from the earlier model to this one.

```
-----------------------------------------------------------------------------
Latent Class / Panel OrdProbs Model
Dependent variable               HEALTH
Log likelihood function    -21779.75836
Estimation based on N =  27326, K =  21
Inf.Cr.AIC  =43601.517 AIC/N =    1.596
Model estimated: Jul 19, 2011, 19:27:39
Unbalanced panel has   7293 individuals
Ordered probability model
Ordered probit (normal) model
LHS variable = values 0,1,..., 2
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HEALTH|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|     1.41223***       .10283    13.73  .0000      1.21070   1.61377
  HHNINC|     -.24084***       .08785    -2.74  .0061      -.41301   -.06866
  HHKIDS|      .02548          .03257      .78  .4340      -.03836    .08932
    EDUC|      .06130***       .00862     7.11  .0000       .04441    .07819
   MU(1)|     1.72679***       .04553    37.93  .0000      1.63756   1.81602
        |Model parameters for latent class 2
Constant|     -.80867***       .12257    -6.60  .0000     -1.04890   -.56845
  HHNINC|      .55004***       .12874     4.27  .0000       .29771    .80236
  HHKIDS|      .11778**        .05227     2.25  .0242       .01533    .22023
    EDUC|      .03595***       .01105     3.25  .0011       .01430    .05760
   MU(1)|     1.93880***       .06839    28.35  .0000      1.80477   2.07284
        |Model parameters for latent class 3
Constant|      .80114***       .07069    11.33  .0000       .66260    .93969
  HHNINC|     -.08541          .07783    -1.10  .2725      -.23796    .06713
  HHKIDS|      .16879***       .02640     6.39  .0000       .11706    .22052
    EDUC|      .04689***       .00614     7.64  .0000       .03487    .05892
   MU(1)|     2.66629***       .02734    97.53  .0000      2.61270   2.71987
```

```
        |Estimated prior probabilities for class membership
   ONE_1|     .81468***      .13922     5.85  .0000      .54181   1.08755
   AGE_1|    -.03807***      .00345   -11.05  .0000     -.04482   -.03131
FEMALE_1|    -.13830*        .07356    -1.88  .0601     -.28247    .00586
   ONE_2|   -3.09023***      .22351   -13.83  .0000    -3.52830  -2.65215
   AGE_2|     .04049***      .00447     9.07  .0000      .03174    .04924
FEMALE_2|    -.01649         .09674     -.17  .8647     -.20609    .17312
   ONE_3|        0.0    .....(Fixed Parameter).....
   AGE_3|        0.0    .....(Fixed Parameter).....
FEMALE_3|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

```
+--------------------------------------------------------------+
|  Prior class probabilities at data means for LCM variables |
|   Class 1     Class 2     Class 3     Class 4     Class 5  |
|    .24199      .15782      .60019      .00000      .00000  |
+--------------------------------------------------------------+
```

The model estimates include the estimates of the prior probabilities of group membership. As shown in Section R25.9, it is also possible to compute the posterior probabilities for the groups, conditioned on the data. The **; List** specification will request a listing of these. The following illustration shows this feature for a small subset of the data used above.

```
Predictions computed for the group with the largest posterior probability
Obs.  Periods Fitted outcomes
=============================================================================
Ind.=    1  J* = 2  P(j)=  .008  .881  .111
Ind.=    2  J* = 2  P(j)=  .401  .491  .109
Ind.=    3  J* = 2  P(j)=  .203  .737  .060
Ind.=    4  J* = 2  P(j)=  .050  .909  .041
Ind.=    5  J* = 2  P(j)=  .186  .702  .113
Ind.=    6  J* = 2  P(j)=  .172  .735  .094
Ind.=    7  J* = 2  P(j)=  .177  .735  .088
Ind.=    8  J* = 2  P(j)=  .039  .869  .092
Ind.=    9  J* = 3  P(j)=  .002  .334  .663
Ind.=   10  J* = 3  P(j)=  .000  .003  .997
Ind.=   11  J* = 2  P(j)=  .106  .836  .057
Ind.=   12  J* = 2  P(j)=  .079  .758  .164
Ind.=   13  J* = 2  P(j)=  .023  .928  .049
Ind.=   14  J* = 2  P(j)=  .017  .959  .024
Ind.=   15  J* = 2  P(j)=  .106  .829  .065
Ind.=   16  J* = 2  P(j)=  .070  .895  .036
Ind.=   17  J* = 2  P(j)=  .388  .497  .114
Ind.=   18  J* = 2  P(j)=  .065  .842  .093
Ind.=   19  J* = 3  P(j)=  .006  .111  .884
Ind.=   20  J* = 3  P(j)=  .017  .391  .592
Ind.=   21  J* = 3  P(j)=  .010  .353  .637
Ind.=   22  J* = 2  P(j)=  .140  .735  .125
Ind.=   23  J* = 3  P(j)=  .003  .422  .575
Ind.=   24  J* = 2  P(j)=  .101  .826  .073
Ind.=   25  J* = 2  P(j)=  .043  .920  .037
```

E36.6 Stratification by Thresholds

A version of the ordered probability model for a particular type of stratified data may be estimated with

```
ORDERED     ; Lhs = your dependent variable
            ; Rhs = ...
            ; Str  = your stratification variable $
```

The Lhs and Rhs variables are as usual and *y* is still coded 0,1,2,... The model assumes that there are $S \leq 9$ strata, indicated for each observation with the value of the stratification variable. The model estimates the slope parameters as usual but allows each stratum to have its own set of cutoff values. These are the μs. They are marked accordingly in the output. For example, suppose that *y* takes values 0,1,2,3, and there are four strata. Variable *stratum* takes values 1,2,3,4. The commands which use an artificial set of data,

```
CALC        ; Ran (12345) $
SAMPLE      ; 1-500 $
CREATE      ; x1 = Rnn(1,4) ; x2 = Rnd(2) - 1 $
CREATE      ; y = 1 + .5*x1 + 1.2*x2 + Rnn(0,1) $
RECODE      ; y ; -10/2.5 = 0 ; 2.501/3 = 1 ; 3.001 /4 = 2 ; 4.001/10 = 3 $
CREATE      ; stratum = Rnd(4)
            ; s2 = (stratum = 2)
            ; s3 = (stratum = 3)
            ; s4 = (stratum = 4) $
ORDERED     ; Lhs = y
            ; Rhs = one, s2, s3, s4, x1, x2
            ; Str = stratum
            ; Partial Effects $
```

estimate parameters β_0, α_2, α_3, α_4, β_1, β_2, and μ_{11}, μ_{21}, μ_{12}, μ_{22}, μ_{13}, μ_{23}, μ_{14}, μ_{24}. The Lhs variable *y* takes four values, so each stratum has a μ_1 and a μ_2. There are four such pairs.

Because the first of the cutoff parameters for each stratum is a free parameter, and the intercept is adjusted as a normalization, you should have a constant term for each stratum. Instead of separate intercepts, you *must* have a basis constant term and then separate dummy variables for *S* - 1 other strata. Omit one of the dummy variables to avoid the dummy variable trap.

With stratification, the vector *mu* contains the *S* times (*J*-1) values, where *S* is the number of strata. Likewise, in the *Last Model* labels vector, the parameters are labeled as shown above, that is *mu*1_1,*mu*2_1, and so on. Aside from these changes, the model is otherwise unchanged. The results below are produced by the preceding set of commands.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                     Y
Log likelihood function      -320.84619
Restricted log likelihood    -574.87353
Chi squared [  11 d.f.]       508.05468
Significance level               .00000
McFadden Pseudo R-squared      .4418839
Estimation based on N =    500, K =  14
Inf.Cr.AIC  =  669.692 AIC/N =    1.339
Model estimated: Jul 12, 2011, 20:35:46
Underlying probabilities based on Normal
   Stratification on STRATUM
 1   135        2   122        3   116
 4   128
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|   -1.68077***      .16762   -10.03  .0000    -2.00930  -1.35224
      S2|     .22883         .18979     1.21  .2279     -.14315    .60081
      S3|     .45747**       .20630     2.22  .0266      .05312    .86182
      S4|     .01482         .18971      .08  .9377     -.35700    .38665
      X1|     .51182***      .03058    16.74  .0000      .45188    .57176
      X2|    1.26894***      .14268     8.89  .0000      .98929   1.54860
        |Threshold parameters for index
 Mu(1,1)|     .38644***      .05855     6.60  .0000      .27168    .50120
 Mu(2,1)|    1.20251***      .09181    13.10  .0000     1.02256   1.38246
 Mu(1,2)|     .68566***      .06306    10.87  .0000      .56206    .80925
 Mu(2,2)|    2.19620***      .12518    17.54  .0000     1.95085   2.44155
 Mu(1,3)|     .51859***      .07248     7.16  .0000      .37654    .66064
 Mu(2,3)|    1.54458***      .09916    15.58  .0000     1.35023   1.73893
 Mu(1,4)|     .51921***      .08191     6.34  .0000      .35866    .67976
 Mu(2,4)|    1.53027***      .09235    16.57  .0000     1.34926   1.71128
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+----------------------------------------------------------------------+
|                CELL FREQUENCIES FOR ORDERED CHOICES                  |
+----------------------------------------------------------------------+
|               Frequency       Cumulative  < =    Cumulative  > =    |
|Outcome      Count    Percent   Count    Percent   Count    Percent |
|----------- -------  ---------  -------  ---------  -------  --------- |
|Y=00           278    55.6000      278    55.6000      500   100.0000 |
|Y=01            43     8.6000      321    64.2000      222    44.4000 |
|Y=02            77    15.4000      398    79.6000      179    35.8000 |
|Y=03           102    20.4000      500   100.0000      102    20.4000 |
+----------------------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+--------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
       Y|      Effect    Elasticity      z      |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
     *S2|    -.08701    .....(Fixed Parameter).....
     *S3|    -.17589    .....(Fixed Parameter).....
     *S4|    -.00555        -.00920      -.67  .5032     -.02178    .01069
      X1|    -.19122***     -.34783    -12.22  .0000     -.22188   -.16056
     *X2|    -.45361***     -.75220     -5.87  .0000     -.60509   -.30213
        |--------------[Partial effects on Prob[Y=01] at means]--------------
     *S2|     .01514***      .13045      3.47  .0005      .00660    .02368
     *S3|     .02666***      .22972      2.93  .0034      .00885    .04448
     *S4|     .00106         .00918       .45  .6494     -.00353    .00566
      X1|     .03694***      .34914      5.12  .0000      .02280    .05108
     *X2|     .07206***      .62088      5.36  .0000      .04573    .09839
        |--------------[Partial effects on Prob[Y=02] at means]--------------
     *S2|     .04242    .....(Fixed Parameter).....
     *S3|     .08412***      .46894     15.80  .0000      .07369    .09455
     *S4|     .00273         .01523       .43  .6641     -.00960    .01506
      X1|     .09425***      .57638      9.49  .0000      .07478    .11373
     *X2|     .20959***     1.16842      7.89  .0000      .15753    .26165
        |--------------[Partial effects on Prob[Y=03] at means]--------------
     *S2|     .02945***      .29004      5.01  .0000      .01792    .04097
     *S3|     .06511***      .64138     11.85  .0000      .05434    .07589
     *S4|     .00175    .....(Fixed Parameter).....
      X1|     .06003***      .64861      4.17  .0000      .03182    .08823
     *X2|     .17196***     1.69388      2.75  .0059      .04950    .29443
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------

Cross tabulation of predictions and actual outcomes
+------+-----+-----+-----+-----+-----+
|y(i,j)|  0  |  1  |  2  |  3  |Total|
+------+-----+-----+-----+-----+-----+
|   0  |  262|    0|    5|   11|  278|
|   1  |   27|    0|    4|   12|   43|
|   2  |   26|    0|    4|   47|   77|
|   3  |    8|    0|    5|   89|  102|
+------+-----+-----+-----+-----+-----+
| Total|  323|    0|   18|  159|  500|
+------+-----+-----+-----+-----+-----+
Row = actual, Column = Prediction, Model = Probit
Prediction is number of the most probable cell.
```

Cross tabulation of outcomes and predicted probabilities.

y(i,j)	0	1	2	3	Total
0	230	15	21	12	278
1	19	5	10	9	43
2	19	7	18	33	77
3	9	5	15	74	102
Total	277	32	63	128	500

Row = actual, Column = Prediction, Model = Probit
Value(j,m)=Sum(i=1,N)y(i,j)*p(i,m).
Column totals may not match cell sums because of rounding error.

E37: Multinomial Logit Models

E37.1 Introduction

Chapters E37 and E38 will describe two forms of the ‘multinomial logit’ model. These models are also known variously as ‘conditional logit,’ ‘discrete choice,’ and ‘universal logit’ models, among other names. All of them can be viewed as special cases of a general model of utility maximization: An individual is assumed to have preferences defined over a set of alternatives (travel modes, occupations, food groups, etc.)

$$U(\text{alternative } 0) = \boldsymbol{\beta}_0'\mathbf{x}_{i0} + \varepsilon_{i0}$$

$$U(\text{alternative } 1) = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1}$$

...

$$U(\text{alternative } J) = \boldsymbol{\beta}_J'\mathbf{x}_{iJ} + \varepsilon_{iJ}$$

$$\text{Observed } Y_i = \text{choice } j \text{ if } U_i(\text{alternative } j) > U_i(\text{alternative } k)\ \forall\ k \neq j.$$

The ‘disturbances’ in this framework (individual heterogeneity terms) are assumed to be independently and identically distributed with identical extreme value distribution; the CDF is

$$F(\varepsilon_j) = \exp(-\exp(-\varepsilon_j)).$$

Based on this specification, the choice probabilities,

$$\text{Prob[choice } j \text{]} = \text{Prob}[U_j > U_k],\ \forall\ k \neq j$$

$$= \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_{ji})}{\sum_{m=0}^{J}\exp(\boldsymbol{\beta}_m'\mathbf{x}_{mi})},\ j = 0,...,J,$$

where ‘i’ indexes the observation, or individual, and ‘j’ and ‘m’ index the choices. We note at the outset, the IID assumptions made about ε_j are quite stringent, and lead to the ‘Independence from Irrelevant Alternatives’ or IIA implications that characterize the model. Much (perhaps all) of the research on forms of this model consists of development of alternative functional forms and stochastic specifications that avoid this feature. We return to that aspect in Chapter E38, and leave it unresolved for the present.

The observed data consist of the Rhs vectors, $\mathbf{x}_{jt}$, and the outcome, or choice, y_t. (We also consider a number of variants.) There are many forms of the multinomial logit, or multinomial choice model supported in *LIMDEP*. The *NLOGIT* program provides the major extensions. *LIMDEP* contains two basic forms of the model that are documented in this and the next chapter of this manual.

This chapter will examine what we call the *multinomial logit* model. In this setting, it is assumed that the Rhs variables consist of a set of individual specific characteristics, such as age, education, marital status, etc. These are the same for all choices, so the choice subscript on **x** in the formula above is dropped. The observation setting is the individual's choice among a set of alternatives, where it is assumed that the determinant of the choice is the *characteristics* of the individual. An example might be a model of choice of occupation. (This is the model originally devised by Nerlove and Press (1973).) For convenience at this point, we label this the multinomial logit model.

Chapter E38 will examine what we call (again, purely for convenience) the *discrete choice* model and, also, to differentiate the command, the *conditional logit* model. In this framework, we observe the *attributes* of the choices, as well as (or, possibly, instead of) the characteristics of the individual. A well known example is travel mode choice. Samples of observations often consist of the attributes of the different modes and the choice actually made. Sometimes, no characteristics of the individuals are observed beyond their actual choice. Models may also contain mixtures of the two types of choice determinants. These are considered in Chapter E38 as well. (We emphasize, these naming distinctions are meaningless in the modeling framework – we just use them here only to organize the applicable parts of *LIMDEP*. In practice, all of the models considered in this chapter and Chapter E38 are multinomial logit models.

E37.2 The Multinomial Logit Model – MLOGIT

The general form of the *multinomial logit* model is

$$\text{Prob[choice } j \text{]} = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_t)}{\sum_{m=1}^{J}\exp(\boldsymbol{\beta}_m'\mathbf{x}_t)}, \; j = 0,...,J,$$

A possible J+1 *unordered* outcomes can occur. In order to identify the parameters of the model, we impose the normalization $\boldsymbol{\beta}_0 = \mathbf{0}$. This model is typically employed for individual or grouped data in which the '**x**' variables are characteristics of the observed individual(s), not the choices. For present purposes, that is the main distinction between this and the *discrete choice* model described in Chapter E38. The characteristics are the same across all outcomes. The study of occupational choice, by Schmidt and Strauss (1975) provides a well known application.

The data will appear as follows:

- Individual data: y_t coded 0, 1, ..., J,
- Grouped data: y_{0t}, y_{1t},...,y_{Jt} give proportions or shares.

In the grouped data case, a weighting variable, n_t, may also be provided if the observations happen to be frequencies. The proportions variables must range from zero to one and sum to one at each observation. The full set must be provided, even though one is redundant. The data are inspected to determine which specification is appropriate. The number of Lhs variables given and the coding of the data provide the full set of information necessary to estimate the model, so no additional information about the dependent variable is needed. There is a single line of data for each individual.

This model proliferates parameters. There are $J \times K$ nonzero parameters in all, since there is a vector $\boldsymbol{\beta}_j$ for each probability except the first. Consequently, even moderately sized models quickly become very large ones if your outcome variable, *y*, takes many values. The maximum number of parameters which can be estimated in a model is 150 as usual with the standard configuration. However, if you are able to forego certain other optional features, the number of parameters can increase to 300. The model size is detected internally. If your configuration contains more than 150 parameters, the following options and features become unavailable:

- marginal effects
- choice based sampling
- **; Rst** = list for imposing restrictions
- **; CML:** specification for imposing linear constraints
- **; Hold** for using the multinomial logit model as a sample selection equation

In addition, if your model size exceeds 150 parameters, the matrices *b* and *varb* cannot be retained. (But, see below for another way to retrieve large parameter matrices.)

The choice set should be restricted to no more than 25 choices. If you have more than 25 choices, the number of characteristics that may be used becomes very small. Nonetheless, it is possible to fit models with up to 100 choices by using **CLOGIT**, as discussed in Chapter E38. In addition, if you are able to make a few other compromises on the model specification, it is possible to fit models with up to 200 choices by using the panel data binary logit estimator – this is a 'trick'– as described below in Section E37.10.

E37.3 Model Command for the Multinomial Logit Model

The command for fitting this form of multinomial logit model is

MLOGIT ; Lhs = y or **y0,y1,...yJ**
; Rhs = regressors $

(The command may also be **LOGIT**, which is what has always been used in previous versions of *LIMDEP*.) All general options for controlling output and iterations are available except **; Keep = name**. (A program which can be used to obtain the fitted probabilities is listed below.) There are internally computed predictions for the multinomial logit model. The command builder for this model is found in Model:Discrete Choice/MLogit.

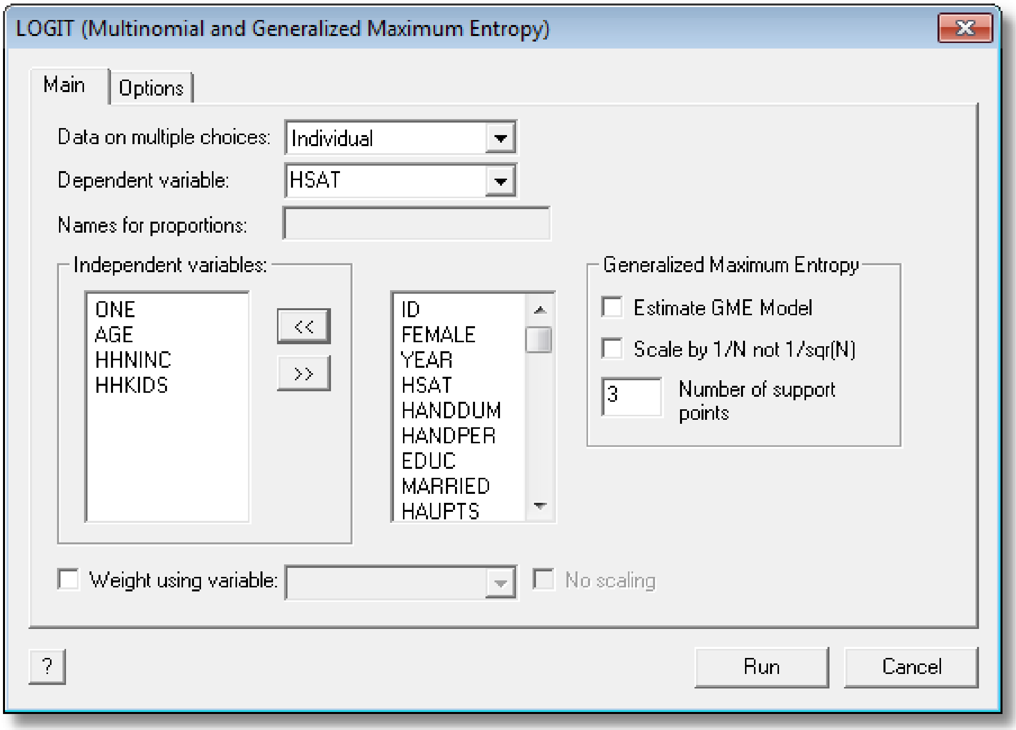


Figure E37.1 Command Builder for Multinomial Logit Models

Standard Model Specifications for the Multinomial Logit (MLOGIT) Model

This is the full list of general specifications that are applicable for this model:

Controlling Output from Model Commands

; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Choice uses choice based sampling (sandwich with weighting) estimated matrix.
; Cluster = name cluster form of corrected covariance estimator.
; Robust requests a ‘sandwich’ estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Prob = name	saves probabilities as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Imposing Constraints on Parameters

The **; Rst = list** form of restrictions is supported for imposing constraints on model parameters, either fixed value or equality. One possible application of the constrained model involves making the entire vector of coefficients in one probability equal that in another. You can do this as follows:

```
NAMELIST   ; x = the entire set of Rhs variables $
CALC       ; k = Col(x) $
LOGIT      ; Lhs  = y
           ; Rhs  = x
           ; Rst = k_b, k_b, ... , k_b $
```

This would force the corresponding coefficients in all probabilities to be equal. You could also apply this to some, but not all of the outcomes, as in

```
           ; Rst = k_b, k_b, k_b2, k_b3
```

> **HINT:** The coefficients in this model are not the marginal effects. But, forcing the coefficient on a characteristic in probability *j* to equal its counterpart in probability *m* also forces the two marginal effects to be equal.

Starting Values

The parameter vector for this model is a $J \times K$ column vector,

$$\Theta = [\beta_1' , \beta_2' , ..., \beta_J']'.$$

You may provide starting values with **; Start = list**.

E37.4 Choice Based Sampling and Robust Covariance Matrices

Choice Based Sampling

The choice based sampling methodology for individual data can be applied here. You must provide a weighting variable which gives the sampling ratios. The variable gives the ratio of the true, population proportion to the sample proportions. This presumes that you know the population proportions, $\phi_0,...,\phi_J$. If you know the sample proportions, $f_0,...,f_J$, as well, then you can calculate the necessary ratios, $w_0,...,w_J = \phi_j/f_j$ needed for the calculations to follow. With these in hand, you can create the weights using **RECODE** as follows:

```
CREATE    ; wts  = y (your dependent variable) $
RECODE    ; wts  ; 0 = weight for 0
                 ; 1 = weight for 1
          ; ... $
```

Perhaps a more convenient way to do the same computation is to create a vector with the weights,

MATRIX ; cbwt = [w_0, w_1,...,w_J] $

then you can use the following:

CREATE ; yplus1 = y + 1 ; wts = cbwt(yplus1) $ Zero is not a valid subscript.

Regardless, you must have the population proportions in hand. If you do not know the appropriate sample proportions, there is a special **MATRIX** function, Prpn(*variable*), for this purpose, which you can use as follows:

```
CREATE    ; yplus1 = y + 1 $
MATRIX    ; f = Prpn (yplus1) $
```

Since you have ϕ_j in hand, you can now proceed as follows:

MATRIX ; phi = [ϕ_0,...,ϕ_J] $ You provide the values.
MATRIX ; cbwt = diag(f) ; cbwt = phi * <cbwt> $
CREATE ; wts = cbwt(yplus1) $

(Note, the Prpn(*variable*) function is used specifically for this purpose. It creates a vector with one column and number of rows equal to the minimum of 100 and the maximum of *yplus*1. Values larger than 100 or less than one are discarded, and not counted in the proportions.)

Be sure to provide a sampling ratio for every outcome. With the weights in place, your **MLOGIT** command is

MLOGIT ; Lhs = y
; Rhs = regressors
; Wts = weights
; Choice Based Sampling $

This adjustment changes the estimator in two ways. First, the observations are weighted in computing the parameter estimates. Second, after estimation, the standard errors are adjusted. The estimator of the asymptotic covariance matrix for the choice based sampling case is

$$\text{Asy.Var}[\mathbf{b}_{CBWT}] = (\mathbf{-H})^{-1}\mathbf{BHHH}\,(\mathbf{-H})^{-1}$$

where the weighted matrices are constructed from the Hessian and first derivatives using

$$\partial^2 \log L/\partial\boldsymbol{\beta}_l\partial\boldsymbol{\beta}_m' = \Sigma_t\, w_t\{-[\mathbf{1}(l{=}m)P_l - P_lP_m\,]\}\mathbf{X'X}.$$

$$\partial \log L/\partial\boldsymbol{\beta}_{\text{j}} = \Sigma_t\, w_i(\text{d}_{tj} - t_{ij})\mathbf{x}_t \text{ where } d_{tj} = 1 \text{ if person } t \text{ makes choice } j;$$

$$\mathbf{BHHH}\text{(in blocks)} = \Sigma_t\, w_i(\text{d}_{tl} - P_{tl})(\text{d}_{tm} - P_{tm})\mathbf{x}_t\mathbf{x}_t'$$

and w_t = population frequency for choice made by individual t divided by sample proportion for choice made by individual t.

Generic Robust Covariance Matrix

It has become common in the literature to compute a ‘robust covariance matrix’ for the MLE. (The misspecification to which the matrix is robust is left unspecified in most cases.) The desired robust covariance matrix would result in the preceding computation if w_i equals one for all observations. This suggests a simple way to obtain it, just by specifying

; Choice Based ; Wts = one.

Alternatively, just use

; Robust

which is equivalent.

Cluster Correction

A related calculation is used when observations occur in groups which may be correlated. This is rather like a panel; one might use this approach in a random effects kind of setting in which observations have a common latent heterogeneity. The parameter estimator is unchanged in this case, but an adjustment is made to the estimated asymptotic covariance matrix. The calculation is done as follows: Suppose the n observations are assembled in C clusters of observations, in which the number of observations in the cth cluster is n_c. Thus,

$$\sum_{c=1}^{C} n_c = n.$$

Denote by $\boldsymbol{\beta}$ the full set of model parameters, $[\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_J']'$. Let the observation specific gradients and Hessians for individual i in cluster c be

$$\mathbf{g}_{ic} = \frac{\partial \log L_{ic}}{\partial \boldsymbol{\beta}}$$

$$\mathbf{H}_{ic} = \frac{\partial^2 \log L_{ic}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

The uncorrected estimator of the asymptotic covariance matrix based on the Hessian is

$$\mathbf{V}_H = -\mathbf{H}^{-1} = \left(-\sum_{c=1}^{C} \sum_{i=1}^{n_c} \mathbf{H}_{ic}\right)^{-1}$$

The corrected asymptotic covariance matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{V}_H \frac{C}{C-1}\left[\sum_{c=1}^{C}\left(\sum_{i=1}^{n_c} \mathbf{g}_{ic}\right)\left(\sum_{i=1}^{n_c} \mathbf{g}_{ic}\right)'\right] \mathbf{V}_H$$

Note that if there is exactly one observation per cluster, then this is $C/(C\text{-}1)$ times the sandwich (robust) estimator discussed above. Also, if you have fewer clusters than parameters, then this matrix is singular – it has rank equal to the minimum of C and JK, the number of parameters. This estimator is requested with

; Cluster = specification

where the specification is either a fixed number of observations per cluster, or an identifier that distinguishes clusters, such as an identification number. This estimator can also be extended to stratified as well as clustered data, using

; Stratum = specification

The full description of using these procedures appears in Chapter R10.

E37.5 Output for the Logit Models

Initial ordinary least squares results are used for the starting values for this model. For individual data, J binary variables are implied by the model. These are used in a least squares regression. For the grouped data case, a minimum chi squared, generalized least squares estimate is obtained by the weighted regression of

$$q_{ij} = \log(P_{ij} / P_{i0})$$

on the regressors, with weights $w_{ij} = (n_i P_{ij} P_{i0})^{1/2}$ (n_i may be 1.0). The OLS estimates based on the individual data are inconsistent, but the grouped data estimates are consistent (and, in the binomial case, efficient). The least squares estimates are included in the displayed results by including

; OLS

in the model command. The iterations are followed by the maximum likelihood estimates with the usual diagnostic statistics. An example is shown below.

NOTE: Minimum chi squared (MCS) is an estimator, not a model. Moreover, the MCS estimator has the same properties as, but is different from the maximum likelihood estimator. Since the MCS estimator in *LIMDEP* is not iterated, it should not be used as the final results of estimation. Without iteration, the MCS estimator is not a fixed point – the weights are functions only of the sample proportions, not the parameters. For current purposes, these are only useful as starting values.

Standard output for the logit model will begin with a table such as the following which results from estimation of a model in which the dependent variable takes values 0,1,2,3,4,5:

```
SAMPLE   ; All $
REJECT   ; hsat > 5 $
LOGIT    ; Lhs = hsat ; Rhs = one,educ,hhninc,age,hhkids $
```

(This is based on the health satisfaction variable analyzed in the preceding chapter. We reduced the sample to those with *hsat* reported zero to five. We would note, though these make for a fine numerical example, the multinomial logit model would be inappropriate for these ordered data.) The restricted log likelihood is computed for a model in which *one* is the only Rhs variable. In this case,

$$\log L_0 = \Sigma_j n_j \log P_j$$

where n_j is the number of individuals who choose outcome j and $P_j = n_j/n$ = the jth sample proportion. The chi squared statistic is $2(\log L - \log L_0)$. If your model does not contain a constant term, this statistic need not be positive, in which case it is not reported. But, even if it is computable, the statistic is meaningless if your model does not contain a constant.

The diagnostic statistics are followed by the coefficient estimates: These are $\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_J$. Recall $\boldsymbol{\beta}_0$ is normalized to zero, and not reported.

```
-----------------------------------------------------------------------------
Multinomial Logit Model
Dependent variable                   HSAT
Log likelihood function    -11246.96937
Restricted log likelihood  -11308.02002
Chi squared [  20 d.f.]       122.10132
Significance level               .00000
McFadden Pseudo R-squared      .0053989
Estimation based on N =   8140, K =  25
Inf.Cr.AIC  =22543.939 AIC/N =    2.770
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|   -1.77566**       .69486    -2.56  .0106    -3.13756   -.41376
    EDUC|      .07326        .04476     1.64  .1017     -.01447    .16099
  HHNINC|      .28572        .58129      .49  .6231     -.85359   1.42503
     AGE|      .00566        .00838      .68  .4996     -.01077    .02209
  HHKIDS|      .27188        .19642     1.38  .1663     -.11311    .65686
        |Characteristics in numerator of Prob[Y = 2]
Constant|    -.54217         .54866     -.99  .3231    -1.61752    .53318
    EDUC|      .06152*       .03617     1.70  .0890     -.00937    .13240
  HHNINC|      .85929*       .44943     1.91  .0559     -.02158   1.74017
     AGE|     -.00090        .00651     -.14  .8903     -.01365    .01185
  HHKIDS|      .13921        .15530      .90  .3700     -.16517    .44359
        |Characteristics in numerator of Prob[Y = 3]
Constant|    -.25433         .49206     -.52  .6053    -1.21876    .71010
    EDUC|      .10996***     .03247     3.39  .0007      .04632    .17359
  HHNINC|     1.54517***     .40167     3.85  .0001      .75791   2.33242
     AGE|     -.00955        .00584    -1.64  .1017     -.02099    .00189
  HHKIDS|      .08178        .14014      .58  .5595     -.19289    .35645
        |Characteristics in numerator of Prob[Y = 4]
Constant|      .09378        .48301      .19  .8461     -.85291   1.04047
    EDUC|      .10453***     .03202     3.26  .0011      .04178    .16729
  HHNINC|     1.74362***     .39382     4.43  .0000      .97175   2.51550
     AGE|     -.01430**      .00571    -2.50  .0123     -.02550   -.00310
  HHKIDS|      .19549        .13660     1.43  .1524     -.07224    .46321
        |Characteristics in numerator of Prob[Y = 5]
Constant|     1.58459***     .45170     3.51  .0005      .69927   2.46991
    EDUC|      .07527**      .03035     2.48  .0131      .01579    .13475
  HHNINC|     1.64030***     .37209     4.41  .0000      .91101   2.36959
     AGE|     -.01481***     .00526    -2.82  .0049     -.02512   -.00450
  HHKIDS|      .19988        .12655     1.58  .1142     -.04815    .44791
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The statistical output for the coefficient estimates is followed by a table of predicted and actual frequencies, such as the following: This table is requested by adding

; Summary

to the **MLOGIT** command.

```
Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted
------  -----------------------------  +  -----
Actual      0    1    2    3    4    5  |  Total
------  -----------------------------  +  -----
  0         0    0    0    0    0  447  |    447
  1         0    0    0    0    0  255  |    255
  2         0    0    0    0    0  642  |    642
  3         0    0    0    0    0 1173  |   1173
  4         0    0    0    0    0 1390  |   1390
  5         0    0    0    0    0 4233  |   4233
------  -----------------------------  +  -----
Total       0    0    0    0    0 8140  |   8140
```

The prediction for any observation is the cell with the largest predicted probability for that observation.

NOTE: If you have more than three outcomes, it is very common, as occurred above, for the model to predict zero outcomes in one or more of the cells. Even in a model with very high t ratios and great statistical significance, it takes a very well developed model to make predictions in all cells.

The **; List** specification produces a listing such as the following:

```
Predicted Values          (* => observation was not in estimating sample.)
Observation         Observed Y    Predicted Y   Residual      MaxPr(i)  Prob[Y*=y]
       20           2.0000000    5.0000000      .000000      .6845695     .0631146
       24             .000000    4.0000000      .000000      .3196778     .0885942
       38           5.0000000    5.0000000      .000000      .6041918     .6041918
       39           2.0000000    5.0000000      .000000      .6439476     .1224276
       57           5.0000000    5.0000000      .000000      .5050133     .5050133
       59           5.0000000    5.0000000      .000000      .4284611     .4284611
       60           5.0000000    5.0000000      .000000      .4173034     .4173034
```

In the listing, the MaxPr(i) is the probability attached to the outcome with the largest predicted probability; the outcome is shown as the Predicted Y. The last column shows the predicted probability for the observed outcome. Residuals are not computed – there is no significance to the reported zero. (The results above illustrate the format of the table. They were complete with a small handful of observations, not the 8,140 used to fit the model shown earlier.)

The results kept for further use are:

Matrices: *b* and *varb*.
b_logit = $(J+1)\times K$.
This additional matrix contains the parameters arranged so that $\boldsymbol{\beta}_j'$ is the jth row. The first row is zero. This matrix can be used to obtain fitted probabilities, as discussed below.

Scalars: *kreg*, *nreg*, *logl*, and *exitcode*.

Labels for **WALD** are constructed from the outcome and variable numbers. For example, if there are three outcomes and **; Rhs = one,x1,x2**, the labels will be

Last Model: [*b*1_1,*b*1_2,*b*1_3,*b*2_1,*b*2_2,*b*2_3].

Last Function: Prob($y = J|\mathbf{x}$).
You may specify other outcomes in the **PARTIALS** and **SIMULATE** commands.

E37.6 Partial Effects

The partial effects in this model are

$$\boldsymbol{\delta}_j = \partial P_j/\partial \mathbf{x},\ \ j = 0,1,...,J.$$

For the present, ignore the normalization $\boldsymbol{\beta}_0 = \mathbf{0}$. The notation P_j is used for Prob[y = j]. After some tedious algebra, we find

$$\boldsymbol{\delta}_j = P_j(\boldsymbol{\beta}_j - \bar{\boldsymbol{\beta}})$$

where
$$\bar{\boldsymbol{\beta}} = \sum_{j=0}^{J} P_j \boldsymbol{\beta}_j.$$

It follows that neither the sign nor the magnitude of $\boldsymbol{\delta}_j$ need bear any relationship to those of $\boldsymbol{\beta}_j$. (This is worth bearing in mind when reporting results.) The asymptotic covariance matrix for the estimator of $\boldsymbol{\delta}_j$ would be computed using

$$\text{Asy.Var.}\left[\hat{\boldsymbol{\delta}}_j\right] = \mathbf{G}_j \,\text{Asy.Var}\left[\hat{\boldsymbol{\beta}}\right]\mathbf{G}_j'$$

where $\boldsymbol{\beta}$ is the full parameter vector. It can be shown that

$$\text{Asy.Var.}\left[\hat{\boldsymbol{\delta}}_j\right] = \Sigma_l \Sigma_m\ \mathbf{V}_{jl}\,\text{Asy.Cov.}[\hat{\boldsymbol{\beta}}_l,\ \hat{\boldsymbol{\beta}}_m']\mathbf{V}_{jm}',\ j=0,...,J,$$

where
$$\mathbf{V}_{jl} = [\mathbf{1}(j = l) - P_l]\{P_j\mathbf{I} + \boldsymbol{\delta}_j\mathbf{x}'\} - P_j\boldsymbol{\delta}_l\mathbf{x}'$$

and
$$\mathbf{1}(j = l) = 1 \text{ if } j = l, \text{ and } 0 \text{ otherwise.}$$

E37.6.1 Internal Computation of Partial Effects

This full set of results is produced automatically when your **LOGIT** command includes

; Partial Effects (or just ; Partials).

NOTE: Marginal effects are computed at the sample averages of the Rhs variables in the model.

There is no conditional mean function in this model, so marginal effects are interpreted a bit differently from the usual case. What is reported are the derivatives of the probabilities. (Note this is the same as the ordered probability models.) These derivatives are saved in a matrix named *partials* which has J+1 rows and K columns. Each row is the vector of partial effects of the corresponding probability. Since the probabilities will always sum to one, the column sums in this matrix will always be zero. That is,

MATRIX ; List ; 1 ' partials $

will display a row matrix of zeros. The elasticities of the probabilities, $(\partial P_j/\partial x_k)\times(x_k/Pj)$ are placed in a $(J+1)\times K$ matrix named *elast_ml*. The format of the results is illustrated in the example below.

```
-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are  All Obs.
A full set is given for the entire set of
outcomes, HSAT     =  0 to HSAT     =   5
Probabilities at the mean values of X are
  0= .052 1= .030 2= .078 3= .145 4= .171
  5= .523
--------+--------------------------------------------------------------------
        |     Partial                         Prob.      95% Confidence
    HSAT|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Marginal effects on Prob[Y = 0]
    EDUC|    -.00415***     -.87310    -2.87  .0042     -.00699   -.00131
  HHNINC|    -.07533***     -.48081    -4.28  .0000     -.10982   -.04085
     AGE|     .00059**       .53969     2.36  .0184      .00010    .00109
  HHKIDS|    -.00875        -.05610    -1.44  .1505     -.02067    .00317
        |Marginal effects on Prob[Y = 1]
    EDUC|    -.00021        -.07636     -.21  .8331     -.00220    .00178
  HHNINC|    -.03570***     -.38652    -2.64  .0083     -.06222   -.00918
     AGE|     .00052***      .80559     2.62  .0087      .00013    .00091
  HHKIDS|     .00313         .03408      .68  .4994     -.00596    .01222
        |Marginal effects on Prob[Y = 2]
    EDUC|    -.00147        -.20405     -.92  .3557     -.00458    .00165
  HHNINC|    -.04677**      -.19725    -2.31  .0211     -.08652   -.00703
     AGE|     .00083***      .49750     2.67  .0075      .00022    .00144
  HHKIDS|    -.00234        -.00993     -.32  .7478     -.01662    .01194
```

```
       |Marginal effects on Prob[Y = 3]
    EDUC|     .00430**       .32277     2.29  .0218      .00063    .00797
  HHNINC|     .01276         .02908      .53  .5938    -.03413    .05965
     AGE|     .00028         .09081      .70  .4822    -.00050    .00106
  HHKIDS|    -.01265        -.02898    -1.35  .1760    -.03097    .00567
        |Marginal effects on Prob[Y = 4]
    EDUC|     .00416**       .26381     2.07  .0385      .00022    .00810
  HHNINC|     .04913**       .09457     1.98  .0482      .00040    .09787
     AGE|    -.00048        -.13248    -1.14  .2552    -.00132    .00035
  HHKIDS|     .00452         .00874      .46  .6444    -.01466    .02370
        |Marginal effects on Prob[Y = 5]
    EDUC|    -.00262        -.05450     -.94  .3475    -.00809    .00285
  HHNINC|     .09591***      .06048     2.78  .0054      .02827    .16355
     AGE|    -.00174***     -.15634    -3.07  .0021    -.00285   -.00063
  HHKIDS|     .01609         .01020     1.23  .2205    -.00965    .04183
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Marginal Effects Averaged Over Individuals

Variable	Y=00	Y=01	Y=02	Y=03	Y=04	Y=05
ONE	-.0377	-.0772	-.0975	-.1380	-.1051	.4556
EDUC	-.0044	-.0002	-.0014	.0043	.0042	-.0025
HHNINC	-.0786	-.0361	-.0459	.0136	.0494	.0977
AGE	.0006	.0005	.0008	.0003	-.0005	-.0018
HHKIDS	-.0092	.0033	-.0023	-.0125	.0045	.0162

Averages of Individual Elasticities of Probabilities

Variable	Y=00	Y=01	Y=02	Y=03	Y=04	Y=05
ONE	-.7050	-2.4807	-1.2472	-.9593	-.6112	.8796
EDUC	-.8732	-.0764	-.2041	.3227	.2638	-.0545
HHNINC	-.4847	-.3904	-.2011	.0252	.0907	.0566
AGE	.5315	.7974	.4894	.0827	-.1406	-.1645
HHKIDS	-.0571	.0330	-.0110	-.0300	.0077	.0092

Matrix - B_LOGIT

[6, 5] Cell:

	1	2	3	4	5
1	0	0	0	0	0
2	-1.77566	0.0732571	0.285721	0.00565832	0.271876
3	-0.542169	0.0615164	0.859294	-0.000897661	0.13921
4	-0.254329	0.109956	1.54517	-0.00955207	0.081778
5	0.0937819	0.104535	1.74362	-0.0143038	0.195486
6	1.58459	0.0752677	1.6403	-0.0148114	0.199883

Matrix - PARTIALS

[6, 5] Cell: -0.0368127

	1	2	3	4	5
1	-0.0368127	-0.00415059	-0.0753323	0.000593779	-0.00874508
2	-0.0758147	-0.000213985	-0.0356972	0.000522449	0.00313091
3	-0.098142	-0.00146816	-0.0467745	0.000828444	-0.00234229
4	-0.139903	0.00429655	0.0127595	0.000279777	-0.0126482
5	-0.105991	0.00415859	0.0491332	-0.000483331	0.00451648
6	0.456663	-0.0026224	0.0959113	-0.00174112	0.0160882

Matrix - ELAST_ML

[6, 5] Cell: -0.712012

	1	2	3	4	5
1	-0.712012	-0.873102	-0.480807	0.539688	-0.0561038
2	-2.48767	-0.0763641	-0.386522	0.805587	0.0340761
3	-1.25418	-0.204054	-0.197249	0.497504	-0.00992853
4	-0.966341	0.322768	0.0290829	0.0908123	-0.0289784
5	-0.61823	0.263811	0.0945706	-0.132481	0.00873817
6	0.872575	-0.054497	0.0604751	-0.156338	0.0101966

Figure E37.2 Matrices Created by MLOGIT

Marginal effects are computed by averaging the effects over individuals rather than computing them at the means. The difference between the two is likely to be quite small. Current practice favors the averaged individual effects, rather than the effects computed at the means. **MLOGIT** also reports elasticities with the marginal effects. An example appears above.

E37.6.2 Partial Effects Using PARTIALS

The **; Partials** specification in the **MLOGIT** command computes the partial effects at the means of the variables. The post estimation command, **PARTIAL EFFECTS** (or just **PARTIALS**), can be used to compute average partial effects, and to compute various simulations of the outcome. For example, we compute the partial effects on Prob(*hsat* = 5|**x**) for the model estimated above with

```
SAMPLE      ; All $
REJECT      ; hsat > 5 $
LOGIT       ; Lhs = hsat ; Rhs = one,educ,hhninc,age,hhkids ; Partials $
PARTIALS    ; Effects: educ / hhninc / age / hhkids ; Summary $
```

The first results below are those reported earlier. The second set are the average partial effects. (The similarity is striking.)

```
-----------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
    HSAT|      Effect    Elasticity      z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Marginal effects on Prob[Y = 5]
    EDUC|    -.00262        -.05450     -.94  .3475     -.00809    .00285
  HHNINC|     .09591***      .06048     2.78  .0054      .02827    .16355
     AGE|    -.00174***     -.15634    -3.07  .0021     -.00285   -.00063
  HHKIDS|     .01609         .01020     1.23  .2205     -.00965    .04183
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------


---------------------------------------------------------------------
Partial Effects for Multinomial Logit Probability Y =  5
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial     Standard
(Delta method)     Effect       Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      EDUC         -.00249      .00279     .89      -.00796       .00298
      HHNINC        .09767      .03445    2.84       .03015       .16519
      AGE          -.00175      .00056    3.11      -.00286      -.00065
   *  HHKIDS        .01592      .01310    1.22      -.00976       .04160
---------------------------------------------------------------------
```

The various optional specifications in **PARTIALS** may be used here. For example,

```
PARTIALS    ; Effects: hhkids & hhninc=.05(.5)3 ; Outcome = 4 ; Plot $
```

plots the effect of *hhkids* on Prob(*hsat*=4) for several values of *hhninc*.

E37.7 Predicted Probabilities

Predicted probabilities can be computed automatically for the multinomial logit model. Since there are multiple outcomes, this must be handled a bit differently from other models. The procedure is as follows: Request the computation with

; Prob = name

as you would normally for a discrete choice model. However, for this model, *LIMDEP* does the following:

1. A namelist is created with name consisting of up to the first four letters of '*name*' and *prob* is appended to it. Thus, if you use **; Prob = pfit**, the namelist will be named *pfitprob*.

2. The set of variables, one for each outcome, are named with the same convention, with *prjj* instead of *prob*.

For example, in a five outcome model, the specification

; Prob = job

produces a namelist

jpbprob = *jobpr*00, *jobpr*01, *jobpr*02, *jobpr*03, *jobpr*04.

For our running example,

; Prob = hsat

produces the namelist named *hsatprob* and variables *hsatpr00*, *hsatpr01*, …, *hsatpr05*. The variables will then contain the respective probabilities. You may also use

; Fill

with this procedure to compute probabilities for observations that were not in the sample. Observations which contain missing data are bypassed as usual.

E37.8 Generalized Maximum Entropy (GME) Estimation

This is an alternative estimator for the multinomial logit model. The GME criterion is based on the entropy of the probability distribution,

$$E(p_0,...,p_J) = -\Sigma_j\, p_j \ln p_j.$$

The implementation of the GME estimator in *LIMDEP*'s multinomial logit model is done by augmenting the likelihood function with a term that accounts for the entropy of the choice probability set. Let

H = the number of support points for the entropy distribution.

and V = an H specific set of weights. These are

$$
\begin{aligned}
V &= -1/\sqrt{N},\ +1/\sqrt{N} && \text{for } H = 2\\
&= -1/\sqrt{N},\ 0,\ +1/\sqrt{N} && \text{for } H = 3\\
&= -1/\sqrt{N},\ -.5/\sqrt{N},\ [0],\ +.5/\sqrt{N},\ +1/\sqrt{N} && \text{for } H = 4 \text{ or } 5\\
&= \ ...\ [0],\ +.33/\sqrt{N},\ +.67/\sqrt{N},\ +1/\sqrt{N} && \text{for } H = 6 \text{ or } 7\\
&= \ ...\ [0],\ +.25/\sqrt{N},\ +.50/\sqrt{N},\ +.75/\sqrt{N},\ +1/\sqrt{N} && \text{for } H = 8 \text{ or } 9
\end{aligned}
$$

(You may optionally choose to scale the entire V by $1/\sqrt{N}$). Then,

$$\Psi_{ij} = \sum_{h=1}^{H} \exp[V_h \boldsymbol{\beta}_j' \mathbf{x}_i]$$

Then, the additional term which augments the contribution to the log likelihood for individual i is

$$F_{\Psi i} = \sum_{j=0}^{J} \ln \Psi_{ij}$$

This estimator is invoked simply by adding

; GME = the number of support points, H

to the **LOGIT** command. You may choose to scale the weighting vector with

; Scale

You may also choose the GME estimator in the command builder, as shown in Figure E37.1 earlier.

In the example below, we have treated the self reported health satisfaction measure as a discrete choice (doubtlessly inappropriately – just for the purpose of a numerical example). The first set of estimates given are the GME results. The model is refit by maximum likelihood in the second set. As can be seen, the GME estimator triggers some additional results in the table of summary statistics. It also brings some relatively modest changes in the estimated parameters.

```
-----------------------------------------------------------------------------
Generalized Maximum Entropy (Logit)
Dependent variable                  HSAT
Log likelihood function   -106287.21094
Estimation based on N =   8140, K =  25
Number of support points =          7
Weights in support scaled to 1/sqr(N)
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|   -1.76249**       .69184    -2.55  .0108    -3.11848   -.40650
    EDUC|     .07199         .04453     1.62  .1059     -.01529    .15926
  HHNINC|     .26975         .57843      .47  .6410     -.86396   1.40346
     AGE|     .00570         .00835      .68  .4951     -.01067    .02207
  HHKIDS|     .26950         .19568     1.38  .1684     -.11402    .65302
        |Characteristics in numerator of Prob[Y = 2]
Constant|    -.53230         .54599     -.97  .3296    -1.60243    .53782
    EDUC|     .06033*        .03595     1.68  .0933     -.01012    .13078
  HHNINC|     .84177*        .44699     1.88  .0597     -.03432   1.71786
     AGE|    -.00083         .00648     -.13  .8986     -.01353    .01188
  HHKIDS|     .13734         .15466      .89  .3745     -.16579    .44047
        |Characteristics in numerator of Prob[Y = 3]
Constant|    -.24497         .48927     -.50  .6166    -1.20392    .71398
    EDUC|     .10879***      .03223     3.38  .0007      .04562    .17197
  HHNINC|    1.52790***      .39910     3.83  .0001      .74567   2.31013
     AGE|    -.00948         .00581    -1.63  .1030     -.02087    .00191
  HHKIDS|     .07994         .13948      .57  .5666     -.19344    .35332
        |Characteristics in numerator of Prob[Y = 4]
Constant|     .10311         .48018      .21  .8300     -.83803   1.04426
    EDUC|     .10338***      .03178     3.25  .0011      .04108    .16567
  HHNINC|    1.72645***      .39122     4.41  .0000      .95966   2.49323
     AGE|    -.01423**       .00569    -2.50  .0124     -.02538   -.00308
  HHKIDS|     .19367         .13593     1.42  .1542     -.07276    .46009
        |Characteristics in numerator of Prob[Y = 5]
Constant|    1.59393***      .44877     3.55  .0004      .71437   2.47350
    EDUC|     .07412**       .03010     2.46  .0138      .01512    .13312
  HHNINC|    1.62344***      .36941     4.39  .0000      .89940   2.34748
     AGE|    -.01474***      .00523    -2.82  .0049     -.02500   -.00448
  HHKIDS|     .19810         .12585     1.57  .1155     -.04857    .44477
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
| Information Statistics for Discrete Choice Model.                    |
|                           M=Model MC=Constants Only   M0=No Model    |
| Criterion F (log L)  -106287.21094     -106347.98256 -109623.17376   |
| LR Statistic vs. MC      121.54324            .00000        .00000   |
| Degrees of Freedom        20.00000            .00000        .00000   |
| Prob. Value for LR          .00000            .00000        .00000   |
| Entropy for probs.     11250.94128       11311.43749   14584.92208   |
| Normalized Entropy          .77141            .77556       1.00000   |
| Entropy Ratio Stat.     6667.96160        6546.96918        .00000   |
| Bayes Info Criterion      26.13692          26.15185      26.95656   |
| BIC(no model) - BIC         .81965            .80472        .00000   |
| Pseudo R-squared            .22859            .00000        .00000   |
| Pct. Correct Pred.        52.00246          52.00246      16.66667   |
| Means:       y=0    y=1    y=2    y=3    y=4    y=5     y=6   y>=7   |
| Outcome     .0549  .0313  .0789  .1441  .1708  .5200  .0000  .0000   |
| Pred.Pr     .0552  .0314  .0788  .1440  .1707  .5199  .0000  .0000   |
| Notes: Entropy computed as Sum(i)Sum(j)Pfit(i,j)*logPfit(i,j).       |
|        Normalized entropy is computed against M0.                    |
|        Entropy ratio statistic is computed against M0.               |
|        BIC = 2*criterion - log(N)*degrees of freedom.                |
|        If the model has only constants or if it has no constants,    |
|        the statistics reported here are not useable.                 |
+----------------------------------------------------------------------+
```

E37.9 Technical Details on Optimization

Newton's method is used to obtain the estimates in all cases. The log likelihood function for the multinomial logit model is

$$\log L = \Sigma_t \Sigma_j d_{tj} \log P_{tj},$$

where P_{tj} is the probability defined earlier and $d_{tj} = 1$ if $y_t = j$, 0 otherwise, $j = 0,...,J$ or d_{tj} equals the proportion for choice j for individual t in the grouped data case. The first and second derivatives are

$$\partial \log L / \partial \boldsymbol{\beta}_j = \Sigma_t (d_{tj} - P_{tj})\mathbf{x}_t.$$

$$\partial^2 \log L / \partial \boldsymbol{\beta}_l \partial \boldsymbol{\beta}_m' = \Sigma_t -[\mathbf{1}(l=m)P_{tl} - P_{tl}P_{tm}]\mathbf{x}_t\mathbf{x}_t'.$$

The negative inverse of the Hessian provides the asymptotic covariance matrix.

The log likelihood function for the multinomial logit model is globally concave. With the exception of OLS and possibly the Poisson regression model, this is the most benign optimization problem in *LIMDEP*, and convergence should always be routine. As such, you should not need to change the default algorithm or the convergence criteria. If you do observe convergence problems, such as more than a handful of iterations, you should suspect the data. Occasionally, a data set will contain some peculiarities that impede Newton's method. In most cases, switching the algorithm to BFGS with

; Alg = BFGS

will solve the problem.

E37.10 Panel Data Multinomial Logit Models

The random parameters model described in Chapter R24 is useful for constructing two types of panel data structures for the multinomial logit model, random effects and a dynamic model.

E37.10.1 Random Effects and Common (True) Random Effects

The structural equations of the multinomial logit model are

$$U_{ijt} = \boldsymbol{\beta}_j'\mathbf{x}_{it} + \varepsilon_{ijt},\ t = 1,...,T_i,\ j = 0,1,...,J,\ i = 1,...,N,$$

where U_{ijt} gives the utility of choice j by person i in period t – we assume a panel data application with $t = 1,...,T_i$. The model about to be described can be applied to cross sections, where $T_i = 1$. Note also that as usual, we assume that panels may be unbalanced. We also assume that ε_{ijt} has a type 1 extreme value (Gumbel) distribution and that the J random terms are independent. Finally, we assume that the individual makes the choice with maximum utility. Under these (IIA inducing) assumptions, the probability that individual i makes choice j in period t is

$$P_{ijt} = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it})}{\sum_{j=0}^{J}\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it})}.$$

Note that this is the MLOGIT form of the model – the Rhs data are in the form of individual characteristics, not attributes of the choices. That would be handled by **CLOGIT**, discussed in Chapter E38. We now suppose that individual i has latent, unobserved, time invariant heterogeneity that enters the utility functions in the form of a random effect, so that

$$U_{ijt} = \boldsymbol{\beta}_j'\mathbf{x}_{it} + \alpha_{ij} + \varepsilon_{ijt},\ t = 1,...,T_i,\ j = 0,1,...,J,\ i = 1,...,N.$$

The resulting choice probabilities, conditioned on the random effects, are

$$P_{ijt} \mid \alpha_{i1},...,\alpha_{iJ} = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it} + \alpha_{ij})}{\sum_{j=0}^{J}\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it} + \alpha_{ij})}.$$

To complete the model, we assume that heterogeneity is normally distributed with zero means and $(J+1)\times(J+1)$ covariance matrix, Σ. For identification purposes, one of the coefficient vectors must be normalized to zero and one of the α_{ij}s is set to zero. We normalize the first element – subscript 0 – to zero. For convenience, this normalization is left implicit in what follows. It is automatically imposed by the software. To allow the remaining random effects to be freely correlated, we write the $J\times 1$ vector of nonzero αs as

$$\boldsymbol{\alpha}_i = \boldsymbol{\Gamma}\,\mathbf{v}_i$$

where $\boldsymbol{\Gamma}$ is a lower triangular matrix to be estimated and $\mathbf{v}_i$ is a standard normally distributed (mean zero, covariance matrix, $\mathbf{I}$) vector.

The preceding extends the random effects model to the multinomial logit framework. It is also of the form of *LIMDEP*'s other random parameter models, which is how we do the estimation, by maximum simulated likelihood. (See Section R24.7.) There are two additional versions of the essential structure:

1. Independent effects: $\boldsymbol{\Gamma}$ = A diagonal matrix.
2. True random effects: $\boldsymbol{\Gamma}$ = A diagonal matrix,
   and $v_{ji} = v_i$ = the same random variable in all utility functions.

Thus, in the second case, the preference heterogeneity is a choice invariant characteristic of the person.

The command structure for this model has two parts. In the first, the logit model is fit without the effects in order to obtain the starting values. In the second, we use a standard form of the random parameters model, as described in Chapter R24.

```
MLOGIT    ; Lhs  = dependent variable
          ; Rhs  = list of variables including one $
MLOGIT    ; Lhs  = dependent variable
          ; Rhs  = list of variables including one
          ; RPM ; Fcn = one(n)
          [; Halton]
          [; Pts = ...]
          ; Pds  = panel specification  $
```

The items in the square brackets are optional. This requests the type 1, independent effects model. To estimate the second model, type 2, true random effects model, add

; Common Effect

to the commands. To fit the general model with freely correlated effects, use, instead,

; Correlated

Partial effects for this model are obtained using the procedure shown in Section E37.6.2.

To illustrate this estimator, we constructed an example using the health care data. The Lhs variable is health satisfaction. We restricted the sample by first, keeping only groups with T_i = 7. We then eliminated all observations with Lhs variable greater than four. This leaves a dependent variable that takes five outcomes, 0,1,2,3,4, and a total sample of 905 observations in 394 groups ranging in size from one to seven. So, the resulting panel is unbalanced. The Rhs variables are *one*, *age*, *income* and *hhkids* that is kids in the household. We fit all three models described above.

The commands are as follows:

```
REJECT      ; _groupti < 7 $
REJECT      ; hsat > 4 $
REGRESS     ; Lhs = one  ; Rhs = one ; Str = id ; Panel $
MLOGIT      ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids $
MLOGIT      ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids
            ; RPM ; Fcn = one(n)
            ; Halton ; Pts = 50
            ; Pds  = _groupti
            ; Common $
MLOGIT      ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids ; Quietly $
MLOGIT      ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids
            ; RPM ; Fcn = one(n)
            ; Halton ; Pts = 50
            ; Pds  = _groupti $
MLOGIT      ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids ; Quietly $
MLOGIT      ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids
            ; RPM ; Fcn = one(n)
            ; Halton ; Pts = 50
            ; Pds  = _groupti
            ; Correlated $
```

These are the initial values, without latent effects.

```
-----------------------------------------------------------------------------
Multinomial Logit Model
Dependent variable                 HSAT
Log likelihood function     -1289.68419
Restricted log likelihood   -1295.05441
Chi squared [  12 d.f.]        10.74042
Significance level               .55129
McFadden Pseudo R-squared      .0041467
Estimation based on N =    905, K =  16
Inf.Cr.AIC  = 2611.368 AIC/N =    2.885
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     -.97586       1.20831     -.81   .4193    -3.34410   1.39238
     AGE|      .00500        .02273      .22   .8259     -.03954    .04954
  HHNINC|      .29496       1.23304      .24   .8109    -2.12176   2.71167
  HHKIDS|      .47793        .42941     1.11   .2657     -.36370   1.31957
        |Characteristics in numerator of Prob[Y = 2]
Constant|     -.58489        .93591     -.62   .5320    -2.41923   1.24946
     AGE|      .01279        .01758      .73   .4667     -.02166    .04724
  HHNINC|     1.48473        .93548     1.59   .1125     -.34877   3.31823
  HHKIDS|      .22135        .33932      .65   .5142     -.44370    .88641
        |Characteristics in numerator of Prob[Y = 3]
Constant|     1.05098        .84361     1.25   .2128     -.60247   2.70442
     AGE|     -.00744        .01590     -.47   .6400     -.03860    .02373
  HHNINC|     1.28703        .87733     1.47   .1424     -.43251   3.00657
  HHKIDS|     -.03754        .31211     -.12   .9043     -.64926    .57419
```

```
        |Characteristics in numerator of Prob[Y = 4]
Constant|     .56268          .83149       .68  .4986    -1.06700   2.19237
     AGE|     .00343          .01564       .22  .8263     -.02723    .03409
  HHNINC|    1.55568*         .85486      1.82  .0688     -.11982   3.23118
  HHKIDS|     .30585          .30374      1.01  .3140     -.28946    .90116
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This model has a separate, independent effect in each utility function.

```
+---------------------------------------------+
| Random Coefficients  MltLogit Model         |
| Dependent variable                 HSAT     |
| Log likelihood function     -1232.79687     |
| Estimation based on N =     905, K =  20    |
| Inf.Cr.AIC  = 2505.594 AIC/N =    2.769     |
| Model estimated: Jul 21, 2011, 22:49:15     |
| Unbalanced panel has    394 individuals     |
+---------------------------------------------+

-----------------------------------------------------------------------------
Random Coefficients  MltLogit Model
All parameters have the same random effect
Multinomial logit with random effects
Simulation based on  50 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .00522          .01994       .26  .7936     -.03387    .04431
  HHNINC|     .18002         1.04166       .17  .8628    -1.86160   2.22165
  HHKIDS|     .48013          .38705      1.24  .2148     -.27848   1.23874
     AGE|     .02077          .01814      1.15  .2520     -.01477    .05632
  HHNINC|    1.20948          .82664      1.46  .1434     -.41070   2.82967
  HHKIDS|     .23686          .35048       .68  .4992     -.45007    .92379
     AGE|     .00077          .01694       .05  .9636     -.03243    .03397
  HHNINC|     .96235          .86369      1.11  .2652     -.73045   2.65516
  HHKIDS|    -.01765          .35090      -.05  .9599     -.70539    .67010
     AGE|     .01048          .01741       .60  .5472     -.02364    .04460
  HHNINC|    1.19343          .87672      1.36  .1734     -.52492   2.91177
  HHKIDS|     .31389          .34815       .90  .3673     -.36847    .99625
        |Means for random parameters
Constant|    -.97734         1.00299      -.97  .3298    -2.94317    .98849
Constant|     .23872          .96599       .25  .8048    -1.65459   2.13202
Constant|    2.06626**        .88897      2.32  .0201      .32392   3.80860
Constant|    1.56019*         .90344      1.73  .0842     -.21052   3.33089
        |Scale parameters for dists. of random parameters
Constant|     .02031          .19069       .11  .9152     -.35343    .39406
Constant|    1.22214***       .17722      6.90  .0000      .87480   1.56948
Constant|    1.73095***       .17833      9.71  .0000     1.38142   2.08048
Constant|    2.55108***       .18704     13.64  .0000     2.18448   2.91768
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This model has the same latent effect in each utility function, though different scale factors.

```
-----------------------------------------------------------------------------
Random Coefficients  MltLogit Model
Dependent variable                HSAT
Log likelihood function     -1258.50063
Estimation based on N =    905, K =  20
Inf.Cr.AIC  = 2557.001 AIC/N =    2.825
Unbalanced panel has    394 individuals
Multinomial logit with random effects
Simulation based on  50 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|    -.00209         .02263     -.09  .9264     -.04644    .04226
  HHNINC|     .48018        1.17852      .41  .6837    -1.82968   2.79003
  HHKIDS|     .29347         .43402      .68  .4989     -.55720   1.14414
     AGE|     .01538         .01558      .99  .3234     -.01515    .04591
  HHNINC|    1.34339*        .70838     1.90  .0579     -.04501   2.73178
  HHKIDS|     .21473         .32248      .67  .5055     -.41733    .84679
     AGE|    -.00776         .01237     -.63  .5304     -.03201    .01649
  HHNINC|    1.19572*        .65055     1.84  .0661     -.07933   2.47077
  HHKIDS|    -.05011         .29433     -.17  .8648     -.62699    .52676
     AGE|     .00310         .01324      .23  .8149     -.02286    .02906
  HHNINC|    1.44279**       .70145     2.06  .0397      .06796   2.81761
  HHKIDS|     .31137         .29645     1.05  .2936     -.26967    .89241
        |Means for random parameters
Constant|   -1.47532        1.20016    -1.23  .2190    -3.82759    .87696
Constant|    -.70734         .82080     -.86  .3888    -2.31608    .90140
Constant|    1.09794*        .62345     1.76  .0782     -.12401   2.31988
Constant|     .64952         .67371      .96  .3350     -.67094   1.96998
        |Scale parameters for dists. of random parameters
Constant|    1.38963***      .18611     7.47  .0000     1.02486   1.75439
Constant|     .40740***      .09464     4.30  .0000      .22192    .59289
Constant|     .26460***      .07701     3.44  .0006      .11367    .41553
Constant|    1.27599***      .10406    12.26  .0000     1.07203   1.47995
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This model has separate, correlated effects in all utility functions.

```
-----------------------------------------------------------------------------
Random Coefficients  MltLogit Model
Dependent variable                HSAT
Log likelihood function     -1228.68780
Estimation based on N =    905, K =  26
Inf.Cr.AIC  = 2509.376 AIC/N =    2.773
Unbalanced panel has    394 individuals
Multinomial logit with random effects
Simulation based on  50 Halton draws
-----------------------------------------------------------------------------
```

HSAT	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
AGE	-.00277	.01900	-.15	.8840	-.04001	.03447
HHNINC	.18258	1.05908	.17	.8631	-1.89318	2.25833
HHKIDS	.44728	.39924	1.12	.2626	-.33522	1.22978
AGE	.01952	.01979	.99	.3239	-.01927	.05832
HHNINC	.99148	.88908	1.12	.2648	-.75109	2.73405
HHKIDS	.19586	.36220	.54	.5887	-.51404	.90577
AGE	-.00134	.01802	-.07	.9407	-.03667	.03398
HHNINC	.74182	.88342	.84	.4011	-.98965	2.47329
HHKIDS	-.06698	.35619	-.19	.8508	-.76510	.63114
AGE	.00795	.01824	.44	.6631	-.02780	.04369
HHNINC	.95944	.89476	1.07	.2836	-.79425	2.71313
HHKIDS	.26625	.34917	.76	.4457	-.41811	.95061
	Means for random parameters					
Constant	-1.44262	.98772	-1.46	.1441	-3.37851	.49327
Constant	.03520	1.05196	.03	.9733	-2.02660	2.09700
Constant	2.00734**	.94721	2.12	.0341	.15083	3.86384
Constant	1.54147	.94470	1.63	.1027	-.31011	3.39305
	Diagonal elements of Cholesky matrix					
Constant	.77973***	.21166	3.68	.0002	.36489	1.19458
Constant	1.02801***	.14489	7.10	.0000	.74403	1.31199
Constant	.22445**	.09346	2.40	.0163	.04127	.40763
Constant	.18188**	.08031	2.26	.0235	.02447	.33929
	Below diagonal elements of Cholesky matrix					
lONE_ONE	.50481***	.18120	2.79	.0053	.14966	.85995
lONE_ONE	1.08605***	.17694	6.14	.0000	.73926	1.43284
lONE_ONE	.94188***	.13768	6.84	.0000	.67204	1.21172
lONE_ONE	1.88987***	.18720	10.10	.0000	1.52296	2.25677
lONE_ONE	1.07104***	.14041	7.63	.0000	.79584	1.34624
lONE_ONE	.37947***	.09765	3.89	.0001	.18807	.57086

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Implied covariance matrix of random parameters

Var_Beta	1	2	3	4
1	.607984	.393614	.846831	1.47359
2	.393614	1.31163	1.51651	2.05506
3	.846831	1.51651	2.11703	3.14646
4	1.47359	2.05506	3.14646	4.89580

Implied standard deviations of random parameters

S.D_Beta	1
1	.779734
2	1.14527
3	1.45500
4	2.21265

Implied correlation matrix of random parameters

Cor_Beta	1	2	3	4
1	1.00000	.440776	.746426	.854121
2	.440776	1.00000	.910072	.810972
3	.746426	.910072	1.00000	.977343
4	.854121	.810972	.977343	1.00000

E37.10.2 A Dynamic Multinomial Logit Model

The preceding random effects model can be modified to produce the dynamic multinomial logit model analyzed in Gong, van Soest and Villagomez (2000). Then

$$P_{ijt} \mid \alpha_{i1},\ldots,\alpha_{iJ} = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \alpha_{ij})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}_j'\mathbf{x}_{it} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \alpha_{ij})} \quad t = 1,\ldots,T_i,\ j = 0,1,\ldots,J,\ i = 1,\ldots,N$$

where $\mathbf{z}_{it}$ contains lagged values of the dependent variables (these are binary choice indicators for the choice made in period t) and possibly interactions with other variables. The $\mathbf{z}_{it}$ variables are now endogenous, and conventional maximum likelihood estimation is inconsistent. The authors argue that Heckman's treatment of initial conditions is sufficient to produce a consistent estimator. (We used this method to set up a dynamic probit model in Section E31.2.6.) The core of the treatment is to treat the first period as an equilibrium, with no lagged effects,

$$P_{ij0} \mid \theta_{i1},\ldots,\theta_{iJ} = \frac{\exp(\boldsymbol{\delta}_j'\mathbf{x}_{i0} + \theta_{ij})}{\sum_{j=1}^{J}\exp(\boldsymbol{\delta}_j'\mathbf{x}_{i0} + \theta_{ij})},\ t = 0,\ j = 0,1,\ldots,J,\ i = 1,\ldots,N$$

where the vector of effects, θ, is built from the same primitives as α in the later choice probabilities. Thus, $\boldsymbol{\alpha}_i = \boldsymbol{\Gamma}\mathbf{v}_i$ and $\boldsymbol{\theta} = \boldsymbol{\Phi}\,\mathbf{v}_i$, for the same $\mathbf{v}_i$, but different lower triangular scaling matrices. This treatment slightly less than doubles the size of the model – it amounts to a separate treatment for the first period.) Full information maximum likelihood estimates of the model parameters, $(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_J,\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_J,\boldsymbol{\Gamma},\Phi)$ are obtained by maximum simulated likelihood, by modifying the random effects model. The likelihood function for individual i consists of the period 0 probability as shown above times the product of the period $1,2,\ldots,T_i$ probabilities defined earlier.

In order to use this procedure, you must create the lagged values of the variables, and the products with other variables if any are to be present – that is, the elements of $\mathbf{z}_{it}$. Then, starting values for both parameter vectors must be provided for the iterations. The program below shows the several steps involved. In terms of the broad command structure, the essential new ingredient will be the addition of

; Rh2 = the variables in z

to the model definition. However, again, several steps must precede this, as shown in the command set below.

To construct this estimator in generic form, we assume the dependent variable is named *y* and the independent variables are to be contained in a namelist *x*. Several commands remain application specific. These are modified for the specific model. We need a time variable first. For convenience, periods are numbered 1,...,*T* with *t* = 1 being the initial period.

```
    NAMELIST   ; x = the x variables in the model, including one $
    SAMPLE     ; All $
    CREATE     ; time = Trn(-T,0)  $    Fixed number of periods
or  CREATE     ; time = Ndx(ID,1) $     Unbalanced panel, variable T(i)
```

Compute the binary variables for the outcomes - endogenous variables.

```
CREATE      ; dit1 = (y=1) ; dit2 = (y=2) ; dit3 = (y=3) ... and so on ... $
```

Create lagged values of the dummy variables and interactions of lagged dummy variables with other variables in the model if desired. You will name variables according to your application. This is just a template. (And repeat likewise for a second, third, ... *x* variable.)

```
CREATE      ; dit1lag = dit1[-1] ; dit2lag = dit2[-1]
            ; dit3lag = dit3[-1] ... and so on $
CREATE      ; d1x1lag = dit1lag*x1 ; d2x1lag = dit2lag*x1 ... $
NAMELIST    ; z = dit1L,dit2L,...,d1x1L,...,... for the z variables $
```

Fit the time invariant model for the first period and retain the coefficients.

```
REJECT      ; time > 1 $
MLOGIT      ; Lhs = y ; Rhs = x $
MATRIX      ; delta = b $
```

Fit the dynamic part for 2,...,T_i and again, save the coefficients.

```
INCLUDE     ; New ; T > 1 $
MLOGIT      ; Lhs = y ; Rhs = x,z $
MATRIX      ; betagama = b $
```

The full model for all periods is a random parameters model.

```
SAMPLE      ; All $
MLOGIT      ; Lhs = y ; Rhs = x
            ; Rh2 = z ?  This indicates the dynamic MNL model.
            ; Start = delta,betagama
            ; RPM ; (options including ; Halton, ; Pts = replications)
            ; Panel specification
            ; Fcn = one(n) ; Common $ (; Correlated may be specified)
```

E38: Conditional Logit Models

E38.1 Introduction

This chapter and Chapters E39 and E40 will describe the major extension of the MLOGIT model of Chapter E37. An individual is assumed to have preferences defined over a set of alternatives (travel modes, occupations, food groups, etc.)

$$U(\text{alternative } 1) = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \boldsymbol{\gamma}_1'\mathbf{z}_i + \varepsilon_{i1}$$

...

$$U(\text{alternative } J) = \boldsymbol{\beta}_J'\mathbf{x}_{iJ} + \boldsymbol{\gamma}_J'\mathbf{z}_i + \varepsilon_{iJ}$$

$$\text{Observed } Y_i = \text{choice } j \text{ if } U_i(\text{alternative } j) > U_i(\text{alternative } k)\ \forall\ k \neq j.$$

In this expanded specification, we use $\mathbf{x}_{ij}$ to denote the *attributes* of choice j that face individual i – attributes generally differ across choices and across individuals. We use $\mathbf{z}_i$ to denote *characteristics* of individual i, such as age, income, gender, etc. Characteristics differ across individuals, but not across choices. The 'disturbances' in this framework (individual heterogeneity terms) are assumed to be independently and identically distributed with identical extreme value distribution; the CDF is

$$F(\varepsilon_j) = \exp(-\exp(-\varepsilon_j))$$

Based on this specification, the choice probabilities,

$$\text{Prob[choice } j \text{]} = \text{Prob}[U_j > U_k],\ \forall\ k \neq j$$

$$= \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ji} + \boldsymbol{\gamma}_j'\mathbf{z}_i)}{\sum_{m=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{x}_{mi} + \boldsymbol{\gamma}_m'\mathbf{z}_i)},\ j = 1,...,J,$$

where 'i' indexes the observation, or individual, and 'j' and 'm' index the choices. We note at the outset, the IID assumptions made about ε_j are quite stringent, and lead to the 'Independence from Irrelevant Alternatives' or IIA implications that characterize the model. Much (perhaps all) the research on forms of this model consists of development of alternative functional forms and stochastic specifications that avoid this feature. We return to that aspect in Section E40.4, and leave it unresolved for the present.

The observed data consist of the vectors, $\mathbf{x}_{jt}$ and $\mathbf{z}_i$ and the outcome, or choice, y_i. (We also consider a number of variants.) A well known example is travel mode choice. Samples of observations often consist of the attributes of the different modes and the choice actually made. Usually, no characteristics of the individuals are observed beyond their actual choice, though survey data may include familiar sociodemographics such as age and income. Models may also contain mixtures of the two types of choice determinants. Chapters E38-E40 present the various aspects of this model contained in *LIMDEP*. This chapter describes basic model specification and estimation. Chapter E39 describes extensions of the model that allow for different types of data, different specifications of the utility functions and a built in feature of the estimation for modeling choice strategy of the individual. Chapter E40 develops the post estimation features, partial effects, prediction and model simulation.

Notes on the Conditional Logit Model, MLOGIT, CLOGIT and NLOGIT

For the present, we have labeled the model estimated by the program described in this chapter as the 'conditional logit model.' It will be clear shortly that this is a meaningless distinction. The only significance to the use of CLOGIT (conditional logit) here and MLOGIT (multinomial logit) in the preceding chapter is to differentiate the commands used in *LIMDEP*. The models are, in fact, the same. We will demonstrate this with an example below. Indeed, in the contemporary literature the model we are examining here above is generically called the 'multinomial logit model,' and the artificial distinction we have drawn based on characteristics vs. attributes has largely faded from view.

The internal programs that do the estimation for MLOGIT and CLOGIT are different, however. MLOGIT is a specific estimation module in *LIMDEP*. CLOGIT is likewise a particular estimation module, but it is also the gateway to *NLOGIT*, a separate package of analysis tools for analysis of discrete choice models. The models estimated by *NLOGIT* Version 5 are extensions of the basic multinomial logit fit by CLOGIT. These include:

- nested logit,
- generalized nested logit,
- nested logit with covariance heterogeneity,
- multinomial probit,
- heteroscedastic extreme value,
- mixed (random parameters) logit,
- latent class logit,
- error components logit,
- generalized mixed logit,
- scaled mixed logit,
- random regret logit model,
- nonlinear random parameters logit,
- Box-Cox nested logit,
- latent class mixed logit,

and a few others. These models are not in *LIMDEP*. (Development of *NLOGIT* began in the late 1990s with the construction of full information maximum likelihood estimators for the nested logit model (hence the name, '*NLOGIT*'). The package has evolved into a large group of estimators for the models listed above, as well as a separate set of tools for estimation and analysis of discrete choice models. *NLOGIT* Version 5 consists of all of *LIMDEP* as described in these manuals plus the analysis tools for discrete choice. Further information about *NLOGIT* and its features may be found on the website for the program, www.nlogit.com.

E38.2 The Conditional Logit Model – CLOGIT

In the multinomial logit model described in Chapter E37, there is a single vector of characteristics that describes the individual, and a set of J parameter vectors. In the 'discrete choice' setting of this chapter, these are essentially reversed. The J (not J+1 – we will be changing the notation slightly here) alternatives are each characterized by a set of K 'attributes,' $\mathbf{x}_{ij}$. Respondent 'i' chooses among the J alternatives. In the example we will use throughout this discussion, a sampled individual making a trip between Sydney and Melbourne chooses one of four modes of travel, air, train, bus or car. The attributes include cost, travel time and terminal time, which differ by mode, and characterize the choice, not the person. The data also include a characteristic of the chooser, household income. It will emerge shortly however, that MLOGIT and CLOGIT are not different models at all. The estimator described here accommodates both cases, and mixtures of the two. For example, for the commuting application just noted, we also have income for the person and traveling party size, both of which are choice invariant.

For the present, we develop the model with a single parameter vector, $\boldsymbol{\beta}$. The model underlying the observed data is assumed to be the following random utility specification:

$$U(\text{choice } j \text{ for individual } i) = U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{z}_i + \varepsilon_{ij},\ j = 1,...,J.$$

The random, individual specific terms, $(\varepsilon_{i1},\varepsilon_{i2},...,\varepsilon_{iJ})$ are once again assumed to be independently distributed across the utilities, each with the same type 1 extreme value distribution

$$F(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij})).$$

Under these assumptions, the probability that individual i chooses alternative j is

$$\text{Prob}[U_{ij} > U_{im}] \text{ for all m} \neq \text{j}.$$

It has been shown that for independent extreme value (Gumbel) distributions, as above, this probability is

$$\text{Prob}[y_i = j] = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i\right)}{\sum_{m=1}^{J}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{im} + \boldsymbol{\gamma}_m'\mathbf{z}_i\right)}$$

where y_i is the index of the choice made. As before, we note at the outset that the IID assumptions made about ε_j are quite stringent, and induce the 'Independence from Irrelevant Alternatives' or IIA features that characterize the model. We will return to this restriction later in Chapter E40. Regardless of the number of choices, there is a single vector of K parameters to be estimated. This model does not suffer from the proliferation of parameters that appears in the MLOGIT model described in Section E37.2.

For convenience in what follows, we will refer to the estimator as CLOGIT, keeping in mind, this refers to a command and class of models in *LIMDEP*, not a separate program.

The basic setup for this model consists of observations on n individuals, each of whom makes a single choice among J_i choices, or alternatives. There is a subscript on J_i because ultimately, we will not restrict the choice sets to have the same number of choices for every individual. The data will typically consist of the choices and observations on K 'attributes' for each choice. The attributes that describe each choice, i.e., the variables that enter the utility functions, may be the same for all choices, or may be defined differently for each utility function. The estimator described in this chapter allows a large number of variations of this basic model. In the discrete choice framework, the observed 'dependent variable' usually consists of an indicator of which among J_i alternatives was *most* preferred by the respondent. All that is known about the others is that they were judged inferior to the one chosen. But, there are cases in which information is more complete and consists of a subjective ranking of all J_i alternatives by the individual. CLOGIT allows specification of the model for estimation with 'ranks data.' In addition, in some settings, the sample data might consist of aggregates for the choices, such as proportions (market shares) or frequency counts. CLOGIT will accommodate these cases as well.

E38.3 Clogit Data for the Applications

The documentation of the CLOGIT program below includes numerous applications based on the data set clogit.dat, that is distributed with *LIMDEP*. These data provide a compact illustration of how data should be arranged for CLOGIT. The data set is a survey of the transport mode chosen by a sample of 210 travelers between Sydney and Melbourne (about 500 miles) and other points in nonmetropolitan New South Wales. As will be shown, the clogit data will generally consist of a record (row of data) for each alternative in the choice set, for each individual. Thus, the data file contains 210 observations, or 840 records. The variables in the data set are as follows:

Original Data

mode = 0/1 for four alternatives: air, train, bus, car
(this variable equals one for the choice made, labeled *choice* below),
ttme = terminal waiting time,
invc = invehicle cost for all stages,
invt = invehicle time for all stages,
gc = generalized cost measure = Invc + Invt × value of time,
chair = dummy variable for chosen mode is air,
hinc = household income in thousands,
psize = traveling party size.

Transformed Variables

aasc = choice specific dummy for air (generated internally),
tasc = choice specific dummy for train,
basc = choice specific dummy for bus,
casc = choice specific dummy for car,
hinca = *hinc* × *aasc*,
psizea = *psize* × *aasc*.

The table below lists the first 10 observations in the data set. In the terms used here, each 'observation' is a block of four rows. The mode chosen in each block is boldfaced.

mode	*choice*	*ttme*	*invc*	*invt*	*gc*	*chair*	*hinc*	*psize*	*aasc*	*tasc*	*basc*	*casc*	*hinca*	*psizea*	*obs.*
Air	0	69	59	100	70	0	35	1	1	0	0	0	35	1	*i*=1
Train	0	34	31	372	71	0	35	1	0	1	0	0	0	0	
Bus	0	35	25	417	70	0	35	1	0	0	1	0	0	0	
Car	**1**	**0**	**10**	**180**	**30**	**0**	**35**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	64	58	68	68	0	30	2	1	0	0	0	30	2	*i*=2
Train	0	44	31	354	84	0	30	2	0	1	0	0	0	0	
Bus	0	53	25	399	85	0	30	2	0	0	1	0	0	0	
Car	**1**	**0**	**11**	**255**	**50**	**0**	**30**	**2**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	69	115	125	129	0	40	1	1	0	0	0	40	1	*i*=3
Train	0	34	98	892	195	0	40	1	0	1	0	0	0	0	
Bus	0	35	53	882	149	0	40	1	0	0	1	0	0	0	
Car	**1**	**0**	**23**	**720**	**101**	**0**	**40**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	64	49	68	59	0	70	3	1	0	0	0	70	3	*i*=4
Train	0	44	26	354	79	0	70	3	0	1	0	0	0	0	
Bus	0	53	21	399	81	0	70	3	0	0	1	0	0	0	
Car	**1**	**0**	**5**	**180**	**32**	**0**	**70**	**3**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	64	60	144	82	0	45	2	1	0	0	0	45	2	*i*=5
Train	0	44	32	404	93	0	45	2	0	1	0	0	0	0	
Bus	0	53	26	449	94	0	45	2	0	0	1	0	0	0	
Car	**1**	**0**	**8**	**600**	**99**	**0**	**45**	**2**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	69	59	100	70	0	20	1	1	0	0	0	20	1	*i*=6
Train	**1**	**40**	**20**	**345**	**57**	**0**	**20**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	
Bus	0	35	13	417	58	0	20	1	0	0	1	0	0	0	
Car	0	0	12	284	43	0	20	1	0	0	0	1	0	0	
Air	**1**	**45**	**148**	**115**	**160**	**1**	**45**	**1**	**1**	**0**	**0**	**0**	**45**	**1**	*i*=7
Train	0	34	111	945	213	1	45	1	0	1	0	0	0	0	
Bus	0	35	66	935	167	1	45	1	0	0	1	0	0	0	
Car	0	0	36	821	125	1	45	1	0	0	0	1	0	0	
Air	0	69	121	152	137	0	12	1	1	0	0	0	12	1	*i*=8
Train	0	34	52	889	149	0	12	1	0	1	0	0	0	0	
Bus	0	35	50	879	146	0	12	1	0	0	1	0	0	0	
Car	**1**	**0**	**50**	**780**	**135**	**0**	**12**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	69	59	100	70	0	40	1	1	0	0	0	40	1	*i*=9
Train	0	34	31	372	71	0	40	1	0	1	0	0	0	0	
Bus	0	35	25	417	70	0	40	1	0	0	1	0	0	0	
Car	**1**	**0**	**17**	**210**	**40**	**0**	**40**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	
Air	0	69	58	68	65	0	70	2	1	0	0	0	70	2	*i*=10
Train	0	34	31	357	69	0	70	2	0	1	0	0	0	0	
Bus	0	35	25	402	68	0	70	2	0	0	1	0	0	0	
Car	**1**	**0**	**7**	**210**	**30**	**0**	**70**	**2**	**0**	**0**	**0**	**1**	**0**	**0**	

E38.3.1 Setting Up the Data

The clogit data are arranged as follows, where we use a specific set of values for the problem to illustrate. Suppose you observe 25 individuals. Each individual in the sample faces three choices and there are two attributes, *q* and *w*. For each observation, we also observe which choice was made. Suppose further that in the first three observations, the choices made were two, three, and one, respectively. The data matrix would consist of 75 rows, with 25 blocks of three rows. Within each block, there would be the set of attributes and a variable *y*, which, at each row, takes the value one if the alternative is chosen and zero if not. Thus, within each block of *J* rows, *y* will be one once and only once. For the hypothetical case, then, we have:

	y	*q*	*w*
i=1	0	$q_{1,1}$	$w_{1,1}$
	—>1	$q_{2,1}$	$w_{2,1}$
	0	$q_{3,1}$	$w_{3,1}$
i=2	0	$q_{1,2}$	$w_{1,2}$
	0	$q_{2,2}$	$w_{2,2}$
	—>1	$q_{3,2}$	$w_{3,2}$
i=3	—>1	$q_{1,3}$	$w_{1,3}$
	0	$q_{2,3}$	$w_{2,3}$
	0	$q_{3,3}$	$w_{3,3}$

and so on, continuing to *i* = 25, where '—>'marks the row of the respondent's actual choice. The clogit.dat data set shown earlier illustrates the general construction of the data set. Note that for purposes of CLOGIT, the data are set up in the same fashion as a panel data set in other settings.

When you **READ** the data for this model, the data set is not treated any differently. *Nobs* would be the total number of rows in the data set, in the hypothetical case, 75, not 25, and 840 for clogit.dat. The separation of the data set into the above groupings would be done at the time this particular model is estimated.

NOTE: Missing values are handled automatically by this estimator. Do not reset the sample or use **SKIP** with **CLOGIT**. Observations which have missing values are bypassed as a group. We note an implication of this: the multiple imputation programs in *LIMDEP* cannot be used to fill missing values in a multinomial choice setting.

Thus far, it is assumed that the observed outcome is an indicator of which choice was made among a fixed set of up to 500 choices. There are numerous possible variations:

- Data on the observed outcome may be in the form of frequencies, market shares or ranks.
- The number of choices may differ across observations.
- The choice set may be extremely large. A method of fitting models with up very large choice sets is discussed below

E38.3.2 Checking Data Validity

CLOGIT does a full check of the data for bad observations (usually coding errors or missing values) before estimation is done. The program output will contain a simple count of the number of invalid observations that have been bypassed. For example, we sprinkled some missing values into the clogit.dat data set, and fit a model. The initial output contains the count:

```
+-------------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found   3 bad observations among     210 individuals. |
|You can use ;CheckData to get a list of these points. |
+-------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -181.67965
Estimation based on N =    207, K =   7
Inf.Cr.AIC  =  377.359 AIC/N =    1.823
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -279.9949  .3511 .3437
Chi-squared[ 4]          =    196.63055
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    3 obs
--------+--------------------------------------------------------------------
```

You may request the program to show you exactly where the problem observations are by adding

; Check Data

to the command. A complete listing of the bad observations is produced – note in a large data set, this could be quite long. For the preceding, we obtained

```
+-----------------------------------------------------------+
| Inspecting the data set before estimation.               |
| These errors mark observations which will be skipped.    |
| Row Individual = 1st row then group number of data block |
+-----------------------------------------------------------+
    1      1  Individual data, LHS variable is not 0 or 1
    9      3  Missing value found for characteristic or attribute in utility
   17      5  Missing value found for LHS variable
```

E38.3.3 Types of Data on the Choice Variable

Data on the dependent (Lhs) variable may come in four forms:

- **Individual Data:** The Lhs variable consists of zeros and a single one which indicates the choice that the individual made. The data sets shown earlier are individual data.
- **Proportions Data:** The Lhs variable consists of a set of sample proportions or market shares. Values range from zero to one, and they sum to 1.0 over the set of choices in the choice set. Observed proportions may equal 1.0 or 0.0 for some observations.
- **Frequency Data:** The Lhs variable consists of a set of frequency counts for the outcomes. Frequencies are nonnegative integers for the outcomes in the choice set and may be zero.
- **Ranks Data:** The Lhs variable consists of a complete set of ranks of the alternatives in the individual's choice set. Thus, if there are *J* alternatives available, the observation will consist of a full set of the integers 1,...,*J* not necessarily in that order, which indicate the individual's ranking of the alternatives. The number of choices may still differ by observation. Thus, we might have [(unranked),0,1,0,0,0] in the usual case, and [(ranked) 4,1,3,2,5] with ranks data. Note that the positions of the ones are the same for both sets, by definition. (See Beggs, Cardell, and Hausman (1981).) You may also have partial rankings. For example, suppose respondents are given 10 choices and asked to rank their top three. Then, the remaining six choices should be coded 4.0. A set of ranks might appear thusly: [1,4,2,4,3,4,4,4,4,4]. The ties must only appear at the lowest level. Ties in the data are detected automatically. No indication is needed. For later reference, we note the following for the model based on ranks data:
  - You may have observation weights, but no choice based sampling.
  - The IIA test described in Chapter E40 is not available.

The first three data types can be detected automatically by **CLOGIT**. You generally do not have to give any additional information about the data set, since the type of data being provided can usually be deduced from the values. The ranks data are an exception for which you must use

CLOGIT ; ... as before ...; Ranks $

If you are using frequency or proportions data, and your data contain zeros or ones, then certain kinds of observations cannot be distinguished from erroneous individual data, and they may be flagged as such. For example, in a frequency data set, the observation [0,0,1,1,0,0] is a valid observation, but for individual data, it looks like a badly coded observation. In order to avoid this kind of ambiguity, if you have frequency data containing zeros, add

; Frequencies

to your **CLOGIT** command. (You may use this in any event to be sure that the data are always recognized correctly.) If you have proportions data, instead, you may use

; Shares

to be sure that the data are correctly marked. (Again, this will only be relevant if your data contain zeros and/or ones.)

Data are checked for validity and consistency. An unrecognizable mixture of the three types will cause an error. For example, a mixture of frequency and proportions data cannot be properly analyzed. For the ranks data, an error will occur if the set of ranks is miscoded or incomplete or if ties are detected for any ranks other than the lowest.

E38.3.4 Simulated Choice Data

For some kinds of experiments and simulations, you might want to draw a random sample of choices given known utility functions. CLOGIT allows simulation of the Lhs variable in a choice model using

$$Y = j^* \text{ from } \text{Max}(U_{ij})$$

where $U_{ij} = v_{ij}$ + a simulated random term. You must provide the utility values as the Lhs variable. The choice outcome is then simulated by adding a type 1 extreme value error term to each utility value, and choosing the j associated with the largest simulated utility. Request this computation by adding

; MCS (for Monte Carlo Simulation)

to the **CLOGIT** command. (The utilities are not lost. You can reuse them, for example to do another simulation. On the other hand, the simulated data are lost at the end of the estimation.) Keep in mind, if you want to reuse the data for a simulation, you have to reset the seed for the random number generator. You might for example want to fit different models with the same simulated data set.

E38.3.5 Entering Data on a Single Line

The clogit data are generally provided as if in a panel data set, in blocks of J_i observations per individual, where J_i is the number of choices. The following describes an alternative format in which data for these models are provided in one line per individual. This construction can only be used for discrete choice models with a fixed number of alternatives available to each individual. This feature is not available for cases in which the choice set varies across individuals. (We have seen this arrangement of data called the ‘wide form,’ with the data arranged as earlier in the ‘long form.’)

In general, discrete choice models require that the data set be arranged with a line of data (observation) for each alternative in the model, essentially as a panel. For purposes of the discussion, it will be useful to consider an example. Suppose individuals choose among four alternatives, (*air,train,bus,car*), and the attributes are *cost* and *traveltime*, which vary across choice, and *income* which is fixed. The actual data for an observation would consist of four variables on four records, arranged as follows: (The y_j variable consists of three zeros and a one to indicate the choice made.)

The arrangement is as follows:

$$\begin{array}{cc} & \begin{array}{cccc} \text{Choice} & \text{Cost} & \text{Time} & \text{Income} \end{array} \\ \begin{array}{c} \text{Air} \\ \text{Train} \\ \text{Bus} \\ \text{Car} \end{array} & \begin{bmatrix} y_{air} & cost_{air} & time_{air} & income \\ y_{train} & cost_{train} & time_{train} & income \\ y_{bus} & cost_{bus} & time_{bus} & income \\ y_{car} & cost_{car} & time_{car} & income \end{bmatrix} \end{array}$$

The model observation would be constructed from the four variables, and would, with alternative specific constants for the first three alternatives, ultimately appear as follows:

$$\begin{array}{cc} & \begin{array}{ccccc} \text{choice} & \text{cost} & \text{time} & \text{constants} & \text{income} \end{array} \\ \mathbf{X}_i = & \begin{bmatrix} y_{air} & c_a & t_a & 1 & 0 & 0 & income & 0 & 0 \\ y_{train} & c_t & t_t & 0 & 1 & 0 & 0 & income & 0 \\ y_{bus} & c_b & t_b & 0 & 0 & 1 & 0 & 0 & income \\ y_{car} & c_c & t_c & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

This setup normally requires four lines of data. But, an alternative way to arrange the same data would be in a single line of data, consisting of

Choice(coded 0,1,2,3) ca ct cb cc ta tt tb tc one income

from which it would be straightforward to construct the observation above.

The command for this arrangement will contain the following to set this up: First, the choice set is specified as follows:

; Lhs = the name of the choice variable (here, *choice***)**
; Choices = the list of J choice labels [coding of Lhs variable]

The coding is contained in square brackets. If the dependent variable is coded as consecutive integers, such as 0,1,2,3, then just put the first value in the brackets. Thus, 0,1,2,3 is indicated with [0], while 1,2,3,4 is [1]. For our example, this is going to appear

; Lhs = choice
; Choices = air,train,bus,car [0]

If the coding is some other set of integers, put the set of integers in the square brackets. Suppose, for example, in our model, we eliminated *train* as a choice. Then, the coding might be [0,2,3].

NOTE: It is only the square brackets in the **; Choices** specification which indicates that you will be using this data arrangement instead of the standard one.

Second, for variables which provide attributes which vary by choice, such as *cost* and *time* above, a **; Rhs** specification must contain blocks of *J* variable names. For the example, this might be

; Rhs = cair,ctrain,cbus,ccar,tair,ttrain,tbus,tcar

For variables which are to be interacted with alternative specific constants, as well as the constants themselves, use **; Rh2** instead of **; Rhs**. Thus, for the example above, we might use

; Rh2 = one,income

NOTE: To request a set of alternative specific constants, include *one* in the Rh2 list, not the Rhs list.

Notice that when these interactions are created, the last one in the set is dropped. In the example above, only three constants and three income terms appear in the four choice model.

Third, for the Rhs groups, a name is created for the group, *attrib*01, *attrib*02, and so on. If you would like to provide your own names for the blocks, use

; Attr = list of k labels

To combine all of these in our example, we might use

```
; Lhs  = mode
; Choices = air,train,bus,car [ 0 ]
; Rhs  = cair,ctrain,cbus,ccar,tair,ttrain,tbus,tcar
; Rh2 = one,income
; Attr = cost,time
```

E38.3.6 Converting Wide Data Sets to the Long Format

The single line format for multinomial choice modeling is clumsy, and will become extremely unwieldy if the choice set has more than a few alternatives or the model has more than two or three attributes. A utility program is provided for you to convert single line choice data to the more convenient format.

We wish to transform the data set so that one observation in the second form shown above becomes three observations in the first form above. The general command is

```
CLCONVERT ; Lhs     = one or more choice variables
          ; Choices = the J names for the choices in the choice set
          ; Rhs     = K sets of J variable names – the attributes
          ; Rh2     = M characteristics variables
          ; Names   = names for new choice variables,
                      names for new attribute variables,
                      names for new characteristic variables $
```

For the example above, the command would be

CLCONVERT ; Lhs = choicei
; Choices = car,train,bus
; Rhs = ctime,ttime,btime,ccost,tcost,bcost
; Rh2 = agei,incomei
; Names = choice,time,cost,age,income $

This command is set up to resemble a model command to make it simple to construct. But, it does nothing but rearrange the data set.

Some points to note about **CLCONVERT** are:

- It is only for choice settings with fixed numbers of choices for every observation.
- You can recode more than one choice variable with the other data.
- You can rearrange the entire data set, not just the variables for a particular model. The appearance of the command as a model command is only for convenience.
- After the data are converted, the new data are placed at the top of the data array, regardless of where they were before. You can, for example, convert rows 201 to 250 in your data set. If this is a three choice setting, the new data will be observations 1 to 150.

There are also several conventions that must be followed:

- The new names must not be in use for anything else already in your project, including other variables. **CLCONVERT** cannot replace existing variables.
- You must provide the **; Names** and **; Choices** specifications. These are mandatory.
- You must provide at least one of **; Rhs** or **; Rh2** variable. Either is optional, but at least one of the two must be present.
- Note that the count of Rhs variables is an exact multiple of the number of choices in the **; Choices** list.
- The number of names in the **; Names** list is the sum of
    - the number of Lhs variables
    - the number of sets of Rhs variables
    - the number of Rh2 variables.

When **CLCONVERT** is executed, the sample is reset to the number of observations in the new sample. There is an additional option with **CLCONVERT**. After the data are converted, you can discard the original data set with

; Clear

This leaves the entire data set consisting of the variables that are in your **; Names** list. (Use this with caution. The operation cannot be reversed.)

To illustrate the operation of this command, suppose the data set consists of these three observations:

$$\begin{bmatrix} choicei1 & choicei2 & ctime & ttime & btime & ccost & tcost & bcost & agei & incomei \\ 2 & 3 & 44 & 29 & 56 & 125 & 40 & 25 & 37 & 56.6 \\ 1 & 1 & 19 & 44 & 20 & 160 & 18 & 50 & 42 & 98.6 \\ 3 & 2 & 28 & 55 & 15 & 85 & 50 & 9 & 10 & 22.0 \end{bmatrix}.$$

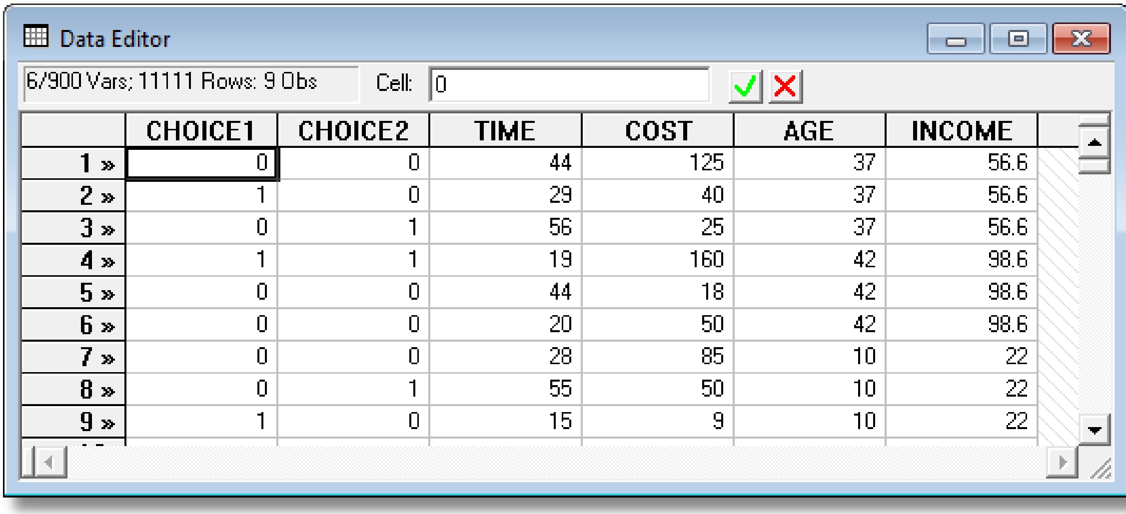
We wish to convert this data set to *NLOGIT*'s multiple line format. There are three choices in the choice set, so there will be three rows of data for each observation. The command and the results are as follows:

IMPORT $

```
choicei1,choicei2,ctime,ttime,btime,ccost,tcost,bcost,agei,incomei
2,3,44,29,56,125,40,25,37,56.6
1,1,19,44,20,160,18,50,42,98.6
3,2,28,55,15, 85,50, 9,10,22.0
```

ENDDATA $
CLCONVERT ; Lhs = choicei1,choicei2
; Choices = car,train,bus
; Rhs = ctime,ttime,btime,ccost,tcost,bcost
; Rh2 = agei,incomei
; Names = Choice1,Choice2,time,cost,age,income ; Clear $

```
=================================================================
Data Conversion from One Line Format for NLOGIT
Original data were cleared. This is now the whole data set.
The new sample contains      9 observations.
=================================================================
Choice set in new data set has  3 choices:
CAR      TRAIN    BUS
-----------------------------------------------------------------
There were  2 choice variables coded 1,..., 3 converted to binary
Old variable = CHOICEI1, New variable = CHOICE1
Old variable = CHOICEI2, New variable = CHOICE2
-----------------------------------------------------------------
There were  2 sets of variables on attributes converted.  Each
set of  3 variables is converted to one new variable
New Attribute variable TIME     is constructed from
CTIME    TTIME    BTIME
New Attribute variable COST     is constructed from
CCOST    TCOST    BCOST
-----------------------------------------------------------------
There were  2 characteristics that are the same for all choices.
Old variable = AGEI    , New variable = AGE
Old variable = INCOMEI , New variable = INCOME
=================================================================
```



Data Editor

6/900 Vars; 11111 Rows: 9 Obs   Cell: 0

	CHOICE1	CHOICE2	TIME	COST	AGE	INCOME
1 »	0	0	44	125	37	56.6
2 »	1	0	29	40	37	56.6
3 »	0	1	56	25	37	56.6
4 »	1	1	19	160	42	98.6
5 »	0	0	44	18	42	98.6
6 »	0	0	20	50	42	98.6
7 »	0	0	28	85	10	22
8 »	0	1	55	50	10	22
9 »	1	0	15	9	10	22

Figure E38.1 Converted Data Set

E38.4 Command for the Discrete Choice Model

The essential command for the discrete choice models is

CLOGIT ; Lhs = variable which indicates the choice made
; Choices = a set of J names for the set of choices
; Rhs = choice varying attributes in the utility functions
; Rh2 = choice invariant variables, including *one* for ASCs $

(The command **DISCRETE CHOICE** may also be used.)

The command builder for this model is found in Model:Discrete Choice/Discrete Choice. The model and the choice set are set up on the Main page. The Rhs variables (attributes) and Rh2 variables (characteristics) are defined on the Options page. Note in the two windows on the Options page, the Rhs of the model is defined in the left window and the Rh2 variables are specified in the right window.

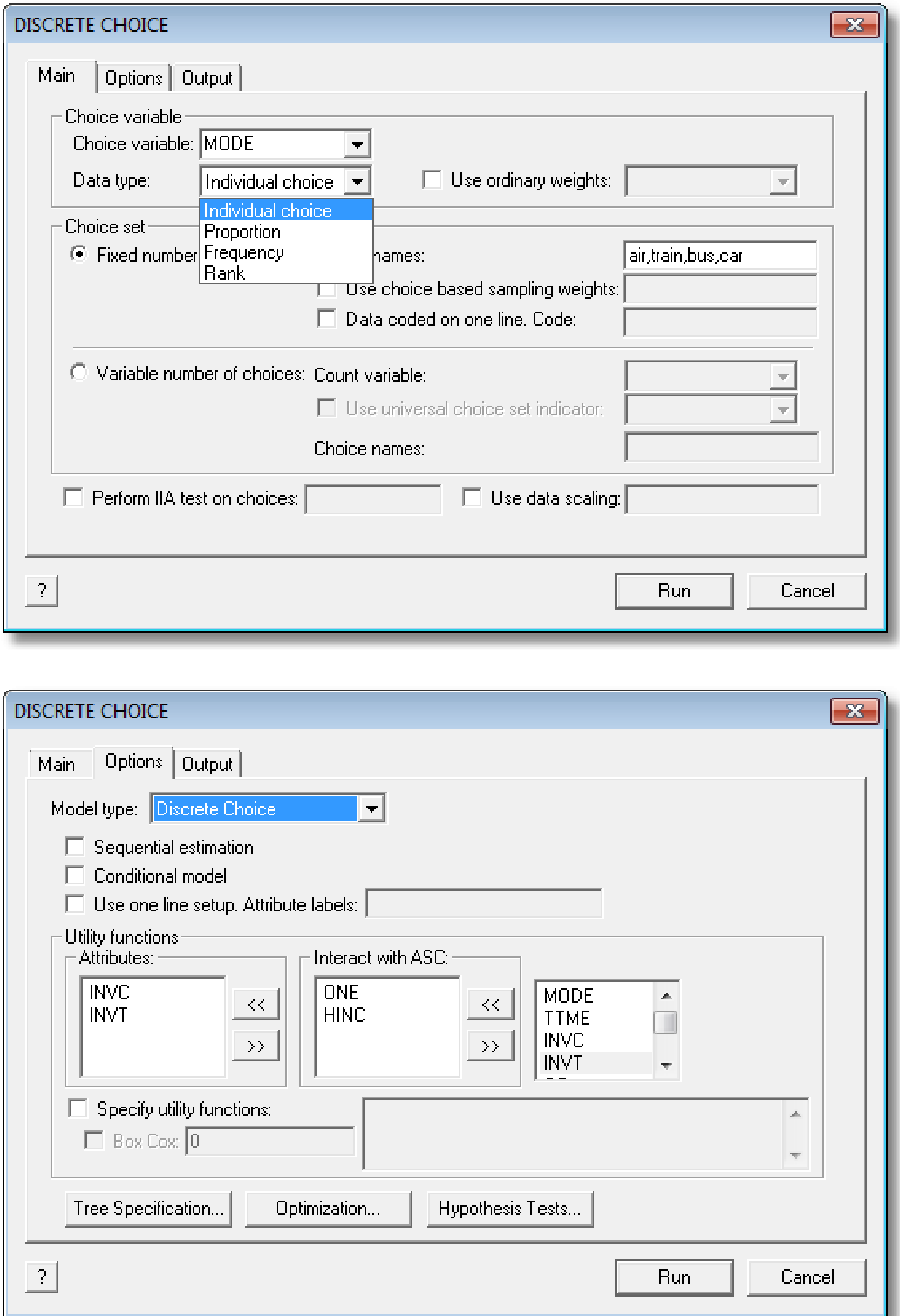


Figure E38.2 Command Builder for the Conditional Logit Model

A set of exactly J choice labels must be provided in the command. These are used to label the choices in the output. The number you provide is used to determine the number of choices there are in the model. Therefore, the set of the right number of labels is essential. Use any descriptor of eight or fewer characters desired – these do not have to be valid names, just a set of labels, separated in the list by commas.

The internal limit on J, the number of choices, is 500.

There are K attributes (Rhs variables) measured for the choices. The next chapter will describe variations of this for different formulations and options. The total number of parameters in the utility functions will include K_1 for the Rhs variables and $(J\text{-}1)K_2$ for the Rh2 variables. The total number of utility function parameters is thus $K = K_1 + (J\text{-}1)K_2$.

The internal limit on K, the number of utility function parameters, is 300.

The random utility model specified by this setup is precisely of the form

$$U_{i,j} = \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_{K1} x_{i,K1} + \gamma_{1,j} z_{i,1} + ... + \gamma_{K2,j} z_{i,K2} + \varepsilon_{i,j}$$

where the x variables are given by the Rhs list and the z variables are in the Rh2 list. By this specification, the same attributes and the same characteristics appear in all equations, at the same position. The parameters, β_k appear in all equations, and so on. There are various ways to change this specification of the utility functions – i.e., the Rhs of the equations that underlie the model, and several different ways to specify the choice set. These will be discussed at various points below.

Unlabeled Choice Sets

In some situations, particularly in choice experiments and survey data, the choices will not be a well defined set of alternatives such as (air, train, bus, car), but, rather will simply be a set of unordered choices distinguished only by the different attributes. For example, in a marketing experiment, the choice set might consist of (first, second, third, none of these). When the choice set does not have natural labels, you may use

; Choices = number_name

to define the list. For our example, we might use

; Choices = 3_brand, none

which produces the list (*brand*1,*brand*2,*brand*3,*none*).

Standard Model Specifications for the Conditional Logit (CLOGIT) Model

This is the full list of general that apply to this model.

Controlling Output from Model Commands

; Partial Effects	displays partial effects. (Use **; Effects: specification**.)
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Choice	uses choice based sampling (sandwich with weighting) estimated matrix. (This is specified in the **; Choices = list** specification for this model.)
; Cluster = name	requests computation of the cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm. Newton's method is best. BFGS is occasionally needed.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Prob = name	saves probabilities as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E38.5 Results for the Conditional Logit Model

The output for the CLOGIT estimator may contain a description of the model before the statistical results. The description consists of a table that shows the sample proportions (and a 'tree' structure that is not useful here) and one that lists the components of the utility functions. You can request these two listings by adding

; Show Model

to your **CLOGIT** command. Starting values for the iterations are either zeros or the values you provide with **; Start = list**. As such, there is no initial listing of OLS results. Output begins with the final results for the model. Here is a sample: The command is

```
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc
          ; Rh2 = one,hinc
          ; Show Model $
```

The full set of results is as follows:

```
Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.
+----------------+------+---
|Choice   (prop.)|Weight|IIA
+----------------+------+---
|AIR       .27619| 1.000|
|TRAIN     .30000| 1.000|
|BUS       .14286| 1.000|
|CAR       .28095| 1.000|
+----------------+------+---
+-----------------------------------------------------------+
| Model Specification:  Table entry is the attribute that    |
| multiplies the indicated parameter.                        |
+--------+------+-------------------------------------------+
| Choice |******| Parameter                                 |
|        |Row  1| INVC     INVT     GC       A_AIR    AIR_HIN1  |
|        |Row  2| A_TRAIN  TRA_HIN2 A_BUS    BUS_HIN3           |
+--------+------+-------------------------------------------+
|AIR     |     1| INVC     INVT     GC       Constant HINC      |
|        |     2| none     none     none     none               |
|TRAIN   |     1| INVC     INVT     GC       none     none      |
|        |     2| Constant HINC     none     none               |
|BUS     |     1| INVC     INVT     GC       none     none      |
|        |     2| none     none     Constant HINC               |
|CAR     |     1| INVC     INVT     GC       none     none      |
|        |     2| none     none     none     none               |
+-----------------------------------------------------------+
Normal exit:   5 iterations. Status=0, F=    246.1098
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function      -246.10979
Estimation based on N =    210, K =   9
Inf.Cr.AIC  =  510.220 AIC/N =    2.430
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .1327 .1201
Chi-squared[ 6]          =     75.29796
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.04613***      .01665    -2.77  .0056     -.07876    -.01349
    INVT|    -.00839***      .00214    -3.92  .0001     -.01258    -.00419
      GC|     .03633**       .01478     2.46  .0139      .00737     .06530
   A_AIR|   -1.31602*        .72323    -1.82  .0688    -2.73353     .10148
AIR_HIN1|     .00649         .01079      .60  .5477     -.01467     .02765
 A_TRAIN|    2.10710***      .43180     4.88  .0000     1.26079    2.95341
TRA_HIN2|    -.05058***      .01207    -4.19  .0000     -.07424    -.02693
   A_BUS|     .86502*        .50319     1.72  .0856     -.12120    1.85125
BUS_HIN3|    -.03316**       .01299    -2.55  .0107     -.05862    -.00770
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

NOTE: (This is one of our frequently asked questions.) The '*R*-squareds' shown in the output are R^2s in name only. They do not measure the fit of the model to the data. It has become common for researchers to report these with results as a measure of the improvement that the model gives over one that contains only a constant. But, users are cautioned not to interpret these measures as suggesting how well the model predicts the outcome variable. It is essentially unrelated to this.

To underscore the point, we will examine in detail the computations in the diagnostic measures shown in the box that precedes the coefficient estimates. Consider the example below, which was produced by fitting a model with five coefficients subject to two restrictions, or three free coefficients – *npfree* = 3. (The effect is achieved by specifying **; Choices = air,(train),(bus),car**.

```
+-------------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found  93 bad observations among     210 individuals. |
|You can use ;CheckData to get a list of these points. |
+-------------------------------------------------------+
Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.
+----------------+------+---
|Choice   (prop.)|Weight|IIA
+----------------+------+---
|AIR       .49573| 1.000|
|TRAIN     .00000| 1.000|*
|BUS       .00000| 1.000|*
|CAR       .50427| 1.000|
+----------------+------+---
```

```
+----------------------------------------------------------------+
| Model Specification:  Table entry is the attribute that        |
| multiplies the indicated parameter.                            |
+--------+------+------------------------------------------------+
| Choice |******| Parameter                                      |
|        |Row  1| GC        TTME      A_AIR     A_TRAIN  A_BUS   |
+--------+------+------------------------------------------------+
|AIR     |     1| GC        TTME      Constant none     none     |
|TRAIN   |     1| GC        TTME      none     Constant none     |
|BUS     |     1| GC        TTME      none     none     Constant |
|CAR     |     1| GC        TTME      none     none     none     |
+----------------------------------------------------------------+
Normal exit from iterations. Exit status=0.
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -62.58418
Estimation based on N =    117, K =   3
Inf.Cr.AIC  =  131.168 AIC/N =    1.121
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only    -81.0939  .2283 .2079
Chi-squared[ 2]          =     37.01953
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped   93 obs
Restricted choice set. Excluded choices are
TRAIN    BUS
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|     .01320*        .00695     1.90  .0574     -.00042    .02682
    TTME|    -.07141***      .01605    -4.45  .0000     -.10286   -.03996
   A_AIR|    3.96117***      .98004     4.04  .0001     2.04032   5.88201
 A_TRAIN|        0.0    .....(Fixed Parameter).....
   A_BUS|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

There are 210 individuals in the data set, but this model was fit to a restricted choice set which reduced the data set to $n = 210 - 93 = 117$ useable observations. The original choice set had $J_i = 4$ choices, but two were excluded, leaving $J_i = 2$ in the sample. The log likelihood of -62.58418 is computed as shown in Section E38.6. The 'constants only' log likelihood is obtained by setting each choice probability to the sample share for each outcome in the choice set. For this application, those are 0.49573 for air and 0.50427 for car. (This computation cannot be done if the choice set varies by person or if weights or frequencies are used.) Thus, the log likelihood for the restricted model is

$$\text{Log } L_0 = 117\,(\,0.49573 \times \log 0.49573 + 0.50427 \times \log 0.50427\,) = -81.09395.$$

The 'R^2' is 1 - (-62.54818/-81.0939) = 0.22869 (including some rounding error). The adjustment factor is

$$K = (\Sigma_i J_i - n) / [(\Sigma_i J_i - n) - npfree] = (234 - 117)/(234 - 117 - 3) = 1.02632.$$

and the '*Adjusted* R^2' is $1 - K(\log L / \text{Log}L_0)$

$$\textit{Adjusted } R^2 = 1 - 1.02632\,(-62.54818/-81.0939) = 0.20794.$$

Results kept by this estimator are:

Matrices:	*b* and *varb*	=	coefficient vector and asymptotic covariance matrix
Scalars:	*logl*	=	log likelihood function
	nreg	=	N, the number of observational units
	kreg	=	the number of Rhs variables
Last Model:	*b_variable*	=	the labels kept for the **WALD** command

NOTE: This estimator does not use **PARTIALS** or **SIMULATE** after estimation. Self contained routines are contained in the estimator. These are described in Chapter E40.

In the *Last Model*, groups of coefficients for variables that are interacted with constants get labels *choice_variable*, as in *trai_gco*. (Note that the names are truncated – up to four characters for the choice and three for the attribute.) The alternative specific constants are *a_choice*, with names truncated to no more than six characters. For example, the sum of the three estimated choice specific constants could be analyzed as follows:

WALD ; Fn1 = a_air + a_train + a_bus $

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors for nonlinear
functions and joint test of nonlinear restrictions.
Wald Statistic               =     16.33643
Prob. from Chi-squared[ 1] =       .00005
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
WaldFcns|  Coefficient     Error      z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
 Fncn(1)|    3.96117***     .98004    4.04  .0001     2.04032   5.88201
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E38.5.1 Robust Standard Errors

The ‘cluster’ estimator described in Chapter R10 is available in CLOGIT. However, this routine does not support hierarchical samples. There may be only one level of clustering. Also, the cluster specification is defined with respect to the CLOGIT groups of data, not the data set. CLOGIT sorts out how many clusters there are and how they are delineated. But, since the row count of the data set is used in constructing the estimator, you must treat a group of NALT observations as one. For example, our sample data used in this section contain 210 groups of four rows of data. Each group of four is an observation. Suppose that these data were grouped in clusters of three choice situations. The estimation command with the cluster estimator would appear

CLOGIT ; ... (the model) ; Cluster = 3 $

The relevant part of the output would appear as follows:

```
+----------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.     |
| Sample of    210 observations contained     70 clusters defined by  |
|      3 observations (fixed number) in each cluster.                 |
+----------------------------------------------------------------------+
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Estimation based on N =    210, K =   9
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.04613**       .01836    -2.51  .0120     -.08211   -.01014
(rows omitted)
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Use **; Cluster** as per the other models in *LIMDEP* – the same construction is used here.

E38.5.2 Descriptive Statistics

You may request a set of descriptive statistics for your model by adding

; Describe

to the model command. For each alternative, a table is given which lists the nonzero terms in the utility function and the means and standard deviations for the variables that appear in the utility function. Values are given for all observations and for the individuals that chose that alternative. For the example shown above, the following tables would be produced:

CLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = invc,invt,gc ; Rh2 = one,hinc
; Describe $

```
+---------------------------------------------------------------------------+
|             Descriptive Statistics for Alternative AIR                    |
|     Utility Function               |                    |     58.0 observs.  |
|     Coefficient                    | All      210.0 obs.|that chose AIR     |
| Name              Value  Variable  | Mean      Std. Dev.|Mean       Std. Dev. |
| -------------------  --------  | ------------------+------------------- |
| INVC          -.0461  INVC     |   85.252     27.409|   97.569     31.733 |
| INVT          -.0084  INVT     |  133.710     48.521|  124.828     50.288 |
| GC             .0363  GC       |  102.648     30.575|  113.552     33.198 |
| A_AIR        -1.3160  ONE      |    1.000       .000|    1.000       .000 |
| AIR_HIN1       .0065  HINC     |   34.548     19.711|   41.724     19.115 |
+---------------------------------------------------------------------------+
```

```
+----------------------------------------------------------------------------+
|             Descriptive Statistics for Alternative TRAIN                   |
|     Utility Function                    |                |     63.0 observs.     |
|     Coefficient                         | All      210.0 obs.|that chose TRAIN   |
| Name              Value  Variable | Mean      Std. Dev.|Mean      Std. Dev. |
| -------------------  -------- | -------------------+------------------- |
| INVC            -.0461  INVC      |  51.338     27.032|  37.460    20.676 |
| INVT            -.0084  INVT      | 608.286    251.797| 532.667   249.360 |
| GC               .0363  GC        | 130.200     58.235| 106.619    49.601 |
| A_TRAIN         2.1071  ONE       |   1.000       .000|   1.000      .000 |
| TRA_HIN2        -.0506  HINC      |  34.548     19.711|  23.063    17.287 |
+----------------------------------------------------------------------------+
+----------------------------------------------------------------------------+
|             Descriptive Statistics for Alternative BUS                     |
|     Utility Function                    |                |     30.0 observs.     |
|     Coefficient                         | All      210.0 obs.|that chose BUS     |
| Name              Value  Variable | Mean      Std. Dev.|Mean      Std. Dev. |
| -------------------  -------- | -------------------+------------------- |
| INVC            -.0461  INVC      |  33.457     12.591|  33.733    11.023 |
| INVT            -.0084  INVT      | 629.462    235.408| 618.833   273.610 |
| GC               .0363  GC        | 115.257     44.934| 108.133    43.244 |
| A_BUS            .8650  ONE       |   1.000       .000|   1.000      .000 |
| BUS_HIN3        -.0332  HINC      |  34.548     19.711|  29.700    16.851 |
+----------------------------------------------------------------------------+
+----------------------------------------------------------------------------+
|             Descriptive Statistics for Alternative CAR                     |
|     Utility Function                    |                |     59.0 observs.     |
|     Coefficient                         | All      210.0 obs.|that chose CAR     |
| Name              Value  Variable | Mean      Std. Dev.|Mean      Std. Dev. |
| -------------------  -------- | -------------------+------------------- |
| INVC            -.0461  INVC      |  20.995     14.678|  15.644     9.629 |
| INVT            -.0084  INVT      | 573.205    274.855| 527.373   301.131 |
| GC               .0363  GC        |  95.414     46.827|  89.085    49.833 |
+----------------------------------------------------------------------------+
```

You may also request a cross tabulation of the model predictions against the actual choices. (The predictions are obtained as the integer part of $\Sigma_t \hat{P}_{jt} y_{jt}$.) Add

; Crosstab

to your model command. For the same model, this would produce

```
+--------------------------------------------------------+
| Cross tabulation of actual choice vs. predicted P(j)  |
| Row indicator is actual, column is predicted.          |
| Predicted total is F(k,j,i)=Sum(i=1,...,N) P(k,j,i).  |
| Column totals may be subject to rounding error.        |
+--------------------------------------------------------+
--------+-------------------------------------------------------------
NLOGIT Cross Tabulation for 4 outcome Multinomial Choice Model
CrossTab|          AIR          TRAIN            BUS            CAR          Total
--------+-------------------------------------------------------------
     AIR|           19             13              8             18             58
   TRAIN|           12             30              9             12             63
     BUS|           10              8              6              6             30
     CAR|           17             12              7             23             59
--------+-------------------------------------------------------------
   Total|           58             63             30             59            210
```

```
+----------------------------------------------------+
| Cross tabulation of actual y(ij) vs. predicted y(ij)  |
| Row indicator is actual, column is predicted.        |
| Predicted total is N(k,j,i)=Sum(i=1,...,N) Y(k,j,i).  |
| Predicted y(ij)=1 is the j with largest probability.  |
+----------------------------------------------------+
--------+----------------------------------------------------------------
NLOGIT Cross Tabulation for 4 outcome Multinomial Choice Model
CrossTab|          AIR         TRAIN           BUS           CAR         Total
--------+----------------------------------------------------------------
     AIR|           23            15             0            20            58
   TRAIN|            8            49             0             6            63
     BUS|           13            12             1             4            30
     CAR|           15            13             0            31            59
--------+----------------------------------------------------------------
   Total|           59            89             1            61           210
```

E38.6 Estimating and Fixing Coefficients

Maximum likelihood estimates are obtained by Newton's method. Since this is a particularly well behaved estimation problem, zeros are used for the start values with little loss in computational efficiency. The gradient and Hessian used in iterations and for the asymptotic covariance matrix are computed as follows:

d_{ji} = 1 if individual i makes choice j and 0 otherwise

$$P_{ji} = \text{Prob}[y_i = j] = \text{Prob}[d_{ji} = 1] = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ji})}{\sum_{m=1}^{J_i}\exp(\boldsymbol{\beta}'\mathbf{x}_{mi})}$$

$$\text{Log } L = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ji}\log P_{ji}$$

$$\overline{\mathbf{x}}_i = \sum_{j=1}^{J_i} P_{ji}\mathbf{x}_{ji},$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ji}(\mathbf{x}_{ji} - \overline{\mathbf{x}}_i),$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} P_{ji}(\mathbf{x}_{ji} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ji} - \overline{\mathbf{x}}_i)',$$

Occasionally, a data set will be such that Newton's method does not work – this tends to occur when the log likelihood is flat in a broad range of the parameter space. There is no way that you can discern this from looking at the data, however. If Newton's method fails to converge in a small number of iterations, unless the data make estimation impossible, you should be able to estimate the model by using

; Alg = BFGS

as an alternative. The BFGS algorithm will take slightly longer, but for most data sets, the difference will be a few seconds. If this method fails as well, you should conclude that your model is inestimable.

You may provide your own starting values with

; Start = list of K values

If you have requested a set of alternative specific constants, you must provide starting values for them as well. *Regardless of where 'one' appears in the Rhs list, the ASCs will be the last J-1 coefficients corresponding to that list. If you have Rh2 variables, the coefficients will follow the Rhs coefficients, including the list of ASCs.*

Coefficients may be fixed at specific values during optimization. Use

; Fix = variable name [value]

for example, **; Fix = ttme [.01]**

The following results are obtained from

```
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,ttme
          ; Rh2 = one
          ; Fix = ttme[.01] $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -287.31412
Estimation based on N =    210, K =   4
Inf.Cr.AIC  =  582.628 AIC/N =    2.774
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588 -.0125-.0190
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|    -.02118***      .00403    -5.26  .0000     -.02908   -.01329
    TTME|     .01000    .....(Fixed Parameter).....
   A_AIR|    -.53263***      .19044    -2.80  .0052     -.90589   -.15937
 A_TRAIN|     .40186*        .22238     1.81  .0708     -.03400    .83773
   A_BUS|    -.66610***      .23961    -2.78  .0054    -1.13572   -.19648
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

E38.7 MLOGIT and CLOGIT

When there are no choice varying attributes, CLOGIT is the same model as MLOGIT. From Chapter E37, the functional form for MLOGIT is

$$\text{Prob}(y_i = j|\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_i)}{\sum_{m=1}^{J}\exp(\boldsymbol{\beta}_m'\mathbf{x}_i)}, \; j = 0,\ldots,J,$$

From the introduction in this chapter,

$$\text{Prob}(\text{choice} = j \mid \mathbf{X}_i,\mathbf{z}_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ji} + \boldsymbol{\gamma}_j'\mathbf{z}_i)}{\sum_{m=0}^{J}\exp(\boldsymbol{\beta}'\mathbf{x}_{mi} + \boldsymbol{\gamma}_m'\mathbf{z}_i)}, \; j = 1,\ldots,J.$$

In the second equation, if $\boldsymbol{\beta}$ equals zero – there are no choice varying attributes – then the second probability is the same as the first, after a simple renaming of the parts; $\boldsymbol{\gamma}_j$ in the second replacing $\boldsymbol{\beta}_j$ in the first, and $\mathbf{z}_i$ replacing $\mathbf{x}_i$. (The alternatives are renumbered, indexing from 1 to J rather than from 0 to J.) The following illustrates the result:

```
? CLOGIT using the original data
CLOGIT      ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = one ; Rh2 = hinc
            ; Effects: hinc(*) $
? Create the dependent variable for MLOGIT, using the first row of clogit data
CREATE      ; pick = mode*(0*aasc+1*tasc+2*basc+3*casc) $
CREATE      ; choice = 3 - (pick+pick[+1]+pick[+2]+pick[+3]) $
? Use only the first row for MLOGIT
MLOGIT      ; If[aasc = 1 ] ; Lhs = choice ; Rhs = one,hinc
            ; Partial Effects
            ; Labels = car,bus,train,air $
```

We have normalized MLOGIT so that *choice* = 0 means pick *car* and *choice* = 3 means pick *air*. The elasticities then correspond to those in the CLOGIT results, and the coefficients are the same.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -261.74506
Estimation based on N =    210, K =   6
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   A_AIR|     .04252          .45456      .09  .9255     -.84840    .93345
 A_TRAIN|    2.00595***       .42180     4.76  .0000     1.17923   2.83266
   A_BUS|     .64169          .49249     1.30  .1926     -.32358   1.60696
AIR_HIN1|    -.00142          .00989     -.14  .8858     -.02081    .01797
TRA_HIN2|    -.06048***       .01169    -5.17  .0000     -.08339   -.03756
BUS_HIN3|    -.03677***       .01282    -2.87  .0041     -.06190   -.01165
--------+--------------------------------------------------------------------

--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Elasticity of Choice Probabilities with Respect to HINC
--------+-----------------------------------
        |     AIR    TRAIN      BUS      CAR
--------+-----------------------------------
    HINC|   .5418  -1.4986   -.6796    .5908

-----------------------------------------------------------------------------
Multinomial Logit Model
Dependent variable               CHOICE
Log likelihood function      -261.74506
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[BUS     ]
Constant|     .64169          .49249     1.30  .1926     -.32358   1.60696
    HINC|    -.03677***       .01282    -2.87  .0041     -.06190   -.01165
        |Characteristics in numerator of Prob[TRAIN   ]
Constant|    2.00595***       .42180     4.76  .0000     1.17923   2.83266
    HINC|    -.06048***       .01169    -5.17  .0000     -.08339   -.03756
        |Characteristics in numerator of Prob[AIR     ]
Constant|     .04252          .45456      .09  .9255     -.84840    .93345
    HINC|    -.00142          .00989     -.14  .8858     -.02081    .01797
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Averages of Individual Elasticities of Probabilities
--------+---------+---------+---------+---------+
Variable|     CAR |     BUS |   TRAIN |     AIR |
--------+---------+---------+---------+---------+
HINC    |   .5908 |  -.6796 | -1.4986 |   .5418 |
--------+---------+---------+---------+---------+
```

E39: Specifications of the Conditional Logit Models

E39.1 Introduction

Chapter E38 described how to fit the generic form of the multinomial logit model for multinomial choice. This chapter presents some modifications of the basic command that accommodate more general choice sets (possibly varying across individuals) and a convenient alternative command format that allows more general specifications of the utility functions. Two modifications of the estimator are described, one for the case in which certain attributes are ignored by some of the sampled individuals and a second that is based on the maximum entropy criterion rather than maximum likelihood.

E39.2 Choice Sets

In the standard case, data on the Lhs variable will consist of a column of J-1 zeros and a one for the choice made, when reading down the J rows of data for the individual. We allow other types of data on the choice variable. If you have grouped data, the values will be proportions or frequencies, instead. For proportions data, within each observation (J data points), the values of the Lhs variable will sum to one when summed *down* the J rows. (This will be the only difference in the grouped data treatment.) With frequencies, the values will simply be a set of nonnegative integers. An example of a setting in which such data might arise would be in marketing, where the proportions might be market shares of several brands of a commodity. Alternatively, the choice variable might be a set of ranks, in which case, instead of zeros and ones, the Lhs variable would take values 1,2,...,J (not necessarily in that order) within, and reading down, each block.

E39.2.1 Fixed and Variable Numbers of Choices

When every individual in the sample chooses from the same choice set, and all alternatives are available to all individuals, then the data set will appear as in the example developed in Chapter E38, and will consist of n sets of J 'observations.' You indicate this case with a command such as:

```
CLOGIT    ; Lhs = the choice variable
          ; Choices = ... a list of J names for the choices
          ; ... the rest of the command $
```

For example,

```
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; etc. $
```

There are many cases in which the choice set will vary from one individual to another. We consider the random choice model first in which the number of choices is not constant from one observation to the next. Ranks data are considered later. Two possible arrangements that might produce variable sized choice sets are as follows:

- There is a *universal choice set*, from which individuals make their choice. But, not all choices are available to all individuals. Consider, for example, the choice of travel mode among *train, bus, car, ferry*. If respondents are observed at many different locations, one or more of the choices, such as *ferry* or *train*, might be unavailable to them, and those might vary from person to person. In this case, there is a fixed set of J alternatives, but each individual chooses among their own J_i choices. This is called a 'labeled' choice set.

- Individuals each choose among their own set of J_i alternatives. However, there is no universal choice set. Consider, for example, the choice of which shopping center to shop at. If observations are taken in many different cities, we will observe numerous different choice sets, but there is no well defined universal choice set. This is called an 'unlabeled' choice set.

Unlabeled choice sets often arise in survey data, or 'stated choice experiments.' In a stated choice experiment, an individual might be offered a set of J_i alternatives that are only differentiated by their attributes. Configurations of features in a choice set of cars or appliances might be such a case. In this instance, the choices are simply numbered, 1,2,…

Either of these cases can be accommodated with **CLOGIT**. For both cases, you will provide a variable which gives the number of choices for each observation. This variable is then a second **; Lhs** specification. The command for an unlabeled choice set, which is the simpler case, becomes

```
CLOGIT    ; Lhs = y,nij
          ; ... specification of the utility functions
          ; ... the rest of the command $
```

Note that the **; Choices = list** is not defined in the command, since in this case, there is no clearly defined choice set. Nothing else need be changed. *LIMDEP* does all of the accounting internally. In this case, it is simply assumed that each individual has their own choice set.

For example, one such data set might appear as follows.

	y	*q*	*w*	*nij*
i=1	0	$q_{1,1}$	$w_{1,1}$	3
	—>1	$q_{2,1}$	$w_{2,1}$	3
	0	$q_{3,1}$	$w_{3,1}$	3
i=2	0	$q_{1,2}$	$w_{1,2}$	4
	0	$q_{2,2}$	$w_{2,2}$	4
	—>1	$q_{3,2}$	$w_{3,2}$	4
	0	$q_{4,2}$	$w_{4,2}$	4
i=3	—>1	$q_{1,3}$	$w_{1,3}$	2
	0	$q_{2,3}$	$w_{2,3}$	2

Note that *nij* is the usual group size variable for a panel in *LIMDEP*. The model command might be

CLOGIT ; Lhs = y,nij ; Rhs = q,w $

Notice, once again, that the command does not contain a definition of the choice set, such as **; Choices = list** specification.

For the case of a universal choice set, suppose that the data set above were, instead:

	Y	*q*	*w*	*nij*	*altij*
i=1	0	$q_{1,1}$	$w_{1,1}$	3	1 (Air)
	—>1	$q_{2,1}$	$w_{2,1}$	3	2 (Train)
	0	$q_{3,1}$	$w_{3,1}$	3	4 (Car)
i=2	0	$q_{1,2}$	$w_{1,2}$	4	1 (Air)
	0	$q_{2,2}$	$w_{2,2}$	4	2 (Train)
	—>1	$q_{3,2}$	$w_{3,2}$	4	3 (Bus)
	0	$q_{4,2}$	$w_{4,2}$	4	4 (Car)
i=3	—>1	$q_{1,3}$	$w_{1,3}$	2	3 (Bus)
	0	$q_{2,3}$	$w_{2,3}$	2	4 (Car)

The specific choice identifier, when it is needed, is provided as a *third* Lhs variable. For this case, the choice set would have to be defined. For example,

CLOGIT ; Lhs = y,nij,altij
; Choices = air,train,bus,car
; Rhs = q,w $

In this case, every individual is assumed to choose from a set of four alternatives, though the *altij* variable indicates that some of these choices are unavailable to some individuals.

Do note that if you are not defining a universal choice set, *LIMDEP* simply uses the largest number of choices for any individual in the sample to determine *J* for the model. As such, an expanded set of choice specific constants is not likely to be meaningful, though you can create one with **; Rh2 = one**. Also, if you do not specify a universal choice set, the variable *altij* will not be meaningful.

E39.2.2 Restricting the Choice Set

The IIA test described later in Section E40.4 is carried out by fitting the model to a restricted choice set, then comparing the two sets of parameter estimates. You can restrict the choice set used in estimation, irrespective of the IIA test, by a slight change in the command. In the **; Choices = list of alternatives** specification, enclose any choices to be excluded in parentheses. For example, in our **CLOGIT** application, the specification

; Choices = air,(train),(bus),car

produces the following display in the model output:

```
+------------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found  93 bad observations among     210 individuals. |
|You can use ;CheckData to get a list of these points. |
+------------------------------------------------------+
Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.
+----------------+------+---
|Choice   (prop.)|Weight|IIA
+----------------+------+---
|AIR       .49573| 1.000|
|TRAIN     .00000| 1.000|*
|BUS       .00000| 1.000|*
|CAR       .50427| 1.000|
+----------------+------+---
Normal exit:   6 iterations. Status=0, F=    52.79148
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function        -52.79148
Estimation based on N =    117, K =   5
Number of obs.=   210, skipped   93 obs
Restricted choice set. Excluded choices are
TRAIN    BUS
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.04871*        .02757    -1.77  .0772     -.10274    .00532
    INVT|    -.01195***      .00395    -3.03  .0025     -.01969   -.00422
      GC|     .08576***      .02654     3.23  .0012      .03374    .13778
    TTME|    -.08222***      .01854    -4.43  .0000     -.11855   -.04588
   A_AIR|    2.12899*       1.20531     1.77  .0773     -.23337   4.49135
 A_TRAIN|        0.0    .....(Fixed Parameter).....
   A_BUS|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
```

Note that as in the IIA test, this procedure results in exclusion of some ‘bad’ observations, that is, the ones that selected the excluded choices. Because of the model specification, the ASCs for bus and train have been fixed at zero.

You may combine the choice based sampling estimator with the restricted choice set. All the necessary adjustments of the weights are made internally. Thus, the specification

; Choices = air,(train),(bus),car / .14,.13,.09,.64

produces the following listing:

```
+----------------+------+---+
|Choice   (prop.)|Weight|IIA|
+----------------+------+---+
|AIR       .49573|  .387|   |
|TRAIN     .00000|  .000| * |
|BUS       .00000|  .000| * |
|CAR       .50427| 1.739|   |
+----------------+------+---+
```

E39.2.3 Very Large Choice Sets

The conditional logit estimator can fit a model with up to 500 choices, which is quite large. However, certain applications, such as home purchase choice, have involved many more than that. CLOGIT and the other estimators in *LIMDEP* are bound by certain internal limits. However, it is possible to stretch the estimator a bit more. It turns out that Chamberlain's fixed effects model for the binary logit model described in Section E30.5 can be used to fit a discrete choice model. The log likelihood function for this model is

$$L_c = \frac{\prod_{t=1}^{T_i} \exp[y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}]}{\sum_{\textit{all arrangements of } T_i \textit{ outcomes with the same sum}} \prod_{s=1}^{T_i} \exp[y_{is}\boldsymbol{\beta}'\mathbf{x}_{is}]}$$

$$= \frac{\exp\left[\sum_{t=1}^{T_i} y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}\right]}{\sum_{\textit{all arrangements of } T_i \textit{ outcomes with the same sum}} \exp\left[\sum_{s=1}^{T_i} d_{is}\boldsymbol{\beta}'\mathbf{x}_{is}\right]}.$$

If the group of observations has exactly one '1' and T_i - 1 '0s,' then this is exactly the log likelihood for the discrete choice model that we have analyzed in this chapter. Thus, if the group of observations for individual *i* is treated as if this were a fixed effects model, then this estimator can be used to obtain parameter estimates. The command setup would be

```
LOGIT       ; Lhs = choice
            ; Rhs = the set of variables
            ; Pds = the number of choices $
```

This arrangement will allow up to 200 choices. The only shortcoming (aside from the greatly restricted number of optional features) is that unless you can provide the actual dummy variables, as we do below, it is not possible to specify a set of choice specific constants with this estimator. Two ways to fit the model in our example would be

```
CLOGIT      ; Lhs = mode
            ; Rhs = invc,invt,gc,ttme
            ; Rh2 = one
            ; Choices = air,train,bus,car $
LOGIT       ; Lhs = mode
            ; Rhs = aasc,tasc,basc,invc,invt,gc,ttme
            ; Pds = 4 $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function      -184.50669
Estimation based on N =    210, K =   7
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.08493***      .01938    -4.38  .0000     -.12292   -.04694
    INVT|    -.01333***      .00252    -5.30  .0000     -.01827   -.00840
      GC|     .06930***      .01743     3.97  .0001      .03513    .10346
    TTME|    -.10365***      .01094    -9.48  .0000     -.12509   -.08221
   A_AIR|    5.20474***      .90521     5.75  .0000     3.43056   6.97893
 A_TRAIN|    4.36060***      .51067     8.54  .0000     3.35972   5.36149
   A_BUS|    3.76323***      .50626     7.43  .0000     2.77098   4.75548
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+---------------------------------------------------+
| Panel Data Binomial Logit Model                   |
| Number of individuals          =     210          |
| Number of periods              =       4          |
| Conditioning event is the sum of MODE             |
| Distribution of sums over the  4 periods:         |
| Sum          0     1     2     3     4     5     6 |
| Number       0   210     0     0     0     5     6 |
| Pct.      .00100.00   .00   .00   .00   .00   .00 |
+---------------------------------------------------+
Normal exit:   6 iterations. Status=0, F=    184.5067
```

```
-----------------------------------------------------------------------------
Logit Model for Panel Data
Dependent variable                  MODE
Log likelihood function      -184.50669
Estimation based on N =    840, K =   7
Fixed Effect Logit Model for Panel Data
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    AASC|    5.20474***      .90521     5.75  .0000     3.43056   6.97893
    TASC|    4.36060***      .51067     8.54  .0000     3.35972   5.36149
    BASC|    3.76323***      .50626     7.43  .0000     2.77098   4.75548
    INVC|    -.08493***      .01938    -4.38  .0000     -.12292   -.04694
    INVT|    -.01333***      .00252    -5.30  .0000     -.01827   -.00840
      GC|     .06930***      .01743     3.97  .0001      .03513    .10346
    TTME|    -.10365***      .01094    -9.48  .0000     -.12509   -.08221
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E39.3 Weighting

You can, in principle, use any weighting variable you wish with this model to weight observations. The model does not require that weights be the same for all outcomes for a given observation. For example, in a grouped data case, you might have at hand the total number of observations which gave rise to each of the proportions in the proportions data. If so, you could use the information to replicate each observation the appropriate number of times. In this case, use the

; Wts = name

option on the **CLOGIT** command, as you would with any other model. Normally, this variable would take the same value for each of the *J* data vectors associated with observation *i*. (Suppose instead of 0,1,0 for the first observation, we observed .4, .5, .1 based on 200 observations. Then, '*name*' would take the value 200 for the first three observations, etc.) (Of course, you could achieve the same result by providing the frequencies as the Lhs variable.)

E39.4 Choice Based Sampling

The weighting may be based on the outcomes. For example, suppose the model predicts mode of travel, *car, train*, or *horse*. The true population proportions are known to be .6, .35, and .05. But, we deliberately oversample the last category so that the sample proportions are, say, .5, .3, and .2. In estimation, to account for the nonrandom sampling, we would use a weighting scheme which gives observations in which outcome 1 (*car*) received a weight of .6/.5 = 1.2, outcome 2 (*train*), .35/.3 = 1.16667, and outcome 3 (*horse*), .05/.2 = .25. Notice that regardless of the number of observations, the weighting variable in this scenario takes only *J* values, where *J* is the number of outcomes. The Lerman-Manski (1981) correction to the variance matrix of the estimates is used at convergence to obtain the appropriate standard errors. The covariance matrix used is $\mathbf{V} = \mathbf{H}^{-1}\mathbf{D}\mathbf{H}^{-1}$, where **H** is the weighted Hessian and **D** is the weighted sum of the outer products of the first derivatives, as opposed to $\mathbf{V} = \mathbf{H}^{-1}$ which would be used normally.

To request this procedure, it is only necessary for you to provide the *J* population weights. Everything else is automated. The weights are provided after the labels for the outcomes following a slash. The following example is consistent with the discussion above. The unweighted specification would be

```
CLOGIT    ; ... ; Choices = car,train,horse $
```

The choice based sampling weights would be provided in

```
CLOGIT    ; ... ; Choices = car,train,horse / .6,.35,.05 $
```

Notice that you only provide the population weights. The program obtains the sample proportions and computes the appropriate weights for the estimator. This is a bit different from the earlier applications (probit and logit – see Section E27.10), and it is the only estimator in *LIMDEP* for which you provide only the population weights, as opposed to the sampling ratios.

Everything else is the same as before. Note, you *do not* use a weighting (**; Wts**) variable here. Your population weights must sum to 1.0; if not, an error occurs and estimation is halted. If you provide population weights, you must give a full set. Thus, if your list has the slash specification, the number of values after the slash must match exactly the number of labels before it.

The data used in our examples in Chapter E38 are choice based. The example below shows the use of this option to make the appropriate corrections to the estimates:

```
CLOGIT    ; Lhs = mode
          ; Rhs = invc,invt,gc,ttme
          ; Rh2 = one
          ; Choices = air,train,bus,car / .14,.13,.09,.64
          ; Show $
```

The **; Show** parameter requests the display of the table below. Otherwise, only the note in the box of diagnostic statistics indicates use of the choice based sampling estimator.)

```
Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.
+----------------+------+---
|Choice   (prop.)|Weight|IIA
+----------------+------+---
|AIR       .27619|  .507|
|TRAIN     .30000|  .433|
|BUS       .14286|  .630|
|CAR       .28095| 2.278|
+----------------+------+---
+----------------------------------------------------------------+
| Model Specification:  Table entry is the attribute that        |
| multiplies the indicated parameter.                            |
+--------+------+------------------------------------------------+
| Choice |******| Parameter                                      |
|        |Row  1| INVC     INVT     GC       TTME     A_AIR       |
|        |Row  2| A_TRAIN  A_BUS                                 |
+--------+------+------------------------------------------------+
|AIR     |     1| INVC     INVT     GC       TTME     Constant    |
|        |     2| none     none                                  |
|TRAIN   |     1| INVC     INVT     GC       TTME     none        |
|        |     2| Constant none                                  |
|BUS     |     1| INVC     INVT     GC       TTME     none        |
|        |     2| none     Constant                              |
|CAR     |     1| INVC     INVT     GC       TTME     none        |
|        |     2| none     none                                  |
+----------------------------------------------------------------+
Normal exit:   6 iterations. Status=0, F=    132.5388

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -132.53879
Estimation based on N =    210, K =   7
Vars. corrected for choice based sampling
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
```

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
INVC	-.11080***	.02336	-4.74	.0000	-.15659	-.06502
INVT	-.01736***	.00299	-5.81	.0000	-.02322	-.01151
GC	.09787***	.01967	4.98	.0000	.05931	.13643
TTME	-.13929***	.02589	-5.38	.0000	-.19003	-.08855
A_AIR	5.68250***	1.58789	3.58	.0003	2.57029	8.79472
A_TRAIN	4.09890***	.90704	4.52	.0000	2.32113	5.87667
A_BUS	3.91452***	.92554	4.23	.0000	2.10050	5.72854

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

These are the parameter estimates computed without the correction for choice based sampling. This is not only a correction to the covariance matrix. The parameter estimates will change as well.

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
INVC	-.08493***	.01938	-4.38	.0000	-.12292	-.04694
INVT	-.01333***	.00252	-5.30	.0000	-.01827	-.00840
GC	.06930***	.01743	3.97	.0001	.03513	.10346
TTME	-.10365***	.01094	-9.48	.0000	-.12509	-.08221
A_AIR	5.20474***	.90521	5.75	.0000	3.43056	6.97893
A_TRAIN	4.36060***	.51067	8.54	.0000	3.35972	5.36149
A_BUS	3.76323***	.50626	7.43	.0000	2.77098	4.75548

E39.5 Building the Utility Functions

The model specification thus far builds the utility functions from the common Rhs and Rh2 specifications. For example, in our four outcome model which contains *cost*, *time*, *one* and *income*, the data for the choice variable and the utility functions are contained in

$$\begin{array}{ccccccccc} \text{choice} & \text{cost} & \text{time} & & \text{constants} & & & \text{income} & \\ \end{array}$$

$$\mathbf{Z}_i = \begin{bmatrix} y_{air} & c_a & t_a & 1 & 0 & 0 & income & 0 & 0 \\ y_{train} & c_t & t_t & 0 & 1 & 0 & 0 & income & 0 \\ y_{bus} & c_b & t_b & 0 & 0 & 1 & 0 & 0 & income \\ y_{car} & c_c & t_c & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The utility functions are all the same;

$$\begin{aligned} U_{i,air} &= \beta_{cost}cost_{i,air} + \beta_{time}time_{i,air} + \alpha_{air} + \gamma_{air}income_i + \varepsilon_{i,air} \\ U_{i,train} &= \beta_{cost}cost_{i,train} + \beta_{time}time_{i,train} + \alpha_{train} + \gamma_{train}income_i + \varepsilon_{i,train} \\ U_{i,bus} &= \beta_{cost}cost_{i,bus} + \beta_{time}time_{i,bus} + \alpha_{bus} + \gamma_{bus}income_i + \varepsilon_{i,bus} \\ U_{i,car} &= \beta_{cost}cost_{i,car} + \beta_{time}time_{i,car} + \alpha_{car} + \gamma_{car}income_i + \varepsilon_{i,car} \end{aligned}$$

In order to prevent a multicollinearity problem, $\alpha_{car} = \gamma_{car} = 0$. One might want to have different attributes appear in the different utility functions, or impose other kinds of constraints on the parameters, or allow a generic coefficient such as β_1 to differ across groups of observations. In general, these sorts of modifications can be obtained by using transformations of the variables. For example, to have β_1 have one value for air and car and a different value for train and bus, we would use

CREATE ; costac = cost*(aasc + casc) ; costtb = cost*(tasc + basc) $

Then, we would replace *cost* with *costac,costtb* in the Rhs specification of the model. The resulting model would be

$$
\begin{aligned}
U_{i,air} &= \beta_{cost1}cost_{i,air} && + \beta_{time}time_{i,air} + \alpha_{air} + \gamma_{air}income_i + \varepsilon_{i,air}\\
U_{i,train} &= \beta_{cost2}cost_{i,train} && + \beta_{time}time_{i,train} + \alpha_{train} + \gamma_{train}income_i + \varepsilon_{i,train}\\
U_{i,bus} &= \beta_{cost2}cost_{i,bus} && + \beta_{time}time_{i,bus} + \alpha_{bus} + \gamma_{bus}income_i + \varepsilon_{i,bus}\\
U_{i,car} &= \beta_{cost1}cost_{i,car} && + \beta_{time}time_{i,car} + \alpha_{car} + \gamma_{car}income_i + \varepsilon_{i,car}
\end{aligned}
$$

This section will describe how to structure the utility functions individually, rather than generically with Rhs and Rh2 and transformations of variables.

E39.5.1 Alternative Specific Constants and Choice Invariant Variables

The CLOGIT model is homogeneous of degree zero in the generic attributes. Any attribute that does not vary across the choices, such as age, marital status, etc., will simply fall out of the probability. Consider, for example, a model that contains an attribute *cost* that varies by choice, and *income* that does not. The generic discrete choice model would specify

$$
\begin{aligned}
Prob(choice = j) &= \frac{\exp(\beta_1 cost_{i,j} + \beta_2 Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \beta_2 Income_i)}\\
&= \frac{\exp(\beta_2 Income_i)\exp(\beta_1 cost_{i,j})}{\exp(\beta_2 Income_i)\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j})}\\
&= \frac{\exp(\beta_1 cost_{i,j})}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j})}
\end{aligned}
$$

Therefore, the model in that form is not estimable. The answer, as we have seen, is to make the coefficient on choice invariant variables vary with the choices. This includes the constant term, *one*. This is how the MLOGIT model of Chapter E37 arises – in that model, all variables are choice invariant. Here, it produces a hybrid model, which can have both types of variables in the utility functions.

$$
Prob(choice = j) = \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}
$$

This is the form of the model in the earlier example,

$$
\begin{aligned}
U_{i,air} &= \beta_{cost}cost_{i,air} + \beta_{time}time_{i,air} + \alpha_{air} + \gamma_{air}income_i + \varepsilon_{i,air}\\
U_{i,train} &= \beta_{cost}cost_{i,train} + \beta_{time}time_{i,train} + \alpha_{train} + \gamma_{train}income_i + \varepsilon_{i,train}\\
U_{i,bus} &= \beta_{cost}cost_{i,bus} + \beta_{time}time_{i,bus} + \alpha_{bus} + \gamma_{bus}income_i + \varepsilon_{i,bus}\\
U_{i,car} &= \beta_{cost}cost_{i,car} + \beta_{time}time_{i,car} + \alpha_{car} + \gamma_{car}income_i + \varepsilon_{i,car}
\end{aligned}
$$

There remains an indeterminacy in the model after it is expanded in this fashion. Suppose the same constant is added to each γ_j, say θ. The resulting model is

$$
\begin{aligned}
Prob(choice = j) &= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + (\gamma_j + \theta)Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + (\gamma_j + \theta)Income_i)}\\
&= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i + \theta Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i + \theta Income_i)}\\
&= \frac{\exp(\theta Income_i)\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\exp(\theta Income_i)\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}\\
&= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}
\end{aligned}
$$

So, the identical model arises for any θ. This means that the model still cannot be estimated in this form. The solution to this remaining issue is to normalize the coefficients so that one of the choice varying parameters is equal to zero. **CLOGIT** sets the last one to zero. The same result applies to the choice specific constant terms that you create with *one*.

The basic four choice model which contains *cost*, *time*, *one* and *income* will have utility functions

$$
\begin{aligned}
U_{i,air} &= \beta_{cost}\, cost_{i,air} + \beta_{time}\, time_{i,air} + \alpha_{air} + \gamma_{air}\, income_i + \varepsilon_{i,air}\\
U_{i,train} &= \beta_{cost}\, cost_{i,train} + \beta_{time}\, time_{i,train} + \alpha_{train} + \gamma_{train}\, income_i + \varepsilon_{i,train}\\
U_{i,bus} &= \beta_{cost}\, cost_{i,bus} + \beta_{time}\, time_{i,bus} + \alpha_{bus} + \gamma_{bus}\, income_i + \varepsilon_{i,bus}\\
U_{i,car} &= \beta_{cost}\, cost_{i,car} + \beta_{time}\, time_{i,car} + \varepsilon_{i,car}
\end{aligned}
$$

The simple device you use to construct utility functions in this fashion is to use

; Rhs = list of attributes that vary across choices

and **; Rh2 = list of variables that do not vary across choices**

The Rh2 variables are automatically expanded into a set of *J*-1 interactions with the choice specific constants, as they are in the matrix $\mathbf{Z}_i$ shown above. The implication is that, generally, you do not need to have these variables in your data set. They are automatically created by your command. (Note that our clogit.dat data set actually does contain the superfluous set of four choice specific constants, *aasc*, *tasc*, *basc* and *casc*.)

NOTE: If you include *one* in your Rhs, it is automatically expanded to become a set of alternative specific constants. That is, *one* is automatically moved to the Rh2 list if it is placed in the Rhs list.

The model specification for the four utility functions shown above would be

; Rhs = cost,time ; Rh2 = one,income

Note that the distinction between Rh2 and Rhs variables is that all variables in the first category are expanded by interacting with the choice specific binary variables. (The last term is dropped.)

HINT: There are many different possible configurations of alternative specific constants (ASCs) and alternative specific variables. In estimating a model, it is not possible to determine a priori if a singularity will arise as a consequence of the specification. You will have to discern this from the estimation results for the particular model.

The constant term, *one* fits the hint above. Recognizing this, *LIMDEP* assumes that if your Rhs list includes *one*, you are requesting a set of alternative specific constants. As such, when the Rhs list includes *one*, *LIMDEP* will create a full set of *J*-1 choice specific constants. Note the earlier examples.

Finally, while it is necessary for choice invariant variables to appear in Rh2, it is not necessary that all variables in the Rh2 list actually be choice invariant. Indeed, one could specify the preceding model with choice specific coefficients on the *cost* variable; it would appear

$$
\begin{aligned}
U_{i,air} &= \gamma_{cost,air}\, cost_{i,air} + \beta_{time}\, time_{i,air} + \alpha_{air} + \gamma_{air}\, income_i + \varepsilon_{i,air}\\
U_{i,train} &= \gamma_{cost,train}\, cost_{i,train} + \beta_{time}\, time_{i,train} + \alpha_{train} + \gamma_{train}\, income_i + \varepsilon_{i,train}\\
U_{i,bus} &= \gamma_{cost,bus}\, cost_{i,bus} + \beta_{time}\, time_{i,bus} + \alpha_{bus} + \gamma_{bus}\, income_i + \varepsilon_{i,bus}\\
U_{i,car} &= \gamma_{cost,car}\, cost_{i,car} + \beta_{time}\, time_{i,car} + \varepsilon_{i,bus}
\end{aligned}
$$

Note also, that there is no need to drop one of the time coefficients because the variable *cost* varies by choices. You *can* estimate a model with four separate coefficients for *cost*, one in each utility function. However, it is not possible to do it by including *cost* in the Rh2 list as described above, because this form will automatically drop the last term (the one in the *car* utility function). You could obtain this form, albeit a bit clumsily, by creating the four interaction terms yourself and including them on the Rhs. We already have the alternative specific constants, so the following would work

```
CREATE     ; cost_a = gc * aasc
           ; cost_t  = gc * tasc
           ; cost_b = gc * basc
           ; cost_c  = gc * casc $
CLOGIT     ; ...  ; Rhs  = time,cost_a,cost_t,cost_b,cost_c
                  ; Rh2 = one,income $
```

Having to create the interaction variables is going to be inconvenient. The alternative method of specifying the model described in the next section will be much more convenient. This method also allows you much greater flexibility in specifying utility functions.

E39.5.2 Building the Utility Functions

The utility functions need not be the same for all choices. Different attributes may enter and the coefficients may be constrained in different ways. The following more flexible format can be used instead of the **; Rhs = list** and **; Rh2 = list** parts of the command described above. This format also provides a way to supply starting values for parameters, so this can also replace the **; Start = list** specification. Finally, you will also be able to use this format to fix coefficients, so it will be an easy way to replace the **; Rst = list** and **; Fix = name[value]** specifications.

We begin with the case of a fixed (and named) set of choices, then turn to the cases of variable numbers of choices. We replace the Rhs/Rh2 setup with explicit definitions of the utility functions for the alternatives. Utility functions are built up from the format

```
; Model:  U(choice 1) = linear equation  /
          U(choice 2) = linear equation /
          ...
          U(choice J) = linear equation $
```

Though we have shown all *J* utility functions, for a given model specification, you could, in principle, not specify a utility function in the list. The implied specification would be $U_{ij} = \varepsilon_{ij}$. The **: U(list)** is mandatory if the command contains **; Model =….** *LIMDEP* now scans for the '*U*' and the parentheses. For example:

```
; Model:  U(air) = ba + bcost * gc
```

Note that the specification begins with '**; Model:**' – the colon ('**:**') is also mandatory. Parameters always come first, then variables. Constant terms need not multiply variables. Thus, *ba* in this *could* be an '*Air specific constant*.' (It depends on whether *ba* appears elsewhere in the model.) Notice that the utility function defines both the variables and the parameters. Usually, you would give an equation for each choice in the model. For example:

```
CLOGIT     ; Lhs = mode
           ; Choices = air,train,bus,car
           ; Model:  U(air)   = ba +  bcost * gc + btime * ttme /
                     U(car)   = bc +  bcost * gc   /
                     U(bus)   = bb +  bcost * gc   /
                     U(train) =       bcost * gc + btime * ttme $
```

Utility functions are separated by slashes. Note also that the alternative specific constants stand alone without multiplying a variable. Your utility definitions now provide the names for the parameters. The estimates produced by this model command are as follows:

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -223.43803
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient     Error         z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      BA|    1.55491***      .37580     4.14  .0000      .81835   2.29147
   BCOST|    -.02021***      .00435    -4.65  .0000     -.02873   -.01168
   BTIME|    -.08680***      .01122    -7.73  .0000     -.10880   -.06481
      BC|   -3.65316***      .46378    -7.88  .0000    -4.56216  -2.74417
      BB|   -3.91983***      .45611    -8.59  .0000    -4.81379  -3.02586
--------+--------------------------------------------------------------------
```

One point that you might find useful to note. The order of the parameters in this list is determined by moving through the model definition from beginning to end. Each time a new parameter name is encountered, it is added to the list. Looking at the model command above, you can now see how the order in the displayed output arose.

The last example in the preceding subsection, which has four separate coefficients on a *cost* variable could be specified using

```
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Model:  U(air)   = bc*invc+bt*invt+aa+cha*hinc+cga*gc /
                    U(train) = bc*invc+bt*invt+at+cht*hinc+cgt*gc  /
                    U(bus)   = bc*invc+bt*invt+ab+chb*hinc+cgb*gc /
                    U(car)   = bc*invc+bt*invt               +cgc *gc $
```

The estimates are

```
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient     Error         z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      BC|    -.04387**       .01713    -2.56  .0104     -.07744   -.01029
      BT|    -.00815***      .00242    -3.37  .0008     -.01289   -.00341
      AA|   -1.37474         .83837    -1.64  .1011    -3.01791    .26844
     CHA|     .00703         .01079      .65  .5145     -.01411    .02818
     CGA|     .03762**       .01677     2.24  .0248      .00476    .07048
      AT|    2.53157***      .60801     4.16  .0000     1.33990   3.72324
     CHT|    -.05097***      .01214    -4.20  .0000     -.07477   -.02717
     CGT|     .03349**       .01506     2.22  .0262      .00397    .06301
      AB|    1.17858         .73949     1.59  .1110     -.27080   2.62795
     CHB|    -.03339**       .01300    -2.57  .0102     -.05886   -.00792
     CGB|     .03456**       .01516     2.28  .0227      .00484    .06428
     CGC|     .03808**       .01524     2.50  .0125      .00821    .06795
--------+--------------------------------------------------------------------
```

E39.5.3 Shorthand Notations for Sets of Utility Functions

There are several shorthands which will allow you to make the model specification much more compact. If the utility functions for several alternatives are the same, you can group them in one definition. Thus,

```
; Model: U(air)  = b0 + bcost * gc   /
         U(car)  = b0 + bcost * gc   $
```

could be specified as **; Model: U(air, car) = b0 + bcost * gc $**

For the model we have been considering, i.e.,

; Choices = air,train,bus,car

all of the following are the same

```
; Model: U(air)    = b1 * ttme + bcost * gc   /
         U(train)  = b1 * ttme + bcost * gc   /
         U(bus)    = b1 * ttme + bcost * gc   /
         U(car)    = b1 * ttme + bcost * gc   $
```

and **; Model: U(air,train,bus,car) = b1 * ttme + bcost * gc $**
and **; Model: U(*) = b1 * ttme + bcost * gc $**
and **; Rhs = ttme, gc**

The last will use the variable names instead of the supplied parameter names for the two parameters, but the models will be the same.

E39.5.4 Alternative Specific Constants and Interactions

You can also specify alternative specific constants in this format, by using a special notation. When you have a **U(a1, a2, ..., aJ)** for *J* alternatives, then you may specify, instead of a single parameter, a list of parameters enclosed in pointed brackets, to signify interaction with choice specific constants. Thus, **<b1,b2,...,bL>** indicates *L* interactions with choice specific dummy variables. *L* may be any number up to the number of alternatives. Use a zero in any location in which the variable does not appear in the corresponding equation. For example,

```
; Choices = air,train,bus,car
; Model: U(air)    = ba + bcost * gc   /
         U(car)    = bc + bcost * gc   /
         U(bus)    =      bcost * gc   /
         U(train)  = bt + bcost * gc   $
```

could be specified as **; Model: U(air,car,bus,train) = <ba,bc,0,bt> + bcost * gc $**

NOTE: Within a < ... > construction, the correspondence between positions in the list is with the **U(... list ...)** list, *not* with the original **; Choices** list. Note these are different (deliberately) in the example above.

Note the considerable savings in notation. The same device may also be used in interactions with attributes. For example:

```
; Model: U(air)   = ba + bcprv * gc  /
         U(car)   = bc + bcprv * gc  /
         U(bus)   =      bcpub * gc  /
         U(train) = bt + bcpub * gc  $
```

There are two cost coefficients, but the variable *gc* is common. This entire model can be collapsed into the single specification

```
; Model: U(air,car,bus,train) = <ba,bc,0,bt> +
         <bcprv,bcprv,bcpub,bcpub> * gc $
```

Parameters inside the brackets need not all be different if you wish to impose equality constraints. The example above imposes the two equality constraints shown in the model specification.

E39.5.5 Equality Constraints

There is no requirement that parameters be unique across any specification. Equality constraints may be imposed anywhere in the model, just by using the same parameter name. For example, nothing precludes

```
; Model: U(air,car,bus,train) = <ba,bc,0,bt> +
         <ba,bc,bcpub,bcpub> * gc $
```

This forces two of the slope coefficients to equal the alternative specific constants. Expanded, this specification would be equivalent to

```
; Model: U(air)   = ba + ba    * gc  /
         U(car)   = bc + bc    * gc  /
         U(bus)   =      bcpub * gc  /
         U(train) = bt + bcpub * gc  $
```

E39.6 Starting and Fixed Values for Parameters

The default starting values for all slope parameters in the utility functions specified as above are 0.0. You may provide a starting value for any parameter defined in a utility equation by including the value in parentheses after the *first* occurrence of the parameter definition.

For example:

```
; Model: U(air)   = ba(.53) + bcprv(-1.25) * gc  /
         U(car)   = bc      + bcprv        * gc  /
         U(bus)   =           bcpub        * gc  /
         U(train) = bt(.04) + bcpub        * gc  $
```

Starting values of 0.53 for *ba*, -1.25 for *bcprv*, and 0.04 for *bt* are given. The other parameters, *bcpub* and *bc* both start at 0.0. Note that the starting value for *bcprv* is given with the first occurrence of this name in the model. It is not necessary to give additional starting values for *bcprv*; the first will suffice. (If a parameter name appears more than once in a model definition, one might inadvertently give different starting values for the definitions. For example, if the second line above were **U(car) = bc+bcprv(1.3)*gc/** then values of -1.25 and 1.3 are being given for the same parameter, *bcprv*. The *last* definition is the one that controls. Thus, in this example, the starting value for *bcprv* would be 1.3, not -1.25. Note that this is not meant to be an option that is used for any purpose. This is only meant to explain how this erroneous specification will be handled.)

In a multiple parameter specification, the same value is given to all parameters that appear in the specification. Thus, in our earlier example:

```
; Model:U(air,car,bus,train) = <ba,bc,0,bt> (1.27439) + bcost * gc
```

the three parameters, *ba*, *bc*, and *bt*, are all started at 1.27439.

E39.6.1 Fixed Values

Any parameter that appears in the model may be fixed at a given value, rather than estimated. This might be useful, for example, for testing hypotheses. To fix a parameter, use the setup described above as if you were providing a starting value. But, instead of enclosing the value in parentheses, enclose it in square brackets. For example, in the model above, the coefficient *bcost* might be fixed at 0.05 with the command

```
; Model: U(air,car,bus,train) = <ba,bc,0,bt> (1.27439) + bcost [0.05] * gc
```

The fixed value will appear in the model output with all of the other estimated results, with a notation that this coefficient has been fixed rather than estimated.

E39.6.2 Starting Values and Fixed Values from a Previous Model

Each time you fit a model with **CLOGIT**, the coefficients and the names that you gave them are stored permanently for later use. (This is separate from the coefficients saved for the **WALD** testing procedure discussed in Section R14.4.) You may reuse these coefficients in the current model by specifying starting or fixed values with a simple '**[ ]**' or '**( )**' with no specific values provided. For example,

bcost ( ) * gc

would instruct **CLOGIT** to examine the previous model that you fit. If you had used the name *bcost* for one of the coefficients, then the estimated value from that model would be used as the starting value for this model.

E39.7 Modeling Choice Strategy

In some occasions in survey data, particularly in stated preference experiments, respondents will indicate that they did not consider certain attributes among a set of attributes in making their choices. When this aspect of the data is known, it has been conventional to insert zeros for the attribute in the choice model, thereby removing that attribute from the utility function. However, in fact, that does not remove the attribute from the choice probability; it forces it to enter with a peculiar, possibly extreme value. Consider, for example, a price variable. If a respondent indicates that they ignored price in a choice, then setting the 'price' to zero in the choice set would force a peculiar value on the choice process. Hensher, Rose, and Greene (2005b) argued that if a respondent truly ignores an attribute in a choice situation, then what should be zero in the choice model is not the attribute, but its coefficient in the utility function. That restriction definitely removes the attribute from the choice consideration by taking it out of the model altogether.

Accommodating this idea requires, in essence, that there be a possibly different model for each respondent. That is, one with possibly different zero restrictions imposed for different individuals. **CLOGIT** allows you to automate precisely this formulation in all discrete choice models with a special data coding.

For respondents who ignore attributes (it must be known in the data), simply code the attribute with value -888 for this respondent.

With this data convention, the program automatically detects these values and adjusts the model accordingly. You do not have to add any other codes to any **CLOGIT** commands to signal this aspect of the data. The model output will contain a diagnostic box noting when this option is being used when **CLOGIT** finds these values in the data.

The following applies to this feature:

1. At least some respondents must actually consider the attribute. It cannot be omitted from the model for everyone.

2. In computing elasticities (see Section E40.2), if **; Means** is used, it may distort the means slightly. How much so depends on how many observations are in use and how often the attribute is ignored. No generalizations are possible.

3. In computing descriptive statistics with the **; Describe** option (see Section E38.5.2), this may distort the means because the -888 values are not skipped, they are changed to 0.0. Output will contain a warning to this effect if it is noticed.

E39.8 Generalized Maximum Entropy Estimator

The CLOGIT multinomial logit model may be estimated using the generalized maximum entropy estimator described in Section E37.8 for the MLOGIT model. The estimator is the same – the difference between there and here is only the constraint on the parameter vectors – there is only a single parameter vector in the CLOGIT model. The computations are identical; the only difference is the format of the data. The estimator is requested by adding

; GME

or **; GME = number of support points**

to the **CLOGIT** command. In the application below, we reestimate the model used in several examples, using GME instead of MLE. The MLE is shown at the end of the results for ease of comparison. The command would be

CLOGIT ; Lhs = mode ; Rhs = one,gc,ttme
; Choices = air,train,bus,car
; GME = 5 $

```
-----------------------------------------------------------------------------
Generalized Maximum Entropy LOGIT Estimator
Dependent variable               Choice
Log likelihood function     -1556.27248
Estimation based on N =    210, K =   5
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
      GC|    -.01014***      .00356    -2.85  .0044      -.01711    -.00316
    TTME|    -.09407***      .01002    -9.38  .0000      -.11371    -.07442
   A_AIR|    5.62289***      .63242     8.89  .0000      4.38337    6.86241
 A_TRAIN|    3.68504***      .41687     8.84  .0000      2.86800    4.50209
   A_BUS|    3.10729***      .43557     7.13  .0000      2.25360    3.96098
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
| Information Statistics for Conditional Logit Model fit by GME       |
| Number of support points =5. Weights in support scaled to 1/sqr(N) |
|                            M=Model MC=Constants Only   M0=No Model |
| Criterion Function     -1556.27248        -1635.80211   -2516.41511 |
| LR Statistic vs. MC      159.05926             .00000        .00000 |
| Degrees of Freedom         2.00000             .00000        .00000 |
| Prob. Value for LR          .00000             .00000        .00000 |
| Entropy for probs.       207.71575          283.75877     291.12182 |
| Normalized Entropy          .71350             .97471       1.00000 |
| Entropy Ratio Stat.      166.81214           14.72609        .00000 |
| Bayes Info Criterion    3133.93338         3292.99265    5054.21865 |
| BIC - BIC(no model)     1920.28527         1761.22600        .00000 |
| Pseudo R-squared            .04862             .00000        .00000 |
| Pct. Correct Prec.        70.47619           30.00000      25.00000 |
| Notes: Entropy computed as Sum(i)Sum(j)Pfit(i,j)*logPfit(i,j).      |
|        Normalized entropy is computed against M0.                   |
|        Entropy ratio statistic is computed against M0.              |
|        BIC = 2*criterion - log(N)*degrees of freedom.               |
|        If the model has only constants or if it has no constants,   |
|        the statistics reported here are not useable.                |
|        If choice sets vary in size, MC and M0 are inexact.          |
+----------------------------------------------------------------------+

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -199.97662
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|    -.01578***      .00438    -3.60  .0003     -.02437    -.00719
    TTME|    -.09709***      .01044    -9.30  .0000     -.11754    -.07664
   A_AIR|    5.77636***      .65592     8.81  .0000     4.49078    7.06193
 A_TRAIN|    3.92300***      .44199     8.88  .0000     3.05671    4.78929
   A_BUS|    3.21073***      .44965     7.14  .0000     2.32943    4.09204
--------+--------------------------------------------------------------------
```

E40: Post Estimation Results for Conditional Logit Models

E40.1 Introduction

This chapter completes the documentation of the conditional logit (CLOGIT) model with four post estimation calculations:

- Partial effects and elasticities
- Predictions of probabilities, utilities and several other variables,
- Specification testing for the IIA assumption
- Model simulation and examination of the effects of changing scenarios on market shares.

E40.2 Partial Effects and Elasticities

In the discrete choice model, the effect of a change in attribute '*k*' of alternative '*j*' on the probability that individual *i* would choose alternative '*m*' (where *m* may or may not equal *j*) is

$$\delta_{im}(k|j) = \partial \text{Prob}[y_i = m]/\partial x_i(k|j) = [\mathbf{1}(j = m) - P_{ij}]P_{im}\beta_k.$$

You can request a listing of the effects of a specific attribute on a specified set of outcomes with

; Effects: attribute [list of outcomes]

The outcomes listing defines the variables '*j*' in the definition above. The attribute is the '*k*th.' A calculated partial effect is then listed for all alternatives (i.e., all '*m*') in the model. You can request additional tables by separating additional specifications with slashes. For example:

; Effects: gc [car, train] / ttme [bus,train]

HINT: It may generate quite a lot of output if your model is large, but you can request an analysis of 'all' alternatives by using the wildcard, **attribute [*]**.

The table below is produced by

```
CLOGIT    ; Lhs  = mode ; Choices = air,train,bus,car
          ; Rhs  = invc,invt,gc
          ; Rh2 = one,hinc
          ; Effects: gc [*] $
```

```
Derivative wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR|   .0060   -.0020   -.0012   -.0028
   TRAIN|  -.0020    .0062   -.0018   -.0024
     BUS|  -.0012   -.0018    .0043   -.0013
     CAR|  -.0028   -.0024   -.0013    .0066
```

The effects are computed by averaging the individual specific results, so the report contains the average partial effects. Since the mean is computed over a sample of observations, we also report the standard deviation of the estimates.

As noted in the tables, the marginal effects are computed by averaging the individual sample observations. An alternative way to compute these is to use the sample means of the data, and compute the effects for this one hypothetical observation. Request this with

; Means

For the table above, the results would be as follows:

```
Derivative wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR|   .0073   -.0030   -.0014   -.0028
   TRAIN|  -.0030    .0076   -.0016   -.0031
     BUS|  -.0014   -.0016    .0044   -.0015
     CAR|  -.0028   -.0031   -.0015    .0073
```

Note that the changes are substantive. The literature is divided on this computation. Current practice favors the first (default) approach.

The results above are only the average partial effects. In order to obtain a full listing of the effects and an estimator of the sample variance, use

; Full

For the preceding, we obtain

```
+---------------------------------------------------+
| Derivative             averaged over observations.|
| Effects on probabilities of all choices in model: |
| * = Direct Derivative effect of the attribute.    |
+---------------------------------------------------+
--------------------------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in AIR
--------+-----------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AIR|     .00604***     .00017    36.54  .0000      .00572    .00637
   TRAIN|    -.00201***   .7814D-04  -25.69  .0000     -.00216   -.00185
     BUS|    -.00124***   .5504D-04  -22.48  .0000     -.00134   -.00113
     CAR|    -.00280***     .00014   -19.84  .0000     -.00307   -.00252
```

```
--------+-------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in TRAIN
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
     AIR|    -.00201***   .7814D-04   -25.69  .0000     -.00216   -.00185
   TRAIN|     .00618***      .00018    34.29  .0000      .00583    .00653
     BUS|    -.00175***   .9502D-04   -18.46  .0000     -.00194   -.00157
     CAR|    -.00242***   .9003D-04   -26.88  .0000     -.00260   -.00224
--------+-------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in BUS
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
     AIR|    -.00124***   .5504D-04   -22.48  .0000     -.00134   -.00113
   TRAIN|    -.00175***   .9502D-04   -18.46  .0000     -.00194   -.00157
     BUS|     .00433***   .9872D-04    43.88  .0000      .00414    .00453
     CAR|    -.00134***   .4473D-04   -29.99  .0000     -.00143   -.00125
--------+-------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in CAR
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
     AIR|    -.00280***      .00014   -19.84  .0000     -.00307   -.00252
   TRAIN|    -.00242***   .9003D-04   -26.88  .0000     -.00260   -.00224
     BUS|    -.00134***   .4473D-04   -29.99  .0000     -.00143   -.00125
     CAR|     .00656***      .00015    44.02  .0000      .00627    .00685
--------+-------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

The ‘standard errors’ in these results are computed as the sample standard deviations of the sample of observations on the derivatives. These are not identical to what would be obtained if the delta method were applied to the nonlinear function used to obtain the elasticities though they should be reasonably close.

E40.2.1 Elasticities

Rather than see the partial effects, you may want to see elasticities,

$$\eta_{im}(k|j) = \partial \log \mathrm{Prob}[y_i = m]/\partial \log x_i(k|j) = x_i(k|j)/P_{im} \times \delta_{im}(k|j)$$

$$= [\mathbf{1}(j = m) - P_{ij}]\, x_i(k|j)\beta_k$$

Notice that this is not a function of P_{im}. The implication is that all the cross elasticities are identical. This will be obvious in the results, as shown in the example below.

You may request elasticities instead of partial effects simply by changing the square brackets above to parentheses, as in

; Effects: attribute (list of outcomes)

The first set of results above would become as shown in the following table:

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR|  2.6002  -1.1293  -1.1293  -1.1293
   TRAIN| -1.2046   3.5259  -1.2046  -1.2046
     BUS|  -.5695   -.5695   3.6181   -.5695
     CAR|  -.8688   -.8688   -.8688   2.5979
```

With **; Full**, the expanded set of elasticities is produced.

```
+---------------------------------------------------+
| Elasticity             averaged over observations.|
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
+---------------------------------------------------+
-----------------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in AIR
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AIR|    2.60021***      .05667   45.89  .0000     2.48915   2.71128
   TRAIN|   -1.12927***      .06414  -17.61  .0000    -1.25498  -1.00356
     BUS|   -1.12927***      .06414  -17.61  .0000    -1.25498  -1.00356
     CAR|   -1.12927***      .06414  -17.61  .0000    -1.25498  -1.00356
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AIR|   -1.20461***      .05673  -21.23  .0000    -1.31580  -1.09343
   TRAIN|    3.52593***      .14909   23.65  .0000     3.23373   3.81813
     BUS|   -1.20461***      .05673  -21.23  .0000    -1.31580  -1.09343
     CAR|   -1.20461***      .05673  -21.23  .0000    -1.31580  -1.09343
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AIR|    -.56952***      .01973  -28.87  .0000     -.60818   -.53086
   TRAIN|    -.56952***      .01973  -28.87  .0000     -.60818   -.53086
     BUS|    3.61811***      .10298   35.13  .0000     3.41627   3.81995
     CAR|    -.56952***      .01973  -28.87  .0000     -.60818   -.53086
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in CAR
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AIR|    -.86881***      .03532  -24.59  .0000     -.93805   -.79958
   TRAIN|    -.86881***      .03532  -24.59  .0000     -.93805   -.79958
     BUS|    -.86881***      .03532  -24.59  .0000     -.93805   -.79958
     CAR|    2.59786***      .10768   24.13  .0000     2.38682   2.80891
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The force of the independence from irrelevant alternatives (IIA) assumption of the multinomial logit model can be seen in the identical cross elasticities in the tables above. The table also shows two other aspects of the model. First, the meaning of the raw coefficients in a multinomial logit model, all of sign, magnitude and significance, are ambiguous. It is always necessary to do some kind of post estimation such as this to determine the implications of the estimates. Second, in light of this, we can see that the particular model estimated must be misspecified. The estimates imply that as the generalized cost of each mode rises, it becomes more attractive. The *gc* coefficient has the 'wrong' sign.

Elasticities and partial effects in the CLOGIT model are computed by averaging the individual observations on these quantities. Observations receive equal weight ($1/n$) in the average. A problem can arise when computing elasticities in this fashion. If an observation in the sample has an extreme configuration of attributes for some reason, then the elasticity or marginal effect for that observation can be extremely large (up to 10,000,000 for some cases). With the simple weighting $w_i = 1/n$, regardless of the rest of the sample, this observation (or observations if it happens more than once), will cause the average to be huge, producing nonsense values. *LIMDEP* provides two alternative methods of computing marginal effects and elasticities:

1. If elasticities are computed just once at the sample means of the attributes, extreme values will almost surely be averaged out, and the end result will almost always be reasonable values. You can request this computation with

   ; Effects:... (as usual) ; Means

2. Some authors have advocated a probability weighted average scheme instead. This uses a weight which differs by alternative. The computation uses

   $$w(t,j) = \text{Estimated } P(t,j) \,/\, \Sigma_t \text{ Estimated } P(t,j)$$

   where t indexes individual observations and j indexes alternatives. By this construction, if an individual probability is very small, the resulting extreme value for the marginal effect is multiplied by a very small probability weight, which offsets the extreme value. This likewise produces reasonable values for elasticities in almost all cases. You can request this computation with

   ; Effects:... (as usual) ; Pwt

   This weighting scheme does cause a problem. In the simple discrete choice model, the elasticities are

   $$\eta_{im}(k|j) = \partial \log \text{Prob}[y_i = m]/\partial \log x_i(k|j) = x_i(k|j)/P_{im}\times\delta_{im}(k|j)$$

   which means that the cross elasticity of change in probability j when the x in the attributes for choice m changes is the same for all of the alternatives. (E.g., the elasticity of the probabilities of alternatives 2,3,... with respect to changes in $x(k)$ in the attributes of alternative 1 are all equal to $\beta_k P(1)x(1,k)$. This will be true for individual observations. But, when probability weights are used, this will not be true for the weighted averages. It is true for the unweighted averages. The implication will be that the elasticities computed with **; Pwt** will suggest that the IIA property of the model has been relaxed. But, it has not. This is a result of the way the elasticity is computed. The IIA property of the model remains. The following shows the comparison of using **; Pwt** to the unweighted case for our example.

(Probability weighted)

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+------------------------------------
     AIR|  2.3722   -.7268   -.9638  -1.0659
   TRAIN|  -.9844   2.4338  -1.3509   -.9442
     BUS|  -.5596   -.6035   3.3527   -.5102
     CAR| -1.0170   -.6356   -.7857   2.0780
```

(Unweighted)

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+------------------------------------
     AIR|  2.6002  -1.1293  -1.1293  -1.1293
   TRAIN| -1.2046   3.5259  -1.2046  -1.2046
     BUS|  -.5695   -.5695   3.6181   -.5695
     CAR|  -.8688   -.8688   -.8688   2.5979
```

E40.2.2 Saving Elasticities in the Data Set

You can save the individual estimates of the own and cross elasticities as a variable in the data set by using

; Effects: attribute(alternative) = variable

This must provide the name of a specific attribute and a specific alternative. Only one variable may be saved by the model command. The following extends our earlier example by saving the elasticities with respect to the generalized cost of air. This saves as a variable the estimates that are averaged to produce the first row of the table of unweighted elasticities above. The table of descriptive statistics confirms the computations. Figure E40.1 shows the first few observations in the data area.

```
CLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
           ; Rhs = invc,invt,gc ; Rh2 = one,hinc
           ; Effects: gc(air) = gcair $
CREATE     ; alt = Trn(-4,0) $
DSTAT      ; Rhs = gcair ; Str = alt $
```

```
-------------------------------------------------------------------------
Descriptive Statistics for GCAIR
Stratification is based on ALT
-----------------+-------------------------------------------------------
Subsample        |        Mean     Std.Dev.    Cases  Sum of wts  Missing
-----------------+-------------------------------------------------------
ALT         =  1 |    2.600215      .823141      210      210.00        0
ALT         =  2 |   -1.129273      .931694      210      210.00        0
ALT         =  3 |   -1.129273      .931694      210      210.00        0
ALT         =  4 |   -1.129273      .931694      210      210.00        0
Full Sample      |    -.196901     1.851636      840      840.00        0
-----------------+-------------------------------------------------------
```

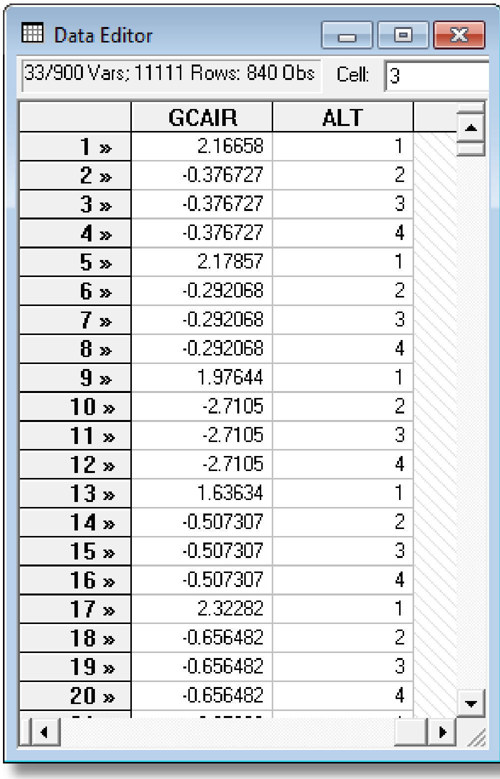


Figure E40.1 Estimated Elasticities

E40.2.3 Exporting Results in a Spreadsheet

Model results and estimated partial effects or elasticities may be exported to a spreadsheet file. Before doing this, you must open the export file with

OPEN ; Export = filename $

The file will be written in the generic .csv format, so you should open the file with a .csv extension, for example

OPEN ; Export = “C:\workspace\elasticities.csv” $

The request to export the results is then done by adding

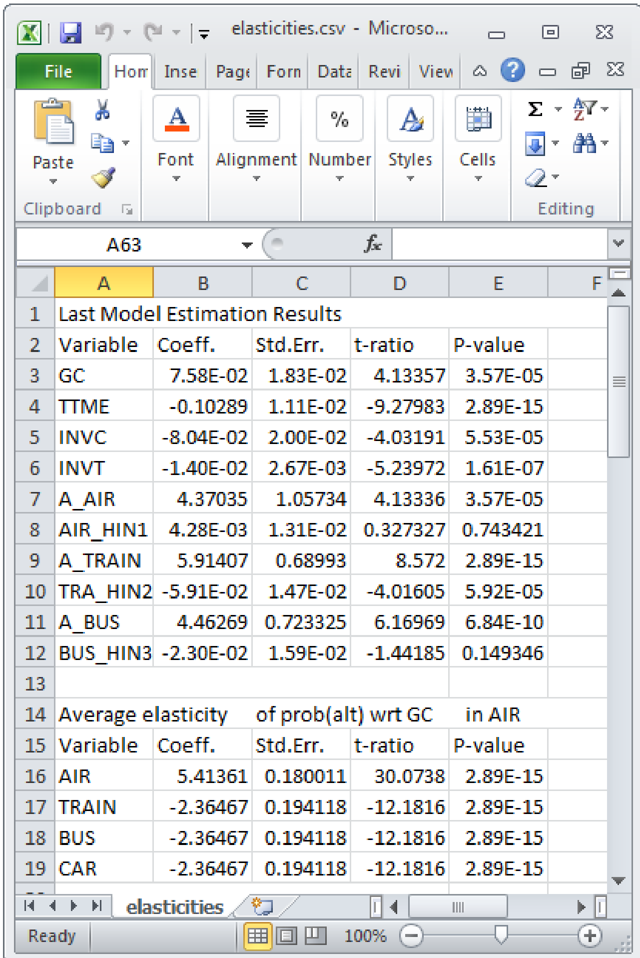
; Export = table

to your model command. Once the export file is open, you can use it for a sequence of models.

The spreadsheet file below was created with this sequence of commands:

```
OPEN     ; Export = "C:\ … \elasticities.csv" $
CLOGIT   ; Lhs = mode ; Choices = air,train,bus,car
         ; Rhs = gc,ttme,invc,invt ; Rh2 = one,hinc
         ; Export output
         ; Export = table
         ; Effects: gc(*),ttme(*) ; Full $
```

The **; Export output** setting requests that the model estimates also be included in the export file. This is followed by the tables of elasticities. The figure shows the results after the file has been read into *Excel*.



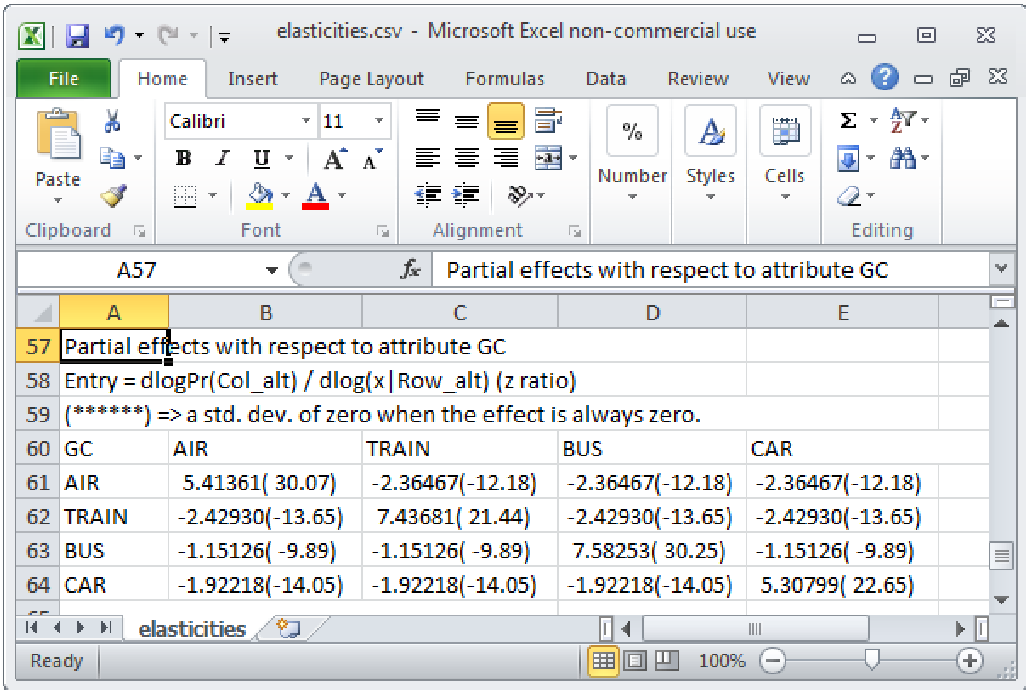
	A	B	C	D	E
1	Last Model Estimation Results				
2	Variable	Coeff.	Std.Err.	t-ratio	P-value
3	GC	7.58E-02	1.83E-02	4.13357	3.57E-05
4	TTME	-0.10289	1.11E-02	-9.27983	2.89E-15
5	INVC	-8.04E-02	2.00E-02	-4.03191	5.53E-05
6	INVT	-1.40E-02	2.67E-03	-5.23972	1.61E-07
7	A_AIR	4.37035	1.05734	4.13336	3.57E-05
8	AIR_HIN1	4.28E-03	1.31E-02	0.327327	0.743421
9	A_TRAIN	5.91407	0.68993	8.572	2.89E-15
10	TRA_HIN2	-5.91E-02	1.47E-02	-4.01605	5.92E-05
11	A_BUS	4.46269	0.723325	6.16969	6.84E-10
12	BUS_HIN3	-2.30E-02	1.59E-02	-1.44185	0.149346
13					
14	Average elasticity		of prob(alt) wrt GC		in AIR
15	Variable	Coeff.	Std.Err.	t-ratio	P-value
16	AIR	5.41361	0.180011	30.0738	2.89E-15
17	TRAIN	-2.36467	0.194118	-12.1816	2.89E-15
18	BUS	-2.36467	0.194118	-12.1816	2.89E-15
19	CAR	-2.36467	0.194118	-12.1816	2.89E-15

Figure E40.2 Exported Model Results and Elasticities

The exported results are in the form of the standard statistical table for estimated parameters. The format of the results in the .csv file may be changed to a matrix format by using

; Export = matrix

instead. Figure E40.3 shows the effect on the table shown in Figure E40.2.



	A	B	C	D	E
57	Partial effects with respect to attribute GC				
58	Entry = dlogPr(Col_alt) / dlog(x|Row_alt) (z ratio)				
59	(******) => a std. dev. of zero when the effect is always zero.				
60	GC	AIR	TRAIN	BUS	CAR
61	AIR	5.41361(30.07)	-2.36467(-12.18)	-2.36467(-12.18)	-2.36467(-12.18)
62	TRAIN	-2.42930(-13.65)	7.43681(21.44)	-2.42930(-13.65)	-2.42930(-13.65)
63	BUS	-1.15126(-9.89)	-1.15126(-9.89)	7.58253(30.25)	-1.15126(-9.89)
64	CAR	-1.92218(-14.05)	-1.92218(-14.05)	-1.92218(-14.05)	5.30799(22.65)

Figure E40.3 Exported Elasticities in Matrix Format

HINT: The export file is created while the computations are being done. However, there is a delay between when results are computed (by *LIMDEP*) and when they arrive in the file (by *Windows*). You should not try to open the export file (for example in *Excel*) while *LIMDEP* is still creating it. The results will be incomplete. Open the export file after you exit *LIMDEP*. Also, you should not try to write to an export file from *LIMDEP* while it is open by another program, such as *Excel*. This will cause a write error. You cannot modify with another program a spreadsheet file that *Excel* is using.

E40.3 Predicted Probabilities and Logsums (Inclusive Values)

There are several variables in addition to the elasticities that you can save in the data area while they are created by **CLOGIT**.

E40.3.1 Fitted Probabilities

There are some models which make use of the predicted probabilities from the discrete choice model. See, for example, Lee (1983). Or, you may have some other use for them. You can compute a column of predicted probabilities for the discrete choice model. Each 'observation' consists of J_i rows of data, where the number of choices may be fixed or variable. Use the command

CLOGIT ; Lhs = ... ; ...
; Prob = name $

The variable *name* will contain the predicted probabilities. The probabilities will sum to 1.0 for each observation, that is, *down* each set of J_i choices. The **; Prob** option will put the probabilities in the right places in your data set regardless of the setting of the current sample. For example, if you happen to be estimating a model after having **REJECTED** some observations, the predictions will be placed with the outcomes for the observations actually used. Unused rows of the data matrix are left undefined.

If your model has 14 or fewer choices, you can also include

; List

in your command to request a listing of the predicted probabilities. These will be listed a full observation at a time, rowwise, with an indicator of the choice that was made by that individual. For example, the first 10 observations (individuals) in the sample for the model above are

CLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = gc,ttme,invc,invt ; Rh2 = one ; Rh2 = hinc
; List $

```
PREDICTED PROBABILITIES (* marks actual, + marks prediction.)
Indiv    AIR       TRAIN     BUS       CAR
    1   .0918     .1574     .1124     .6384*+
    2   .1110     .1481     .0790     .6618*+
    3   .4621 +   .1106     .0953     .3320*
    4   .2112     .2639     .1240     .4008*+
    5   .1976     .2711     .1379     .3935*+
    6   .0901     .1306*    .1181     .6612 +
    7   .8128*+   .0462     .0392     .1018
    8   .3101     .0908     .0868     .5123*+
    9   .1098     .1867     .1312     .5724*+
   10   .1892     .2881     .1840     .3387*+
```

The '+' and '*' indicate the actual and predicted choices, respectively. Where these mark the same probability, the model predicted the outcome correctly. The predicted choice is the one that has the largest fitted probability

E40.3.2 Computing and Listing Model Probabilities

You can use an estimated model to compute (list and/or save) all probabilities, utilities, elasticities, and all descriptive statistics and crosstabulations for any specified set of observations, whether they were used in estimating the model or not. For example, this feature will allow you to compute predicted probabilities for a 'control' sample, to assess how well the model predicts outcomes for observations outside the estimation sample. To use this feature, use the following steps.

Step 1. Set up the full model for estimation, and estimate the model parameters.

Step 2. Reset the sample to specify the observations for which you wish to simulate the model.

Step 3. Use the *identical* **CLOGIT** command, but add the specification **; Prlist** to the command.

The sample that you specify at Step 2 may contain as many observations as you wish; it may be just one individual or it may be an altogether different set of data – as long as the variables match in name and form the variables in the original model.

NOTE: The observations in the new sample must be consistent with the specification of the model. The usual data checking is done to ensure this.

WARNING: You must not change the specification of the model between Steps 1 and 3. The coefficient vector produced by Step 1 is used for the simulation at Step 3. But it is not possible to check whether the coefficient vector used at Step 3 is actually the correct one for the model command used at Step 3. It will be if your model commands at Steps 1 and 3 are identical.

The following sequence fits the model in the preceding examples using the first 200 observations (800 data rows), then simulates the probabilities for the remaining 10 observations in the full sample:

```
SAMPLE      ; 1-800 $
CLOGIT      ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = invc,invt,gc,ttme ; Rh2 = one $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -174.83929
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
    MODE|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
    INVC|    -.08826***     .01987    -4.44  .0000     -.12721    -.04931
    INVT|    -.01344***     .00257    -5.23  .0000     -.01847    -.00841
      GC|     .07053***     .01778     3.97  .0001      .03568     .10539
    TTME|    -.10176***     .01117    -9.11  .0000     -.12366    -.07986
   A_AIR|    5.33347***     .92159     5.79  .0000     3.52720    7.13975
 A_TRAIN|    4.44686***     .52778     8.43  .0000     3.41244    5.48129
   A_BUS|    3.69334***     .52916     6.98  .0000     2.65620    4.73048
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The following commands produce an out of sample listing.

```
SAMPLE    ; 801-840 $
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc,ttme ; Rh2 = one
          ; Prlist $
```

```
+-----------------------------------------------+
| Discrete Choice (One Level) Model             |
| Model Simulation Using Previous Estimates     |
| Number of observations              10        |
+-----------------------------------------------+
PREDICTED PROBABILITIES (* marks actual, + marks prediction.)
Indiv    AIR       TRAIN     BUS       CAR
    1   .0543     .0445     .7540*+   .1472
    2   .2402     .2189     .2014     .3395*+
    3   .0137     .0885     .8571*+   .0406
    4   .0203     .0890     .8287*+   .0620
    5   .4058 +   .1092     .3745*    .1105
    6   .2766     .3248 +   .2785     .1201*
    7   .6129*+   .1446     .1240     .1185
    8   .0824     .5444 +   .0648*    .3084
    9   .1815     .3629 +   .1795     .2761*
   10   .1958     .1863     .0514     .5665*+
```

This arrangement of the model may also include

; Describe
; Show Model to display the model configuration
; Effects: desired elasticities or marginal effects
; Prob = name to save probabilities
; Ivb = name to save inclusive values

All of these computations are done for the current sample. This process is the same as the full model computations listed earlier. But, with **; Prlist** in place, the model estimated previously is used; it is not reestimated.

E40.3.3 Utilities and Inclusive Values

The utility functions used to compute the probabilities are

$$U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij}.$$

These may be saved in the data set as a new variable with the specification

; Utility = name

The *inclusive value*, or *log sum*, for the discrete choice model is

$$IV_i = \log \Sigma_j \exp(\boldsymbol{\beta}'\mathbf{x}_{i,j}).$$

Inclusive values are used for a number of purposes, including computing consumer surplus measures. You can keep the inclusive values for your model and data with the specification

; Ivb = name

The specification Ivb stands for 'inclusive value for branch.' Inclusive values are stored the same way that predicted probabilities are stored. Since each observation has only one inclusive value, the same value will be stored for all rows (choices) for the observation (person). An example is given below

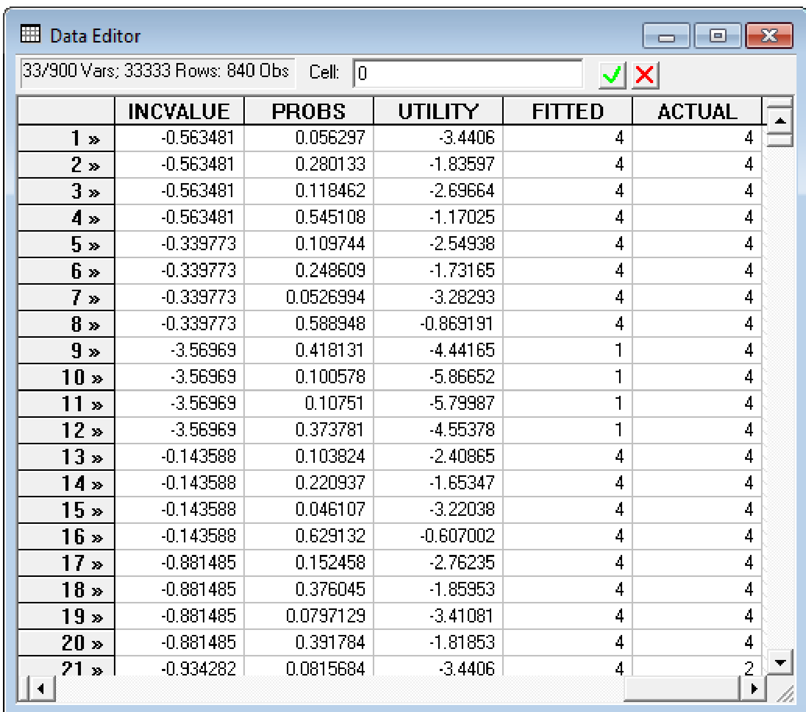
E40.3.4 Fitted Values of the Choice Variable

The actual and predicted outcomes for the model are saved with

; Fittedy = name and **; Actualy = name**

The actual value is the index of the choice actually made, repeated in each row of the choice set for the observation. The fitted value is the index of the alternative that has the largest probability based on the estimated model. The example below combines all of these features in a single command.

```
SAMPLE    ; All $
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc,ttme ; Rh2 = one
          ; Utility = utility ; Prob = probs ; Ivb = incvalue
          ; Actualy = actual ; Fittedy = fitted $
```



Data Editor

33/900 Vars; 33333 Rows: 840 Obs Cell: 0

	INCVALUE	PROBS	UTILITY	FITTED	ACTUAL
1 »	-0.563481	0.056297	-3.4406	4	4
2 »	-0.563481	0.280133	-1.83597	4	4
3 »	-0.563481	0.118462	-2.69664	4	4
4 »	-0.563481	0.545108	-1.17025	4	4
5 »	-0.339773	0.109744	-2.54938	4	4
6 »	-0.339773	0.248609	-1.73165	4	4
7 »	-0.339773	0.0526994	-3.28293	4	4
8 »	-0.339773	0.588948	-0.869191	4	4
9 »	-3.56969	0.418131	-4.44165	1	4
10 »	-3.56969	0.100578	-5.86652	1	4
11 »	-3.56969	0.10751	-5.79987	1	4
12 »	-3.56969	0.373781	-4.55378	1	4
13 »	-0.143588	0.103824	-2.40865	4	4
14 »	-0.143588	0.220937	-1.65347	4	4
15 »	-0.143588	0.046107	-3.22038	4	4
16 »	-0.143588	0.629132	-0.607002	4	4
17 »	-0.881485	0.152458	-2.76235	4	4
18 »	-0.881485	0.376045	-1.85953	4	4
19 »	-0.881485	0.0797129	-3.41081	4	4
20 »	-0.881485	0.391784	-1.81853	4	4
21 »	-0.934282	0.0815684	-3.4406	4	2

Figure E40.4 Model Predictions

E40.4 Hypothesis and Specification Tests of IIA

We consider two types of hypothesis tests. The first is a specification test of the IID extreme value specification. The model assumptions induce the most prominent shortcoming of the multinomial logit model, the *independence from irrelevant alternatives* (IIA) property. The fact that the ratio of any two probabilities in the model involves only the utilities for those two alternatives produces a number of undesirable implications, including the striking pattern in the elasticities in the model shown earlier. We consider a test of the IIA assumption. The second part of this section considers more conventional hypothesis tests about the coefficients in the model.

E40.4.1 Testing the IIA Assumption

Hausman and McFadden (1984) proposed a specification test for this model to test the inherent assumption of the independence from irrelevant alternatives (IIA). (IIA is a consequence of the initial assumption that the stochastic terms in the utility functions are independent and extreme value distributed. Discussion may be found in standard texts on qualitative choice modeling, such as Hensher, Rose and Greene (2005a) and Greene (2011).) The procedure is, first, to estimate the model with all choices. The alternative specification is the model with a smaller set of choices. Thus, the model is estimated with this restricted set of alternatives and the same model specification. The set of observations is reduced to those in which one of the smaller set of choices is made. The test statistic is

$$q = [\mathbf{b}_r - \mathbf{b}_u]'[\mathbf{V}_r - \mathbf{V}_u]^{-1}[\mathbf{b}_r - \mathbf{b}_u]$$

where '*u*' and '*r*' indicate unrestricted and restricted (smaller choice set) models and **V** is an estimated variance matrix for the estimates. To use *LIMDEP* to carry out this test, it is necessary to estimate both models. In the second, it is necessary to drop the outcomes indicated. This is done with the

; Ias = list

specification. The list gives the names of the outcomes to be dropped. This procedure is automated as shown in the following example:

```
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc,ttme $
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Ias = car
          ; Rhs = invc,invt,gc,ttme $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -244.13419
Estimation based on N =    210, K =   4
Inf.Cr.AIC  =  496.268 AIC/N =    2.363
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .1396 .1341
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.02243         .01435    -1.56  .1181     -.05056    .00570
    INVT|    -.00634***      .00184    -3.45  .0006     -.00995   -.00274
      GC|     .03183**       .01373     2.32  .0204      .00492    .05874
    TTME|    -.03481***      .00469    -7.42  .0000     -.04401   -.02561
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+-------------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found  59 bad observations among     210 individuals. |
|You can use ;CheckData to get a list of these points. |
+-------------------------------------------------------+
Normal exit:   6 iterations. Status=0, F=    103.2012

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -103.20124
Estimation based on N =    151, K =   4
Inf.Cr.AIC  =  214.402 AIC/N =    1.420
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -159.0502  .3511 .3424
Response data are given as ind. choices
Number of obs.=   210, skipped   59 obs
Hausman test for IIA. Excluded choices are
CAR
ChiSqrd[ 4] =  51.9631, Pr(C>c) =  .000000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    INVC|    -.04642**       .02109    -2.20  .0277     -.08775   -.00508
    INVT|    -.00963***      .00271    -3.55  .0004     -.01495   -.00432
      GC|     .04116**       .01984     2.07  .0380      .00227    .08005
    TTME|    -.07939***      .00992    -8.01  .0000     -.09882   -.05996
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

In order to compute the coefficients in the restricted model, it is necessary to drop those observations that choose the omitted choice(s). In the example above, 59 observations were skipped. They are marked as bad data because with *car* excluded, no choice is made for those observations. As a consequence, the log likelihood functions are not comparable. The Hausman statistic is used to carry out the test. In the preceding example, the large value suggests that the IIA restriction should be rejected.

Note that you can carry out several tests with different subsets of the choices without refitting the benchmark model. Thus, in the example above, you could follow with a third model in which **; Ias = bus** instead of **car**.

There is a possibility that restricting the choice set can lead to a singularity. It is possible that when you drop one or more alternatives, some attribute will be constant among the remaining choices. Thus, you might induce the case in which there is a 'regressor' which is constant across the choices. In this case, *LIMDEP* will send up a diagnostic about a singular Hessian (it is). Hausman and McFadden (1984) suggest estimating the model with the smaller number of choice sets *and* a smaller number of attributes. There is no question of consistency, or omission of a relevant attribute, since if the attribute is always constant among the choices, variation in it is obviously not affecting the choice. After estimation, the subvector of the larger parameter vector in the first model can be measured against the parameter vector from the second model using the Hausman statistic given earlier. This possibility arises in the model with alternative specific constants, so it is going to be a common case. The examples below suggest one way you might proceed in such as case.

The first step is to fit the original model using the entire sample and retrieve the results.

```
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc,ttme,one $
MATRIX    ; bu = b(1:4) ; vu = Varb(1:4,1:4) $
```

The variable choice takes values 1,2,3,4,1,2,3,4... indicating the indexing scheme for the choices.

```
CREATE    ; choice = Trn(-4,0) $
```

Chair is a dummy variable that equals one for all four rows when choice made is *air*. Now restrict the sample to the observations for choices *train, bus, car*.

```
REJECT    ; chair = 1 | choice = 1 $
```

Fit the model with the restricted sample (choice set) and one less constant term.

```
CLOGIT    ; Lhs = mode
          ; Choices = train,bus,car
          ; Rhs = invc,invt,gc,ttme,one $
```

Retrieve the restricted results and compute the Hausman statistic.

```
MATRIX      ; br = b(1:4) ; vr = Varb(1:4,1:4)
            ; db = br - bu ; vdb = Nvsm(vr,-vu) $
CALC        ; List
            ; q = Qfr(db,vdb)
            ; 1 - Chi(q,4) $
```

The results are:

```
[CALC] Q       =     33.7844338
[CALC] *Result*=       .0000008
Calculator: Computed   2 scalar results
```

NOTE: (We've been asked this one several times.) The difference matrix in this calculation, *vdb*, might be nonsingular (have an inverse), but not be positive definite. In such a case, the chi squared can be negative. If this happens, the right conclusion is probably that it should be zero.

E40.4.2 Lagrange Multiplier, Wald, and Likelihood Ratio Tests

CLOGIT keeps the usual statistics for the classical hypothesis tests. After estimation, the matrices *b* and *varb* will be kept and can be further manipulated for any purposes, for example, in the **WALD** command. You can use

; Test: ... restrictions

as well within the **CLOGIT** command to set up Wald tests of linear restrictions on the parameters. Likelihood ratio tests can be carried out by using the scalar *logl*, which will be available after estimation. The value of the log likelihood function for a model which contains only *J*-1 alternative specific constants will be reported in the output as well (see the sample outputs above). If your model actually contains the ASCs, *LIMDEP* will also report the chi squared test statistic and its significance level for the hypothesis that the other coefficients in the model are all 0.0.

HINT: *LIMDEP* can detect that a model contains a set of ASCs if you have used *one* in an **; Rhs** specification. But, it cannot determine from a set of dummy variables that you, yourself, provide, if they are a set of ASCs, because it inspects the model, not the data, to make the determination. As such, there is an advantage, when possible, to letting *LIMDEP* set up the set of alternative specific constants for you.

Finally, an LM statistic for testing the hypothesis that the starting values are not significantly different from the MLEs (the standard LM test) is requested by adding

; Maxit = 0

to the **CLOGIT** command.

E40.5 Examining Scenarios and Model Simulations

Another way to analyze the estimated model is to examine the effect on predicted ‘market’ shares of changes in the attribute levels. We compute the shares as

$$S(\textit{alternative } j) = N \times \sum_{i=1}^{N} \hat{P}_{ij}$$

Thus, save for the rounding error which is distributed, the model predicts the number of individuals in the sample who will choose each alternative. The crosstab described earlier summarizes this calculation. For our application,

```
CLOGIT    ; Lhs  = mode
          ; Choices = air,train,bus,car
          ; Rhs  = invc,invt,gc,ttme
          ; Rh2 = one,hinc
          ; Crosstab $
```

```
+--------------------------------------------------------+
| Cross tabulation of actual choice vs. predicted P(j)  |
| Row indicator is actual, column is predicted.          |
| Predicted total is F(k,j,i)=Sum(i=1,...,N) P(k,j,i).  |
| Column totals may be subject to rounding error.        |
+--------------------------------------------------------+
--------+------------------------------------------------------------------------
NLOGIT Cross Tabulation for 4 outcome Multinomial Choice Model
CrossTab|          AIR          TRAIN            BUS            CAR          Total
--------+------------------------------------------------------------------------
     AIR|            7             13             18              3             42
   TRAIN|            3             19             10              2             34
     BUS|            5             11             24              2             42
     CAR|            6             10             14              4             34
--------+------------------------------------------------------------------------
   Total|           21             53             66             12            152
+--------------------------------------------------------+
| Cross tabulation of actual y(ij) vs. predicted y(ij)  |
| Row indicator is actual, column is predicted.          |
| Predicted total is N(k,j,i)=Sum(i=1,...,N) Y(k,j,i).  |
| Predicted y(ij)=1 is the j with largest probability.  |
+--------------------------------------------------------+
--------+------------------------------------------------------------------------
NLOGIT Cross Tabulation for 4 outcome Multinomial Choice Model
CrossTab|          AIR          TRAIN            BUS            CAR          Total
--------+------------------------------------------------------------------------
     AIR|            5             10             27              0             42
   TRAIN|            1             27              4              2             34
     BUS|            4              7             29              2             42
     CAR|            5             10             18              1             34
--------+------------------------------------------------------------------------
   Total|           15             54             78              5            152
```

The feature described here is used to examine what becomes of these predictions when the value of an attribute changes. For example, how the predictions change when the generalized cost of air travel changes.

The simulator is used as follows:

Step 1. Fit the model.

Step 2. Use the identical model specification, but add to the command

; Simulation [= a subset of the choices, if desired – see below]
; Scenario = what changes and how

We take the base case first, in which all alternatives are considered in the simulation. A scenario is defined using

; Scenario: attribute (choices in which it appears) = the change

The change is defined using

= specific *value* to force the attribute to take this value in all cases
or = [*] *value* to multiply observed values by the value
or = [+] *value* to add '*value*' to the observed values.

The results of the computation will show the market shares before and after the change.

For example, we will refit our transport mode model, then examine the effect of increasing by 25% the terminal time spent waiting for air transport.

```
SAMPLE    ; 1-840 $
CLOGIT    ; Lhs = mode ; Rhs = one,gc,ttme
          ; Choices = air,train,bus,car $
CLOGIT    ; Lhs = mode ; Rhs = one,gc,ttme
          ; Choices = air,train,bus,car
          ; Simulation ; Scenario: ttme (air) = [*]1.25 $
```

Results are shown below.

```
+----------------------------------------------+
| Discrete Choice (One Level) Model            |
| Model Simulation Using Previous Estimates    |
| Number of observations              210      |
+----------------------------------------------+
+-------------------------------------------------------+
|Simulations of Probability Model                       |
|Model: Discrete Choice (One Level) Model               |
|Simulated choice set may be a subset of the choices.   |
|Number of individuals is the probability times the     |
|number of observations in the simulated sample.        |
|Column totals may be affected by rounding error.       |
|The model used was simulated with    210 observations.|
+-------------------------------------------------------+
```

```
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected           Change type            Value
---------  ------------------------------  ------------------- ---------
TTME       AIR                             Scale base by value     1.250
-------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 27.619    58 | 15.118    32 |-12.501%      -26 |
|TRAIN     | 30.000    63 | 33.694    71 |  3.694%        8 |
|BUS       | 14.286    30 | 16.126    34 |  1.841%        4 |
|CAR       | 28.095    59 | 35.061    74 |  6.966%       15 |
|Total     |100.000   210 |100.000   211 |   .000%        1 |
+----------+--------------+--------------+------------------+
```

The model predicts the base case using the actual data, shown in the left side and what would become of this case if the scenario is assumed. In this case, each person's *ttme* for *air* travel is increased by 25%, and the probabilities are recomputed. We see a fairly strong effect is predicted; 26 of 58 people who chose *air* are now expected to take other modes, eight changing to *train*, four to *bus*, and 15 to *car* (and one apparently deciding to walk – this is rounding error).

You may combine up to five scenarios in each simulation. This allows you to have simultaneous changes in attributes. Use

```
; Scenario: attribute (choices in which it appears) = the change /
            attribute (choices in which it appears) = the change /
  ...
```

For example, suppose terminal time for both *air* and *train* increased by 25%. We would extend our previous setup as follows:

```
SAMPLE      ; 1-840 $
CLOGIT      ; Lhs = mode ; Rhs = one,gc,ttme
            ; Choices = air,train,bus,car $
CLOGIT      ; Lhs = mode ; Rhs = one,gc,ttme
            ; Choices = air,train,bus,car
            ; Simulation ; Scenario:  ttme (air)     = [*] 1.25  /
                                      ttme (train)  = [*] 1.25  $
```

```
+-----------------------------------------------+
| Discrete Choice (One Level) Model             |
| Model Simulation Using Previous Estimates     |
| Number of observations              210       |
+-----------------------------------------------+
+--------------------------------------------------------+
|Simulations of Probability Model                        |
|Model: Discrete Choice (One Level) Model                |
|Simulated choice set may be a subset of the choices.    |
|Number of individuals is the probability times the      |
|number of observations in the simulated sample.         |
|Column totals may be affected by rounding error.        |
|The model used was simulated with    210 observations.|
+--------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type           Value
---------  -------------------------------  ------------------- ---------
TTME       AIR                              Scale base by value     1.250
TTME       TRAIN                            Scale base by value     1.250
-------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 27.619    58 | 16.417    34 |-11.202%      -24 |
|TRAIN     | 30.000    63 | 23.178    49 | -6.822%      -14 |
|BUS       | 14.286    30 | 18.796    39 |  4.510%        9 |
|CAR       | 28.095    59 | 41.609    87 | 13.514%       28 |
|Total     |100.000   210 |100.000   209 |   .000%       -1 |
+----------+--------------+--------------+------------------+
```

You may also compare the effects of different scenarios as well. For example, rather than assume that *ttme* for both *air* and *train* changed, you might compare the two scenarios. To do a pairwise comparison of scenarios, separate them with ‘&’ in the command. For example,

```
CLOGIT      ; Lhs = mode ; Rhs = one,gc,ttme
            ; Choices = air,train,bus,car
            ; Simulation ; Scenario: ttme (air)    = [*] 1.25 &
                                     ttme (train) = [*] 1.25  $
```

produces the following:

```
+--------------------------------------------+
| Discrete Choice (One Level) Model          |
| Model Simulation Using Previous Estimates  |
| Number of observations              210    |
+--------------------------------------------+
+------------------------------------------------------+
|Simulations of Probability Model                      |
|Model: Discrete Choice (One Level) Model              |
|Simulated choice set may be a subset of the choices.  |
|Number of individuals is the probability times the    |
|number of observations in the simulated sample.       |
|Column totals may be affected by rounding error.      |
|The model used was simulated with    210 observations.|
+------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
TTME       AIR                              Scale base by value     1.250
-------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 27.619    58 | 15.118    32 |-12.501%      -26 |
|TRAIN     | 30.000    63 | 33.694    71 |  3.694%        8 |
|BUS       | 14.286    30 | 16.126    34 |  1.841%        4 |
|CAR       | 28.095    59 | 35.061    74 |  6.966%       15 |
|Total     |100.000   210 |100.000   211 |   .000%        1 |
+----------+--------------+--------------+------------------+
-------------------------------------------------------------------------
Specification of scenario 2 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
TTME       TRAIN                            Scale base by value     1.250
-------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:

+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 27.619    58 | 30.168    63 |  2.548%        5 |
|TRAIN     | 30.000    63 | 20.787    44 | -9.213%      -19 |
|BUS       | 14.286    30 | 16.383    34 |  2.097%        4 |
|CAR       | 28.095    59 | 32.662    69 |  4.567%       10 |
|Total     |100.000   210 |100.000   210 |   .000%        0 |
+----------+--------------+--------------+------------------+
```

```
The simulator located    209 observations for this scenario.
Pairwise Comparisons of Specified Scenarios
Base     for this comparison is scenario 1.
Scenario for this comparison is scenario 2.
+----------+-------------+-------------+-----------------+
|Choice    |     Base    |   Scenario  | Scenario - Base |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+-------------+-------------+-----------------+
|AIR       | 15.118    32 | 30.168    63 | 15.049%       31 |
|TRAIN     | 33.694    71 | 20.787    44 |-12.907%      -27 |
|BUS       | 16.126    34 | 16.383    34 |   .257%        0 |
|CAR       | 35.061    74 | 32.662    69 | -2.399%       -5 |
|Total     |100.000   211 |100.000   210 |   .000%       -1 |
+----------+-------------+-------------+-----------------+
```

Simulations and scenarios can be combined and extended. You may have multiple scenarios and each scenario can involve several attributes. Separate the specifications within a scenario with slashes (/) and separate scenarios with ampersands (&). Finally, you can use the simulator to restrict the choice set. The computed probabilities are computed assuming only the specified alternatives are available. To do this, use

; Simulation = the subset of alternatives

To continue the example, we simulate the model assuming that people could not drive, and examine what the effect of increasing terminal time in airports would do to the market shares for the remaining three alternatives.

```
SAMPLE      ; 1-840 $
CLOGIT      ; Lhs = mode ; Rhs = one,gc,ttme
            ; Choices = air,train,bus,car $
CLOGIT      ; Lhs = mode
            ; Rhs = one,gc,ttme
            ; Choices = air,train,bus,car
            ; Simulation = air,train,bus
            ; Scenario: ttme (air) = [*] 1.25  $
```

```
+---------------------------------------------+
| Discrete Choice (One Level) Model           |
| Model Simulation Using Previous Estimates   |
| Number of observations              210     |
+---------------------------------------------+
+-------------------------------------------------------+
|Simulations of Probability Model                       |
|Model: Discrete Choice (One Level) Model               |
|Simulated choice set may be a subset of the choices.   |
|Number of individuals is the probability times the     |
|number of observations in the simulated sample.        |
|Column totals may be affected by rounding error.       |
|The model used was simulated with    210 observations.|
+-------------------------------------------------------+
```

```
------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
TTME       AIR                              Scale base by value     1.250
------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 39.353    83 | 22.933    48 |-16.420%      -35 |
|TRAIN     | 40.985    86 | 52.281   110 | 11.297%       24 |
|BUS       | 19.662    41 | 24.786    52 |  5.123%       11 |
|Total     |100.000   210 |100.000   210 |   .000%        0 |
+----------+--------------+--------------+------------------+
```

E41: Models for Count Data

E41.1 Introduction

This chapter and Chapters E42-E44 describe estimators for models for count data. Applications are discussed in Cameron and Trivedi (1986), Winkelmann (2008) and Hilbe (2011). Another important reference is Hausman, Hall, and Griliches (1984). Major surveys for practitioners are Winkelmann (2008), Cameron and Trivedi (1998, 2005) and Hilbe (2011).

The basic formulation is the *Poisson regression model*. For a discrete random variable, Y, observed over a period of length T_i, and observed frequencies, y_i, i =1,...,n, where y_i is a nonnegative integer count, and regressors $\mathbf{x}_i$, the Poisson regression model is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\exp(-T_i\lambda_i)(T_i\lambda_i)^{y_i}}{y_i!}, \; y_i = 0,1,...; \; \log \lambda_i = \boldsymbol{\beta}'\mathbf{x}_i.$$

In this model, λ_i is both the mean and variance of y_i *per unit of time*; that is

$$\text{E}(Y/T_i|\mathbf{x}_i) = \lambda_i.$$

The scale variable, T_i, might measure the size of a population observed, instead, as it would be if the model were one of the incidence of a disease in a set of locations. As long as the intensity variable, T_i is observed, the model may be conveniently defined in terms of

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \log T_i).$$

Then, we revert to the familiar linear index model, in which $\log T_i$ enters the regression with a coefficient of one. An example appears below. A multiplicative model is obtained if any of the components of $\mathbf{x}_i$ enter λ_i logarithmically. (For an application, see McCullagh and Nelder, (1983).) The partial effects in this nonlinear regression model are,

$$\frac{\partial E[y_i \,|\, \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i\boldsymbol{\beta}$$

The Poisson model has the restrictive equidispersion property that

$$\text{Var}[Y|\mathbf{x}_i] = E[Y|\mathbf{x}_i] = \lambda_i.$$

The *negative binomial regression model* is an extension of the Poisson regression model that allows the variance of the process to differ from the mean. An alternative interpretation that also fits well with several of the extensions considered in the next chapter is that the negative binomial model results from the introduction of a certain kind of unobserved individual heterogeneity into the Poisson regression model.

The probabilities in the negative binomial model are given by

$$\text{Prob}(\text{Y} = y_i|\mathbf{x}_i) \quad = \quad \frac{\theta^\theta \lambda_i^{y_i}}{\Gamma(\theta) y_i!} \frac{\Gamma(y_i + \theta)}{(\lambda_i + \theta)^{y_i + \theta}}$$

where θ is the overdispersion parameter. The connection between the two models is that the Poisson model results if $\alpha = 1/\theta = 0$. (A derivation appears in Section E41.4.5, the technical details section for the negative binomial model.) The formulation of the density that we use for optimization is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) \quad = \quad \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i + 1)} u_i^\theta (1 - u_i)^{y_i}$$

where $u_i \quad = \quad \theta / (\theta + \lambda_i)$

and $\theta \quad = \quad 1/\alpha.$

The negative binomial model has the property that

$$\text{Var}[y_i] \;=\; \text{E}[y_i]\{1 \;+\; \alpha\text{E}[y_i]\}.$$

This is a natural form of 'overdispersion' in that the overdispersion rate is

$$\text{Var}[y_i]/\text{E}[y_i] \;=\; 1 \;+\; \alpha\text{E}[y_i].$$

We have reparameterized the probability distribution in terms of θ because this simplifies the formulation and computation of the log likelihood and its derivatives. Greene (2008), defines the class of Negbin P models by a relationship between mean and variance functions,

$$E[y_i|\mathbf{x}_i] = \lambda_\text{i} \text{ and } \text{Var}[y_i|\mathbf{x}_i] = \lambda_i + \alpha\lambda_i^P.$$

The model already considered, the standard case, is Cameron and Trivedi's model Negbin 2, or NB2. An alternative form labeled Negbin 1 or NB1 is obtained by using $P = 1$. The density is obtained by replacing θ with $\theta\lambda_\text{i}$ in $\text{Prob}(Y = y_i|\mathbf{x}_i)$. More generally, replacing θ with $\theta\lambda_i^{2-P}$ produces the Negbin P family. For NB2, this produces, after a bit of manipulation,

$$\text{Prob}(Y = y_i|\mathbf{x}_i) \quad = \quad \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(\theta\lambda_i)\Gamma(y_i + 1)} w_i^{\theta\lambda_i} (1 - w_i)^{y_i}$$

where $w_i \quad = \quad \theta / (\theta + 1).$

This is not a simple reparameterization of the model; it is a different model. An example given in Section E41.4.4 demonstrates. We also consider the fully general form of the negative binomial model, NBP in Section E41.4.4.

We also provide many variants of each model. (The list of different functional forms that have been derived is surprisingly long. Hilbe (2011), for example, lists more than ten.) Several of these different formulations arise in a model of the sort that produces the negative binomial model, but in which the heterogeneity term derives from a normal distribution, rather than a log-gamma distribution. The resulting models are a bit simpler to estimate and appear to be more stable with respect to ill-behaved data sets. Other formulations arise through the effects of different sampling mechanisms, such as censoring, and other functional forms such as the gamma, generalized Poisson and Poly-Aeppli models.

Data for the count data models are often censored or truncated. The data are said to be censored if a range of values of the dependent variable is collapsed into a single value. Consider, for example, a survey which asks how many times an individual visited a certain facility. The responses might be 0, 1, 2, and 3 or more. Values above three are converted to three, so the data are censored. We allow censoring to be 'on the right,' as in our example, or 'on the left,' which would be the case if all values of y_i less than or equal to a certain value were converted to that value. Data are said to come from a truncated distribution, or be 'truncated,' (for convenience – it is the distribution, not the data that is truncated) if values in a certain range are simply not observed. To continue our example, if the analyst discarded observations with values of three or more, the remaining observations would come from a truncated distribution. The range of y_i for this example would be 0,1,2 instead of 0,1,2,... as in the original population. Another common application is 'on site sampling.' A visitor to a site of some sort, such as a recreation site, is asked how many times they have visited the site. By construction, on site samples are truncated at zero. Like censoring, we allow truncation to be on the right or the left.

This chapter will develop the various functional forms of the models for count data. Chapter E42 will document models that contain heterogeneity, censoring or truncation. Chapter E43 extends the Poisson and negative binomial models to two part formulations such as zero inflation, hurdle and sample selection models. Finally, Chapter E44 documents the panel data estimators.

E41.2 The Poisson Regression Model

The basic command for the Poisson regression model is

```
POISSON     ; Lhs = dependent variable
            ; Rhs = regressors
            ; ... other options $
```

The default model assumes that the time period or unit of space in which the outcome is observed is the same for all observations. When this is not the case, and the scaling is observed, use

```
            ; Exposure = scale variable
```

to provide it. An example appears below.

All of the general options for nonlinear models for controlling the iterative process and listing and keeping fitted values are available. These include:

; List	to display fitted values
; Keep = name	to retain predicted values
; Res = name	to retain residuals
; Maxit = n	to set maximum iterations or **; Maxit = 0** for LM tests
; Tlf, ; Tlb, ; Tlg	to set the convergence criteria
; Output = value	to control technical output during iterations
; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**
; Test: spec	to define Wald tests of linear restrictions

and so on. You may provide starting values and impose fixed values and restrictions in this model with

; Start = list	to give starting values
; Rst = list	to specify constraints
; CML: spec	to define a constrained maximum likelihood estimator

The coefficient vector is $\boldsymbol{\beta} = [\beta_1,\beta_2,...,\beta_K]$. *LIMDEP* uses zeros for the starting values for estimation. The estimated Hessian for the Poisson model is based on the actual second derivatives of the log likelihood. Partial effects are requested with

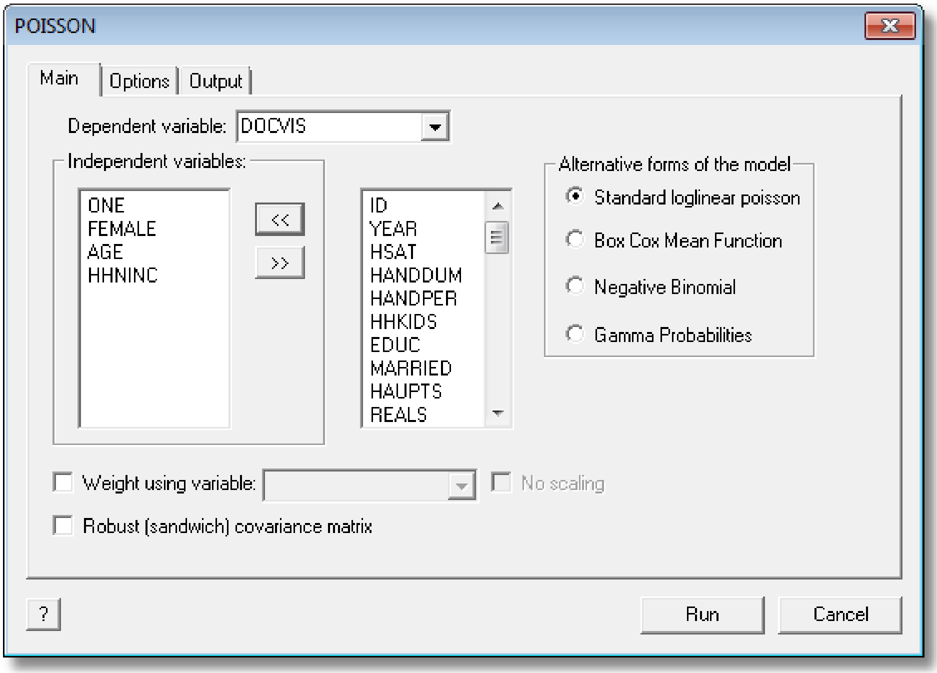
; Partial Effects

Partial effects are computed by averaging the individual estimates. A simpler estimator can be produced by doing the entire computation at the means of the data. Request this by adding

; Means

to the model command.

The command builder for this model can be found in Model:Count Data/Poisson. The basic model is specified on the Main and Options pages which are shown in Figure E41.1.



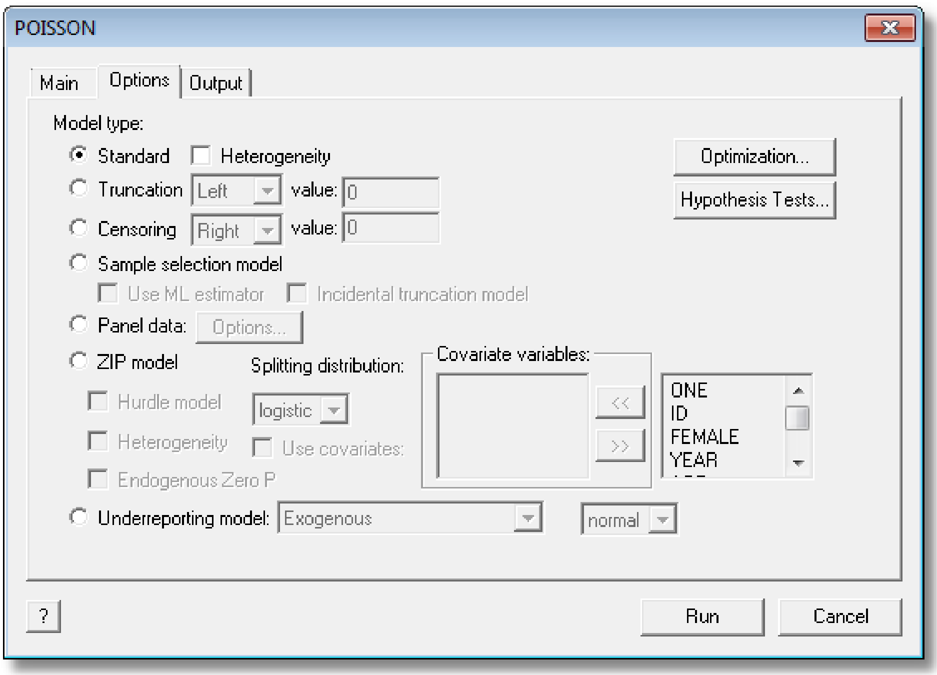


Figure E41.1 Command Builder for the Poisson Model

E41.2.1 Results for the Poisson Model

Estimation for the Poisson model begins with an ordinary least squares regression of the Lhs variable on the regressors. These results are presented only for comparison purposes, if you request them with

; OLS

and are not used as the starting values for the iterations. (Experience has shown clearly that **0** is a superior starting point for the iterations.) Perhaps a still better point would replace the starting value for the constant with the log of the mean of the Lhs variable. However, the model is so simple to estimate that is of little consequence. The model output consists of the standard results for maximum likelihood estimators, including the iterations, log likelihood function, restricted log likelihood function, and two goodness of fit statistics,

$$\text{Chi squared} = \Sigma_i(y_i - \hat{\lambda}_i)^2/\hat{\lambda}_i$$

and

$$\text{G squared} = 2\Sigma_i y_i \log(y_i/\hat{\lambda}_i) \text{ (with } 0\log(0) = 0))$$

(See Agresti (1984). Note, $y\log y = 0$ when $y = 0$.) Significance values are not computed for these because the degrees of freedom is dependent on the application. **CALC** can be used to compute the appropriate probability. We do, however, present the chi squared statistic for testing the hypothesis that the slopes are all zero, including the significance level and degrees of freedom. This computation assumes that there is a constant term in the model. (It is easily shown that in this case, the MLE of λ is $\bar{y}$, from which it follows that the MLE of β_0 is $\log\bar{y}$, and the remaining computations follow.)

The output for the Poisson model also contains two R^2 measures based on these fit measures,

$$R_p^2 = 1 - [\Sigma_i(y_i - \hat{\lambda}_i)^2/\hat{\lambda}_i] / [\Sigma_i(y_i - \bar{y})^2/\bar{y}] \quad \text{(p = Pearson)}$$

$$R_d^2 = 1 - [\Sigma_i y_i \log(y_i/\hat{\lambda}_i)] / [\Sigma_i y_i \log(y_i/\bar{y})] \quad \text{(d = deviance)}$$

In both cases, the fit measure assesses the improvement in the fit that results from using $\hat{\lambda}_i$ instead of $\bar{y}$ to predict y_i.

The **; List** specification produces a listing of:

1. actual y_i,
2. predicted y_i = estimate of $E[y_i] = \hat{\lambda}_i$,
3. residual = $y_i - \hat{\lambda}_i$,
4. ‘var1’ = estimate of $\boldsymbol{\beta}'\mathbf{x}$,
5. ‘var2’ = computed probability for observed y_i.

The partial effects in the Poisson (and negative binomial model) are

$$\partial E[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i = \lambda_i \boldsymbol{\beta}.$$

The **; Partial Effects** specification will produce a listing of these slopes computed at the sample means of the data.

Results saved for the Poisson model are:

Matrices: *b* and *varb* as usual

Scalars: *nreg*, *kreg*, *logl*, and *exitcode* for the model
ybar and *sy* = mean and standard deviation for dependent variable

Last Model: *b_variable*

Last Function: Conditional mean function, $\lambda = \exp(\boldsymbol{\beta}'\mathbf{x})$

The exponential regression function is used for **PARTIALS** and **SIMULATE**.

E41.2.2 Application of the Poisson Model

To illustrate the Poisson and related models, we will use the German health care data introduced in Section E2.4 and used in several earlier applications. The examples below will fit count data models to the count of doctor visits, *docvis*. Poisson regression of *docvis* on *one*, *age*, *hhninc* and *educ* produces the results below:

An example (developed further below) is the following:

```
SAMPLE      ; All $
NAMELIST    ; x = one,age,hhninc,educ,female $
POISSON     ; Lhs = docvis ; Rhs = x
            ; OLS ; Partial Effects $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=DOCVIS   Mean                 =       3.18352
             Standard deviation   =       5.68969
             No. of observations  =         27326  Degrees of freedom
Regression   Sum of Squares       =       30389.0          4
Residual     Sum of Squares       =       854192.      27321
Total        Sum of Squares       =       884581.      27325
             Standard error of e  =       5.59151
Fit          R-squared            =        .03435  R-bar squared =   .03421
Model test   F[  4, 27321]        =     242.99522  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =  -85806.29089  Akaike I.C.   =  3.44268
             Restricted (b=0)     =  -86283.92356  Bayes  I.C.   =  3.44419
             Chi squared [  4]    =     955.26534  Prob C2 > C2* =   .00000
Model was estimated on Jul 26, 2011 at 10:25:47 PM
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.29829***      .23965     5.42  .0000       .82858   1.76800
     AGE|     .06734***      .00304    22.18  .0000       .06139    .07329
  HHNINC|   -1.67038***      .19832    -8.42  .0000    -2.05908  -1.28169
    EDUC|    -.08058***      .01554    -5.19  .0000     -.11103   -.05013
  FEMALE|     .94932***      .06897    13.76  .0000       .81415   1.08449
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                DOCVIS
Log likelihood function   -103923.54929
Restricted log likelihood -108662.13583
Chi squared [   4 d.f.]      9477.17308
Significance level               .00000
McFadden Pseudo R-squared      .0436084
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  =********* AIC/N =    7.607
Model estimated: Jul 26, 2011, 22:25:47
Chi- squared =255750.59514  RsqP= .0796
G  - squared =154808.51777  RsqD= .0577
Overdispersion tests: g=mu(i)  : 21.372
Overdispersion tests: g=mu(i)^2: 21.373
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .57813***      .02630    21.98  .0000       .52659    .62968
     AGE|     .02057***      .00031    67.30  .0000       .01997    .02117
  HHNINC|    -.52855***      .02189   -24.14  .0000     -.57146   -.48565
    EDUC|    -.02868***      .00173   -16.57  .0000     -.03208   -.02529
  FEMALE|     .29405***      .00700    42.00  .0000       .28033    .30777
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1835
Scale Factor for Marginal Effects  3.1835
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .06550***      .00100    65.62  .0000       .06354    .06745
  HHNINC|   -1.68266***      .06992   -24.06  .0000    -1.81971  -1.54562
    EDUC|    -.09132***      .00552   -16.54  .0000     -.10214   -.08050
  FEMALE|     .93023***      .02210    42.10  .0000       .88693    .97354   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

To examine the use of the duration, or exposure variable, T_i discussed earlier, we consider a constructed example. As noted earlier, the data are an unbalanced panel. If we consider not the year by year count of doctor visits but the total number of visits by the individual, then the counts will differ partly because of the differing number of years. The following shows how to account for this effect. (We are also making use of the result that the sum of Poisson variables has a Poisson distribution.)

```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = one,age,hhninc,educ,female $
CREATE      ; sumy = Group Sums(docvis, Pds =_groupti) $
CREATE      ; sumy = Int(sumy + .1) ; date  = Ndx(id,1) $
REJECT      ; date > 1 $
POISSON     ; Lhs = sumy ; Rhs = x $
POISSON     ; Lhs = sumy ; Rhs = x ; Exposure = ti $
```

The results show that accounting for the length of exposure does, indeed, change the results noticeably.

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                  SUMY
Log likelihood function    -68730.19426
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    SUMY|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2.27241***     .02550    89.11  .0000     2.22243   2.32239
     AGE|     .02117***     .00028    76.32  .0000      .02062    .02171
  HHNINC|    -.81944***     .02509   -32.67  .0000     -.86861   -.77027
    EDUC|    -.04808***     .00177   -27.10  .0000     -.05155   -.04460
  FEMALE|     .21781***     .00698    31.22  .0000      .20414    .23148
--------+--------------------------------------------------------------------
Poisson Regression
Dependent variable                  SUMY
Log likelihood function    -56253.06502
Exposure variable for count data = TI
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    SUMY|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .67028***     .02597    25.81  .0000      .61939    .72117
     AGE|     .02021***     .00031    65.55  .0000      .01961    .02082
  HHNINC|    -.48585***     .02567   -18.93  .0000     -.53616   -.43555
    EDUC|    -.03318***     .00176   -18.85  .0000     -.03663   -.02973
  FEMALE|     .29277***     .00700    41.81  .0000      .27904    .30649
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E41.2.3 Testing for Overdispersion

Cameron and Trivedi (1990) have proposed a number of tests for over- or underdispersion in the Poisson regression model. Probably the simplest, and by their results, the optimal test in the set they considered, involves simple least squares regressions. The crux of the test is that under the hypothesis of the Poisson model, $(y - E[y])^2 - E[y]$ has mean 0. The testing framework is built around

$$H_0\text{: } \mathrm{Var}[y_i|\mathbf{x}_i] = \mu_i$$

$$\text{vs.} \quad H_1\text{: } \mathrm{Var}[y_i|\mathbf{x}_i] = \mu_i + \alpha g(\mu_i).$$

They detail the several assumptions needed to carry out the tests. Among them is the important one that under either hypothesis, the Poisson model gives consistent estimates of $E[y_i] = \mu_i$. The reader is referred to their paper for the necessary background. The test they propose, their T_{opt}, is carried out by testing the significance of the single coefficient in the linear OLS regression of

$$z_i = [(y_i - \mu_i)^2 - y_i] / (\mu_i\sqrt{2})$$

on

$$w_i = g(\mu_i) / (\mu_i\sqrt{2}).$$

They suggest two possibilities:

$$g(\mu_i) = \mu_i$$

and

$$g(\mu_i) = \mu_i^2.$$

Under the null hypothesis of equidispersion, the statistics have limiting chi squared distributions with one degree of freedom.

The two statistics are reported in the standard output for the Poisson model, as shown in the example below.

```
Poisson Regression
Dependent variable                    SUMY
Log likelihood function    -56253.06502
Restricted log likelihood  -74728.12052
Chi squared [    4 d.f.]    36950.11101
Significance level               .00000
McFadden Pseudo R-squared      .2472303
Estimation based on N =    7293, K =   5
Inf.Cr.AIC  =********* AIC/N =   15.428
Exposure variable for count data = TI
Chi- squared =123910.94497  RsqP= .3443
G  - squared = 88390.28781  RsqD= .2948
Overdispersion tests: g=mu(i)  : 14.829
Overdispersion tests: g=mu(i)^2: 15.415
```

Since the critical value from the chi squared table for one degree of freedom is 3.84, we would reject the null hypothesis on this basis, and proceed to a less restrictive model.

E41.2.4 Robust Covariance Matrices

The estimator of the asymptotic covariance matrix for the Poisson model based on the actual and expected (they are the same) second derivatives is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \left[-\frac{\partial^2 \log L}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'}\right]^{-1} = \left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X}]^{-1},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of predicted values. The BHHH estimator is the inverse of

$$\text{OPG} = \left[\sum_{i=1}^{n} \left(\frac{\partial \log P_i}{\partial\hat{\boldsymbol{\beta}}}\right)\left(\frac{\partial \log P_i}{\partial\hat{\boldsymbol{\beta}}}\right)'\right]^{-1} = \left[\sum_{i=1}^{n} (y_i - \hat{\lambda}_i')^2 \mathbf{x}_i\mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\mathbf{D}^2\mathbf{X}]^{-1}$$

where **D** is a diagonal matrix of residuals. The Poisson model is one in which the MLE is robust to certain misspecifications of the model, such as the failure to incorporate latent heterogeneity in the mean (i.e., one fits the Poisson model when the negative binomial is appropriate.) In this case, a robust covariance matrix,

$$\text{Robust Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = [\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{D}^2\mathbf{X}]\,[\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X}]^{-1}$$

is appropriate to accommodate this failure of the model. This computation is requested with

; Robust or ; HC2 (heteroscedasticity correction 2)

added to the command. For the model estimated earlier, the command produces the following results. The rather large increase in the standard errors produced by the robust estimator suggests that, indeed, there is something missing in the Poisson specification. As noted earlier, there is ample evidence of overdispersion in the data. The corrected results appear second.

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable              DOCVIS
Log likelihood function   -103923.54929
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .57813***      .02630    21.98  .0000      .52659     .62968
     AGE|      .02057***      .00031    67.30  .0000      .01997     .02117
  HHNINC|    -.52855***       .02189   -24.14  .0000    -.57146    -.48565
    EDUC|    -.02868***       .00173   -16.57  .0000    -.03208    -.02529
  FEMALE|      .29405***      .00700    42.00  .0000      .28033     .30777
--------+--------------------------------------------------------------------
Robust (sandwich) estimator used for VC
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .57813***      .08107     7.13  .0000      .41924     .73703
     AGE|      .02057***      .00095    21.72  .0000      .01872     .02243
  HHNINC|     -.52855***      .06580    -8.03  .0000     -.65752    -.39958
    EDUC|     -.02868***      .00484    -5.92  .0000     -.03817    -.01920
  FEMALE|      .29405***      .02240    13.13  .0000      .25014     .33796
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

A related calculation is used when observations occur in groups which may be correlated. This is rather like a panel; one might use this approach in a random effects kind of setting in which observations have a common latent heterogeneity. The parameter estimator is unchanged in this case, but an adjustment is made to the estimated asymptotic covariance matrix. The calculation is done as follows: Suppose the n observations are assembled in G clusters of observations, in which the number of observations in the ith cluster is n_i. Thus,

$$\sum_{i=1}^{G} n_i = n.$$

Denote by $\boldsymbol{\beta}$ the full set of model parameters in whatever variant of the model is being estimated. Let the observation specific gradients and Hessians be

$$\mathbf{g}_{ij} = \frac{\partial \log L_{ij}}{\partial \boldsymbol{\beta}}$$

$$\mathbf{H}_{ij} = \frac{\partial^2 \log L_{ij}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

The uncorrected estimator of the asymptotic covariance matrix based on the Hessian is

$$\mathbf{V}_H = -\mathbf{H}^{-1} = \left(-\sum_{i=1}^{G} \sum_{j=1}^{n_i} \mathbf{H}_{ij}\right)^{-1}.$$

The corrected asymptotic covariance matrix is

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{V}_H \left(\frac{G}{G-1}\right)\left[\sum_{i=1}^{G} \left(\sum_{j=1}^{n_i} \mathbf{g}_{ij}\right)\left(\sum_{j=1}^{n_i} \mathbf{g}_{ij}\right)'\right] \mathbf{V}_H.$$

Note that if there is exactly one observation per cluster, then this is $G/(G-1)$ times the sandwich (robust) estimator discussed above. Also, if you have fewer clusters than parameters, then this matrix is singular – it has rank equal to the minimum of G and JK, the number of parameters. This estimator is requested with

; Cluster = variable (as in panel data setups) or
number of observations in a cluster

(Further details on this estimator appear in Section R10.2) An extension for stratified and clustered (within strata) data may also be requested with

; Stratum = the specification

Since our data set is a panel, these results apply to the models estimated here. Using *id* as the clustering variable, we obtain the results below:

```
+---------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of  27326 observations contained   7293 clusters defined by |
| variable ID       which identifies by a value a cluster ID.         |
+---------------------------------------------------------------+
--------+-------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  DOCVIS|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
Constant|      .57813***      .11131     5.19  .0000       .35997    .79630
     AGE|      .02057***      .00129    15.97  .0000       .01805    .02310
  HHNINC|     -.52855***      .08191    -6.45  .0000      -.68908   -.36802
    EDUC|     -.02868***      .00688    -4.17  .0000      -.04218   -.01519
  FEMALE|      .29405***      .03212     9.15  .0000       .23109    .35701
--------+-------------------------------------------------------------
```

The continued increases in the standard errors compared to the results with the robust covariance matrix shown earlier suggest that the grouping of the observations is distorting the estimated covariance matrix.

E41.2.5 Scaling the Asymptotic Covariance Matrix MLE

In order to correct the Poisson estimator's asymptotic covariance matrix, a scale factor is suggested,

$$W = [1/(n\text{-}K)]\Sigma_i[(y_i - \exp(\boldsymbol{\beta}'\mathbf{x}_i))^2 / \exp(\boldsymbol{\beta}'\mathbf{x}_i).$$

This correction factor will account for over or underdispersion as well as degrees of freedom. To request this estimator, use

; HC1 (Heteroscedasticity correction 1)

An example appears below. In the example, the scaling makes a considerable difference in the estimated standard errors. In fact, for that model of doctor visits, there is an extreme preponderance of zeros so that the Lhs variable actually is considerably overdispersed. The two tests listed in the diagnostic statistics box shown earlier are consistent with this.

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	.57813***	.08046	7.19	.0000	.42043	.73583
AGE	.02057***	.00094	22.00	.0000	.01874	.02241
HHNINC	-.52855***	.06698	-7.89	.0000	-.65982	-.39728
EDUC	-.02868***	.00530	-5.42	.0000	-.03907	-.01830
FEMALE	.29405***	.02142	13.73	.0000	.25207	.33603

These estimated standard errors are based on the unscaled covariance matrix.

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	.57813***	.02630	21.98	.0000	.52659	.62968
AGE	.02057***	.00031	67.30	.0000	.01997	.02117
HHNINC	-.52855***	.02189	-24.14	.0000	-.57146	-.48565
EDUC	-.02868***	.00173	-16.57	.0000	-.03208	-.02529
FEMALE	.29405***	.00700	42.00	.0000	.28033	.30777

E41.2.6 Technical Details for the Poisson Model

The log likelihood and its derivatives for the Poisson regression model are:

$$\log L = \Sigma_i[-\lambda_i + \boldsymbol{\beta}'\mathbf{x}_i y_i - \ln y_i!],$$

$$\mathbf{g} = \Sigma_i \partial \log \text{Prob}[Y = y_i]/\partial\boldsymbol{\beta} = \Sigma_i(y_i - \lambda_i)\mathbf{x}_i$$

$$\mathbf{H} = \Sigma_i \partial^2 \log \text{Prob}[Y = y_i]/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \Sigma_i[-\lambda_i \mathbf{x}_i \mathbf{x}_i']$$

Estimation is by Newton's method,

$$\mathbf{b}_{k+1} = \mathbf{b}_k - [\mathbf{H}_k]^{-1}\mathbf{g}_k,$$

which converges readily. For this model, the iteration is equivalent to iteratively reweighted least squares,

$$\mathbf{b}_{k+1} = [\mathbf{X}'\boldsymbol{\Lambda}_k\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Lambda}_k\mathbf{z}],$$

where $\boldsymbol{\Lambda}_k$ is a diagonal matrix of fitted variances, λ_i, at iteration k, while

$$z_i = \log\lambda_i + (y_i - \lambda_i)/\lambda_i$$

based on the current parameter estimates.

E41.3 Quantile Regression for Count Data

The quantile regression estimator for count data was proposed by Machado and Silva (2005). The approach is not a Poisson model. Rather, the estimator develops conditional quantiles, $Q(y|\mathbf{x},\alpha)$ where α is the desired quantile of the distribution. The estimator uses a loglinear, i.e., exponential, predictor for the model. The linear programming methods are similar to those used for QREG for continuous data. A difference for the count data case is that the authors provide an analytic approach for estimating the asymptotic covariance matrix while bootstrapping is used in the continuous case. Methods used for computing this estimator are provided by Machado and Silva (2005).

The model is requested with

```
QCREG    ; Lhs = dependent variable
         ; Rhs = independent variables (including one) $
```

The default model is the conditional median, quantile = 0.5. Other quantiles are requested by adding

; Qnt = the quantile, strictly between 0.0 and 1.0.

You may produce results for multiple quantiles by specifying several quantiles in the **; Qnt** specification. For example, in our last application below, we use

; Qnt = .4, .5, .6, .7, .8.

The exponential function, λ_i is the conditional quantile here, not the conditional mean. Note that the count distribution is unlikely to be symmetric, so the conditional median will not equal the conditional mean in any event. Partial effects may be requested with

; Partial Effects.

The **PARTIALS** and **SIMULATE** commands may be used after estimation. However, it should be noted, once again, that these estimators are operating on the conditional quantile function, not the conditional mean.

We applied these to the health care data, and estimated the 0.4 quantile, the median (0.5) and the 0.75 quantiles. The Poisson model is compared to the conditional median.

```
QCREG    ; Lhs = docvis ; Rhs = one,age,educ,hhninc,female,public
         ; Qnt = .4 ; Partial Effects $
QCREG    ; Lhs = docvis ; Rhs = one,age,educ,hhninc,female,public
         ; Qnt = .5 ; Partial Effects $
POISSON  ; Lhs = docvis ; Rhs = one,age,educ,hhninc,female,public
         ; Partials Effects $
QCREG    ; Lhs = docvis ; Rhs = one,age,educ,hhninc,female,public
         ; Qnt = .75 ; Partial Effects $
```

```
--------------------------------------------------------------------------------
Quantile Regression Model. Quantile =       .400000
Quantile Regression Estimator for Count Data
LHS=DOCVIS   Mean                   =       3.18352
             Standard deviation     =       5.68959
             Number of observs.     =         27326
             Minimum                =        .00000
             t= .40000 quantile     =       1.00000
             Maximum                =     121.00000
Model size   Parameters             =             6
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
Constant|   -1.98244***      .13323   -14.88  .0000    -2.24358   -1.72131
     AGE|     .03206***      .00153    20.93  .0000      .02906     .03506
    EDUC|    -.01180         .00718    -1.64  .1004     -.02588     .00228
  HHNINC|    -.18265**       .09090    -2.01  .0445     -.36081    -.00448
  FEMALE|     .78127***      .03618    21.59  .0000      .71035     .85218
  PUBLIC|     .26660***      .05312     5.02  .0000      .16249     .37072
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

+----------------------------------------------------------+
| Partial Effects for Quantile Count Regression            |
| Variable     Value     Partial Effect    Semi-Elasticity |
|  AGE        43.527            .106               .029    |
|  EDUC       11.321           -.039              -.011    |
|  HHNINC       .352           -.605              -.163    |
| *FEMALE       .000           3.922              1.057    |
| *PUBLIC       .000           1.012               .273    |
+----------------------------------------------------------+
| * = Dummy variable. Other variables fixed at means.      |
+----------------------------------------------------------+
```

These are the partial effects produced by **PARTIALS**. They differ from the results above partly because they are treating λ_i as E[y|**x**] while the results are for the conditional median.

```
----------------------------------------------------------------------
Partial Effects for Exponential Regression Function
Partial Effects Computed at data Means
* ==> Partial Effect for a Binary Variable
----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect       Error      |t|  95% Confidence Interval
----------------------------------------------------------------------
      AGE          .02741      .00128    21.39       .02489       .02992
      EDUC        -.01540      .00608     2.53      -.02732      -.00349
      HHNINC      -.24722      .08152     3.03      -.40700      -.08744
      FEMALE       .63557      .03156    20.14       .57372       .69743
      PUBLIC       .24353      .04546     5.36       .15442       .33263
----------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Quantile Regression Model. Quantile =      .500000
Quantile Regression Estimator for Count Data
LHS=DOCVIS   Mean                 =        3.18352
             Standard deviation   =        5.68959
             Number of observs.   =          27326
             Minimum              =         .00000
             t= .50000 quantile   =        1.00000
             Maximum              =      121.00000
Model size   Parameters           =              6
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -1.38214***      .15582    -8.87  .0000     -1.68755  -1.07674
     AGE|     .03074***      .00151    20.30  .0000       .02777    .03371
    EDUC|    -.01953**       .00871    -2.24  .0249      -.03659   -.00247
  HHNINC|    -.21747**       .09591    -2.27  .0234      -.40546   -.02948
  FEMALE|     .65814***      .03582    18.37  .0000       .58793    .72836
  PUBLIC|     .33036***      .07036     4.70  .0000       .19245    .46826
--------+--------------------------------------------------------------------
(Poisson)
--------+--------------------------------------------------------------------
Constant|     .25069***      .03206     7.82  .0000       .18785    .31353
     AGE|     .02059***      .00031    67.42  .0000       .01999    .02119
    EDUC|    -.01983***      .00180   -11.03  .0000      -.02336   -.01631
  HHNINC|    -.48298***      .02194   -22.01  .0000      -.52598   -.43998
  FEMALE|     .29248***      .00700    41.80  .0000       .27877    .30619
  PUBLIC|     .23566***      .01330    17.71  .0000       .20959    .26174
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+----------------------------------------------------------+
| Partial Effects for Quantile Count Regression            |
| Variable     Value    Partial Effect    Semi-Elasticity |
|  AGE        43.527              .087               .026  |
|  EDUC       11.321             -.055              -.017  |
|  HHNINC       .352             -.615              -.185  |
| *FEMALE       .000             2.635               .791  |
| *PUBLIC       .000             1.108               .333  |
+----------------------------------------------------------+
| * = Dummy variable. Other variables fixed at means.      |
+----------------------------------------------------------+
(Poisson model)
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .06556***      .00100    65.73  .0000       .06360    .06751
    EDUC|    -.06314***      .00573   -11.02  .0000      -.07437   -.05191
  HHNINC|   -1.53758***      .07004   -21.95  .0000     -1.67486  -1.40030
  FEMALE|     .92522***      .02208    41.90  .0000       .88194    .96850   #
  PUBLIC|     .68233***      .03497    19.51  .0000       .61380    .75086   #
--------+--------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
Quantile Regression Model. Quantile =       .750000
Quantile Regression Estimator for Count Data
LHS=DOCVIS   Mean                  =       3.18352
             Standard deviation    =       5.68959
             Number of observs.    =         27326
             Minimum               =        .00000
             t= .75000 quantile    =       4.00000
             Maximum               =     121.00000
Model size   Parameters            =             6
--------+---------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error        z    |z|>Z*         Interval
--------+---------------------------------------------------------------------
Constant|     .21734*         .12905     1.68  .0922      -.03560     .47028
     AGE|     .02188***       .00134    16.36  .0000       .01926     .02450
    EDUC|    -.01569**        .00715    -2.20  .0281      -.02970    -.00168
  HHNINC|    -.30590***       .08688    -3.52  .0004      -.47619    -.13561
  FEMALE|     .38686***       .03140    12.32  .0000       .32532     .44839
  PUBLIC|     .19154***       .04929     3.89  .0001       .09494     .28814
--------+---------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
------------------------------------------------------------------------------
```

```
+----------------------------------------------------------+
| Partial Effects for Quantile Count Regression            |
| Variable    Value    Partial Effect    Semi-Elasticity |
|  AGE       43.527              .043               .016   |
|  EDUC      11.321             -.031              -.011   |
|  HHNINC      .352             -.596              -.221   |
| *FEMALE      .000              .920               .341   |
| *PUBLIC      .000              .411               .152   |
+----------------------------------------------------------+
| * = Dummy variable. Other variables fixed at means.      |
+----------------------------------------------------------+
```

The full set of results can be obtained for several quantiles with **; Qnt = list of values**. A summary table will also be produced. For our example, we obtained

```
------------------------------------------------------------------------
Quantile Regression Coefficients Summary
------------------------------------------------------------------------
Variable  b(0.40)  b(0.50)  b(0.60)  b(0.70)  b(0.80)
 ONE       -1.968   -1.345    -.610    -.029     .444
 AGE         .032     .030     .025     .022     .022
 EDUC       -.014    -.018    -.010    -.013    -.016
 HHNINC     -.287    -.224    -.312    -.372    -.397
 FEMALE      .767     .661     .490     .426     .354
 PUBLIC      .317     .298     .226     .198     .223
------------------------------------------------------------------------
```

The quantile estimator is estimated by perturbing the sample data slightly with random draws to make the data continuous. This creates some simulation 'chatter' (noise) in that the results are slightly dependent on the random draws. To reduce this outcome, the authors suggest averaging the results over several series of draws. The default in *LIMDEP*'s estimator is not to do this average – implicitly using m = 1 random sample and one estimator. You can use additional estimates by specifying m with

; Pts = m.

In the results below, we have fit the model with m = 1, then a second time, averaging the results over m = 5 repetitions.

CALC ; Ran(123457) $
QCREQ ; … $
CALC ; Ran(123457) $
QCREQ ; … ; Pts = 5 $

```
-----------------------------------------------------------------------------
Quantile Regression Model. Quantile =      .400000
Quantile Regression Estimator for Count Data
LHS=DOCVIS    Mean                =        3.18352
              Standard deviation  =        5.68959
              Number of observs.  =          27326
              Minimum             =         .00000
              t= .40000 quantile  =        1.00000
              Maximum             =      121.00000
Model size    Parameters          =              6
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -1.97688***      .13173   -15.01  .0000    -2.23507   -1.71869
     AGE|     .03247***      .00149    21.84  .0000      .02955     .03538
    EDUC|    -.01448**       .00709    -2.04  .0411     -.02838    -.00059
  HHNINC|    -.33054***      .09027    -3.66  .0003     -.50748    -.15361
  FEMALE|     .80459***      .03594    22.39  .0000      .73415     .87504
  PUBLIC|     .31003***      .05458     5.68  .0000      .20306     .41699
--------+--------------------------------------------------------------------
-----------------------------------------------------------------------------
Quantile Regression Model. Quantile =      .400000
Quantile Regression Estimator for Count Data
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -1.97692***      .13095   -15.10  .0000    -2.23358   -1.72027
     AGE|     .03236***      .00151    21.37  .0000      .02939     .03533
    EDUC|    -.01432**       .00718    -1.99  .0461     -.02840    -.00025
  HHNINC|    -.25299***      .09583    -2.64  .0083     -.44080    -.06517
  FEMALE|     .77923***      .03736    20.86  .0000      .70601     .85245
  PUBLIC|     .30690***      .05319     5.77  .0000      .20264     .41115
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E41.4 Overdispersion: The Negative Binomial Model

The negative binomial model has served as the most common extension of the Poisson model to allow for overdispersion or latent heterogeneity. We consider several other models as well an extension of the negative binomial model to allow individual variation in the overdispersion parameter and, in the next section, three models based on different functional forms that allow underdispersion as well.

E41.4.1 The Negative Binomial Model

The negative binomial model can be obtained by introducing heterogeneity into the conditional mean of the Poisson. Thus, if

$$f(y_i \mid \lambda_i, \varepsilon_i) = \text{Poisson with } \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i)$$

where

$$\exp(\varepsilon_i) = v_i \sim \text{Gamma with mean 1},$$

$$f(v_i) = \frac{\theta^\theta}{\Gamma(\theta)}\exp(-\theta v_i)v_i^{\theta-1},$$

then the unconditional density is

$$\log \text{Prob}[Y_i = j] = \log L_i$$

$$= \log \Gamma(\theta + y_i) - \log \Gamma(\theta) - \log[\Gamma(y_i+1)] + \theta \log u_i + y_i \log(1 - u_i),$$

where

$$\theta = 1/\alpha$$

and

$$u_i = \theta / (\theta + \lambda_i).$$

The crucial element of the result is that whereas in the Poisson model, $\text{Var}[y_i|\lambda_i] = E[y_i|\lambda_i]$, in the negative binomial model,

$$\text{Var}[y_i|\lambda_i] = E[y_i|\lambda_i] + \alpha\, E[y_i|\lambda_i] > E[y_i|\lambda_i];$$

the model has overdispersion.

The negative binomial regression model is requested by extending the Poisson model. Use

```
NEGBIN      ; Lhs  = ...
            ; Rhs  = ... $

or  POISSON ; Lhs  = ...
            ; Rhs  = ...
            ; Model = Negbin $
```

The full set of estimates for the Poisson model will be given first, followed by the negative binomial estimates. These can be compared for evidence of overdispersion. (The Poisson results will contain two regression based test statistics for the hypothesis of no overdispersion as well. See Section E41.2.3.) The negative binomial model is estimated using only the BFGS algorithm. All other parts of the basic command are identical to those for the Poisson model.

Starting values for the slopes are the Poisson regression parameters estimated earlier. To compute an initial estimate of the overdispersion parameter, α, *LIMDEP* computes the OLS slope in an artificial regression based on the relationship between the Poisson and negative binomial models,

$$[(y_i - \lambda_i)^2 / \lambda_i - 1] = \alpha\lambda_i + w_i.$$

(Certainly, w_i is heteroscedastic, but we are only interested in consistency.) If the resulting estimate is not positive, this suggests that the data are inconsistent with the model. But, *LIMDEP* then uses a value of α = .2, and continues. You may provide your own starting values, as well, with

; Start = slope parameters, value for 1/α

(Be sure to provide the last value.)

NOTE: If you wish to provide your own starting values for the negative binomial model, provide the K values for $\boldsymbol{\beta}$ and $\theta = 1/\alpha$, not α.

Fixed value and linear restrictions may be imposed with

; Rst = list

or **; CML: specification**

Once again, the list in the constraints specification must have a setting for $\theta = 1/\alpha$, either a fixed value or a parameter name.

NOTE: The restrictions are not imposed on the initial Poisson model when it is fit for starting values.

The command builder for the negative binomial model is found at Model:Count Data/NegBin. The dialog for the model specification and options for the model are identical to those for the Poisson model; the model command differs only in the command name.

The negative binomial model occasionally presents convergence problems in estimation, particularly when the data are censored or truncated. To deal with this, or for purposes of hypothesis testing or specification analysis, you may fix the value of α (not θ) with the specification,

; Dsp = value for α

The parameters of the negative binomial model will be estimated by maximum likelihood with α held fixed at this value. The value will be clearly marked as fixed in the final output.

The retrievable results for this model are:

Matrices: *b* and *varb* as usual
Adding **; Par** requests that the estimate of α (not θ) be included with **β** in *b* and *varb*.

Scalars: *nreg*, *kreg*, and *logl* for the model
ybar and *sy* = mean and standard deviation for dependent variable
alpha for the estimate of α for the negative binomial model

Last Function: Conditional mean function, $\lambda = \exp(\boldsymbol{\beta}'\mathbf{x})$

The exponential regression function is used for **PARTIALS** and **SIMULATE**.

E41.4.2 Application

The following refits the Poisson model estimated using, instead, the negative binomial specification. The base Poisson model is shown as well to allow a comparison. These results decisively reject the Poisson model in favor of the negative binomial. The reported results indicate that the NB2 form of the model has been used. It also shows the hypothesis test of the Poisson model as a restriction on the NB model. The hypothesis is decisively rejected by several tests.

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                 DOCVIS
Log likelihood function   -103923.54929
Chi- squared =255750.59514  RsqP= .0796
G  - squared =154808.51777  RsqD= .0577
Overdispersion tests: g=mu(i)  : 21.372
Overdispersion tests: g=mu(i)^2: 21.373
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .57813***      .02630    21.98  .0000      .52659    .62968
     AGE|     .02057***      .00031    67.30  .0000      .01997    .02117
  HHNINC|    -.52855***      .02189   -24.14  .0000     -.57146   -.48565
    EDUC|    -.02868***      .00173   -16.57  .0000     -.03208   -.02529
  FEMALE|     .29405***      .00700    42.00  .0000      .28033    .30777
--------+--------------------------------------------------------------------
Negative Binomial Regression
Log likelihood function    -60164.22014
Restricted log likelihood -103923.54929
Chi squared [   1 d.f.]     87518.65830
Significance level               .00000
NegBin form 2; Psi(i) = theta
Tests of Model Restrictions on Neg.Bin.
Model                Logl ChiSquared[df]
Poisson(b=0)  -108662.14  ******** [**]
Poisson       -103923.55    9477.2 [ 4]
Negative Bin.  -60164.22   87518.7 [ 1]
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
Constant|     .62857***      .05457    11.52  .0000       .52162    .73553
     AGE|     .02042***      .00070    29.07  .0000       .01904    .02179
  HHNINC|    -.48779***      .04520   -10.79  .0000     -.57637   -.39921
    EDUC|    -.03539***      .00378    -9.36  .0000     -.04281   -.02798
  FEMALE|     .32673***      .01588    20.58  .0000       .29561    .35784
        |Dispersion parameter for count data model
   Alpha|    1.90309***      .01984    95.94  .0000     1.86421   1.94197
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These estimates are for a negative binomial with the dispersion parameter forced to equal 1.5 with

NEGBIN ; Lhs = docvis ; Rhs = x ; Dsp = 1.5 $

```
-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function    -60375.82426
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
Constant|     .62664***      .04454    14.07  .0000       .53934    .71394
     AGE|     .02043***      .00058    35.48  .0000       .01930    .02156
  HHNINC|    -.49059***      .03736   -13.13  .0000     -.56381   -.41736
    EDUC|    -.03514***      .00312   -11.25  .0000     -.04126   -.02902
  FEMALE|     .32555***      .01298    25.09  .0000       .30011    .35099
        |Dispersion parameter for count data model
   Alpha|    1.50000    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------
```

Based on the model with free dispersion parameter, the likelihood ratio statistic for this restriction would be -2(-60164.22014.14 - (-60375.82426)) = 423.21. This is far larger than the critical chi squared with one degree of freedom of 3.84, so we would reject the hypothesis that α equals 1.5.

E41.4.3 Heterogeneous Negative Binomial Model

The negative binomial model may be extended to allow observed heterogeneity in the dispersion parameter. The structural model is

$$\text{Prob}[Y = y_i] = \frac{\Gamma(\theta_i + y_i)}{\Gamma(\theta_i)\Gamma(y_i + 1)} u_i^{\theta_i}(1 - u_i)^{y_i}$$

$$u_i = \theta_i / (\theta_i + \lambda_i)$$

$$\alpha_i = 1 / \theta_i = \alpha \exp(\boldsymbol{\delta}'\mathbf{z}_i)$$

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

The mean of the underlying gamma variate which produces the negative binomial model is one while its variance is α_i. Therefore, this extension is equivalent to allowing heteroscedasticity in the latent heterogeneity. The command for adding this specification to the model is

NEGBIN ; ... as before ; Hfn = variables in $\mathbf{z}_i$ $

NOTE: Do not include *one* in the Hfn list. The leading α provides the constant term.

The heterogeneity model retains all the options of the standard model, including fitted values, lists of predictions, saving results, etc. The new parameters, $\boldsymbol{\delta}$, are kept as a matrix named *deltanb*. The heterogeneity does not affect the conditional mean function, so the partial effects are still based on λ_i.

Note, that $\partial\log L_i/\partial\boldsymbol{\delta} = \partial\log L_i/\alpha_i \times \alpha_i \times \mathbf{z}_i$. This is a minor complication added to the model as already developed. Once again, the BHHH estimator is used for the asymptotic covariance matrix. The following estimates the model shown earlier, now with variance function that is a function of whether the individual is married.

```
-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable               DOCVIS
Log likelihood function    -60155.92470
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .62399***      .05532    11.28  .0000      .51556    .73241
     AGE|     .02056***      .00071    29.00  .0000      .01917    .02195
  HHNINC|    -.48776***      .04598   -10.61  .0000     -.57787   -.39764
    EDUC|    -.03555***      .00384    -9.27  .0000     -.04307   -.02803
  FEMALE|     .32528***      .01592    20.43  .0000      .29408    .35648
        |Dispersion parameter for count data model
   Alpha|    2.06990***      .04103    50.45  .0000     1.98948   2.15031
        |Heterogeneity in dispersion parameter
 MARRIED|    -.11067***      .02258    -4.90  .0000     -.15493   -.06642
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1879
Scale Factor for Marginal Effects  3.1879
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .06554***      .00248    26.44  .0000      .06068    .07039
  HHNINC|   -1.55492***      .14763   -10.53  .0000    -1.84427  -1.26558
    EDUC|    -.11334         .76186     -.15  .8817    -1.60657   1.37989
  FEMALE|    1.02846*        .58182     1.77  .0771     -.11189   2.16881   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E41.4.4 Negbin 1, Negbin 2 and Negbin P

The literature, mostly associating the result with Cameron and Trivedi's early (1986) work, defines two familiar forms of the negative binomial model. Where

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i),$$

the Negbin 2 form of the probability is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i + 1)} u_i^{\theta}(1 - u_i)^{y_i}$$

where $u_i = \theta / (\theta + \lambda_i)$

and $\theta = 1/\alpha.$

This is the default form of the model in most (if not all) of the received econometrics packages that provide an estimator for this model. This is the form of the model we have used up to this point. The Negbin 1 form of the model results if θ in the preceding is replaced with $\theta_i = \theta\lambda_i$. Then, u_i becomes $u = \theta/(1+\theta)$, and the density becomes

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(\theta\lambda_i)\Gamma(y_i + 1)} w^{\theta\lambda_i}(1 - w\ )^{y_i}$$

where $w = \theta / (\theta + 1)$

LIMDEP will fit the model with this specification by adding

; Model = NB1

to the **NEGBIN** model command. An example appears below.

We note, this is somewhat more than a simple reparameterization of the model. The results below show that the likelihood function is not quite equal at the maximum, and the parameters are not simple transformations in one model vs. the other. We are not aware of a theory that justifies using one form or the other for the negative binomial model. The two are not nested, so we cannot carry out a likelihood ratio test of one versus the other. The Negbin P family does nest both of them, so this may provide a more general, encompassing approach to finding the right specification. This is examined below.

The results below refit our model using the Poisson specification, Negbin 1 and Negbin 2. Since the conditional mean function in all three cases is

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i),$$

the three sets of parameter estimates should be similar, as they are. However, we have already rejected the Poisson model in favor of either negative binomial model.

```
----------------------------------------------------------------------------
Poisson Regression
Dependent variable                DOCVIS
Log likelihood function   -103923.54929
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|     .57813***      .02630    21.98  .0000      .52659    .62968
     AGE|     .02057***      .00031    67.30  .0000      .01997    .02117
  HHNINC|    -.52855***      .02189   -24.14  .0000     -.57146   -.48565
    EDUC|    -.02868***      .00173   -16.57  .0000     -.03208   -.02529
  FEMALE|     .29405***      .00700    42.00  .0000      .28033    .30777
--------+-------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function    -60164.22014
NegBin form 2; Psi(i) = theta
Tests of Model Restrictions on Neg.Bin.
Model               Logl ChiSquared[df]
Poisson(b=0)  -108662.14  ******** [**]
Poisson       -103923.55    9477.2 [ 4]
Negative Bin.  -60164.22   87518.7 [ 1]
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|     .62857***      .05457    11.52  .0000      .52162    .73553
     AGE|     .02042***      .00070    29.07  .0000      .01904    .02179
  HHNINC|    -.48779***      .04520   -10.79  .0000     -.57637   -.39921
    EDUC|    -.03539***      .00378    -9.36  .0000     -.04281   -.02798
  FEMALE|     .32673***      .01588    20.58  .0000      .29561    .35784
        |Dispersion parameter for count data model
   Alpha|    1.90309***      .01984    95.94  .0000     1.86421   1.94197
--------+-------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function    -60063.78559
NegBin form 1;Psi(i) = theta*exp[bx(i)]
Tests of Model Restrictions on Neg.Bin.
Model               Logl ChiSquared[df]
Poisson(b=0)  -108662.14  ******** [**]
Poisson       -103923.55    9477.2 [ 4]
Negative Bin.  -60063.79   87719.5 [ 1]
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|     .47184***      .05439     8.67  .0000      .36523    .57845
     AGE|     .01710***      .00065    26.24  .0000      .01582    .01838
  HHNINC|    -.22813***      .04417    -5.16  .0000     -.31471   -.14155
    EDUC|    -.01544***      .00356    -4.34  .0000     -.02241   -.00847
  FEMALE|     .32894***      .01477    22.28  .0000      .30000    .35788
        |Dispersion parameter for count data model
   Alpha|    6.11096***      .06715    91.00  .0000     5.97934   6.24258
--------+-------------------------------------------------------------------
```

The more general Negbin P model is obtained by replacing θ in

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i+1)} u_i^{\theta}(1-u_i)^{y_i}$$

where $u_i = \theta / (\theta + \lambda_i)$,

with $\theta\lambda_i^{2-P}$. We have examined the cases of P = 1 and P = 2. For convenience, let $Q = 2 - P$, Then, the density is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i^Q + y_i)}{\Gamma(\theta\lambda_i^Q)\Gamma(y_i+1)} \left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q + \lambda_i}\right)^{\theta\lambda_i^Q} \left(\frac{\lambda}{\theta\lambda_i^Q + \lambda_i}\right)^{y_i}$$

This model is also built into *LIMDEP*. To request it, use

NEGBIN ; all as usual for your count data model
; Model = NBP $

The following reestimates the negative binomial in this more general form.

```
-----------------------------------------------------------------------------
Negative Binomial (P) Model
Dependent variable               DOCVIS
Log likelihood function    -60029.85010
Restricted log likelihood -103923.54929
Chi squared [   1 d.f.]     87787.39838
Significance level               .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .42597***      .06020     7.08  .0000       .30799     .54396
     AGE|      .02028***      .00073    27.87  .0000       .01886     .02171
  HHNINC|     -.34404***      .05023    -6.85  .0000     -.44250    -.24558
    EDUC|     -.02284***      .00411    -5.56  .0000     -.03089    -.01479
  FEMALE|      .36359***      .01643    22.13  .0000       .33139     .39580
        |Dispersion parameter for count data model
   Alpha|     3.83035***      .14966    25.59  .0000     3.53702    4.12367
        |Negative Binomial. General form, NegBin P
       P|     1.39570***      .03249    42.96  .0000     1.33203    1.45938
--------+--------------------------------------------------------------------
```

Note that the log likelihood function continues to increase. For this model, the likelihood ratio test against the NB2 model gives chi squared of -2(-60164.22 - (-60029.97)) = 268.5, which far exceeds the critical value of 3.84. The Wald (t) test would be (1.3975 - 2)/.03249 = -18.31, which is likewise significant.

For exploring the functional form, it may be useful to fix the value of P in the estimation. You can use **; Rst = list** in general. For the NBP model, a convenient shorthand if P is the only parameter to be restricted is

; Scale = the desired value.

In the example below, we have used **; Scale = 1.5**.

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                 DOCVIS
Log likelihood function   -103923.54929
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .57813***      .02630    21.98  .0000       .52659    .62968
     AGE|      .02057***      .00031    67.30  .0000       .01997    .02117
  HHNINC|     -.52855***      .02189   -24.14  .0000      -.57146   -.48565
    EDUC|     -.02868***      .00173   -16.57  .0000      -.03208   -.02529
  FEMALE|      .29405***      .00700    42.00  .0000       .28033    .30777
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:  10 iterations. Status=0, F=    60164.22
Normal exit:  11 iterations. Status=0, F=    60033.20
-----------------------------------------------------------------------------
Negative Binomial (P) Model
Dependent variable                 DOCVIS
Log likelihood function    -60033.20247
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .43551***      .06064     7.18  .0000       .31665    .55437
     AGE|      .02088***      .00073    28.64  .0000       .01945    .02230
  HHNINC|     -.37749***      .05080    -7.43  .0000      -.47705   -.27793
    EDUC|     -.02527***      .00414    -6.10  .0000      -.03339   -.01714
  FEMALE|      .36737***      .01650    22.26  .0000       .33503    .39971
        |Dispersion parameter for count data model
   Alpha|     3.39373***      .03382   100.35  .0000      3.32744   3.46001
        |Negative Binomial. General form, NegBin P
       P|     1.50000    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

The model in the NBP form is built into *LIMDEP*, but it is also easy to formulate it as a user defined procedure with **MAXIMIZE**. The general form would be as follows:

```
SAMPLE     ; whatever is appropriate for your application $
NAMELIST   ; x = your set of independent variables
CALC       ; k = Col(x) $
CREATE     ; y = your dependent variable $
NEGBIN     ; Lhs = y ; Rhs = x $
MATRIX     ; b0 = b ; t0 = 1/alpha $
CALC       ; q0 = 0 $
MAXIMIZE   ; Start = b0,t0,q0
           ; Labels = k_c,t,q
           ; Fcn = al  = Exp(c1'x)    |
                   tlq = t*(al^q)      |
                   w  = tlq/(al+tlq)  |
            Lgm(y+tlq) - Lgm(tlq) - Lgm(y+1) + tlq*Log(w) + y*Log(1-w) $
```

A good starting value for Q is helpful. One strategy that might be used is to fix Q in the model at some specific values, by providing specific starting values and using

; Fix = q

In the models already estimated, we fit Q with $Q = 0$ (Negbin 2) and $Q = 1$ (Negbin 1). Experimenting with values between zero and one may be useful. In our estimates below, the estimated value of Q is 0.60429 (consistent with P = 1.3957 using the built in procedure earlier). The data set in use for these applications is particularly rich and well behaved. The estimator for this model was actually quite routine. Parameter estimates for the fully general model (Negbin 1.3957) are shown below.

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function    -60029.85010
Estimation based on N =  27326, K =   7
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      C1|     .42597***      .06020     7.08  .0000      .30799     .54396
      C2|     .02028***      .00073    27.87  .0000      .01886     .02171
      C3|    -.34404***      .05023    -6.85  .0000     -.44250    -.24558
      C4|    -.02284***      .00411    -5.56  .0000     -.03089    -.01479
      C5|     .36359***      .01643    22.13  .0000      .33139     .39580
       T|     .26107***      .01020    25.59  .0000      .24108     .28107
       Q|     .60429***      .03249    18.60  .0000      .54062     .66797
--------+--------------------------------------------------------------------
```

In the results above, t, the estimate of θ, is an estimate of $1/\alpha$. To compare it to the negative binomial model, we could use the delta method to estimate α and an asymptotic standard error. The estimate would be 3.83039, which suggests much more dispersion than implied by the Negbin 2 model.

E41.4.5 Technical Details

The negative binomial model arises is as a modification of the Poisson model in which the mean is now μ_i, respecified so that

$$\log \mu_i = \log\lambda_i + \varepsilon_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

where $\exp(\varepsilon_i)$ has a gamma distribution with mean 1.0 and variance α. (This is one of several variants of the negative binomial model discussed by Cameron and Trivedi (1986).) The resulting conditional probability distribution is

$$\text{Prob}(Y = y_i|\varepsilon_i,\mathbf{x}_i) = \frac{\exp[-(\exp(\varepsilon_i)\lambda_i)][\exp(\varepsilon_i)\lambda_i]^{y_i}}{y_i!},\ y_i = 0,1,...$$

where
$$f[\exp(\varepsilon_i)] = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta\exp(\varepsilon_i)}[\exp(\varepsilon_i)]^{\theta-1},\ \theta = 1/\alpha,\ \exp(\varepsilon_i) > 0.$$

The unconditional distribution of y_i is obtained by taking the expectation with respect to $\exp(\varepsilon_i)$ of the conditional probability. For convenience, let $\tau_i = \exp(\varepsilon_i)$. Then

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \int_0^\infty \frac{\exp[-\tau_i\lambda_i][\tau_i\lambda_i]^{y_i}}{y_i!}\frac{\theta^\theta}{\Gamma(\theta)}e^{-\theta\tau_i}\tau_i^{\theta-1}d\tau_i$$

This is a gamma integral that can be simplified considerably. Collecting terms and using results for the gamma integral, this reduces to

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\theta^\theta\lambda_i^{y_i}}{\Gamma(\theta)y_i!}\frac{\Gamma(y_i+\theta)}{(\lambda_i+\theta)^{y_i+\theta}}$$

We have reparameterized the probability distribution in terms of θ because this simplifies the formulation and computation of the log likelihood and its derivatives. The formulation of the result that we use for optimization is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta+y_i)}{\Gamma(\theta)\Gamma(y_i+1)}u_i^\theta(1-u_i)^{y_i}$$

where $u_i = \theta / (\theta + \lambda_i)$

and $\theta = 1/\alpha$.

For optimization and forming the BHHH estimator, we have

$$\partial\log L_i/\partial\lambda_i = [\theta/u_i - y_i/(1-u_i)]\partial u_i/\partial\lambda_i.$$

$$\partial u_i/\partial\lambda_i = -u_i / (\theta + \lambda_i) = -u_i(1 - u_i) / \lambda_i$$

$$\partial u_i/\theta = u_i(1-u_i)/\theta$$

$$\partial\lambda_i/\partial\boldsymbol{\beta} = \lambda_i\,\mathbf{x}_i.$$

Combining terms, $\partial\log L_i/\partial\boldsymbol{\beta} = [y_iu_i - \theta(1 - u_i)]\mathbf{x}_i.$

Also, $\partial\log L_i/\partial\theta = \Psi(\theta + y_i) - \Psi(\theta) + \log u_i + (1 - u_i) - y_iu_i/\theta$

where $\Psi(z) = d\log\Gamma(z)/dz.$

The Hessian is

$$\partial^2 \log L_i/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -(\theta + y_i)u_i\,(1 - u_i)\,\mathbf{x}_i\,\mathbf{x}_i'$$

$$\partial^2 \log L_i/\partial\boldsymbol{\beta}\partial\theta = [-(1-u_i)^2 + y_i u_i(1 - u_i)/\theta]\,\mathbf{x}_i$$

$$\partial^2 \log L_i/\partial\theta^2 = \Psi'(\theta + y_i) - \Psi'(\theta) + (1 - u_i)^2 / \theta + y_i\,(u_i/\theta)^2$$

These are used in computation of the log likelihood function, gradient, and estimate of the asymptotic covariance matrix.

Greene (2008), define the class of Negbin P models by the relationship between mean and variance functions,

$$E[y_i|\mathbf{x}_i] = \lambda_i \text{ and } \mathrm{Var}[y_i|\mathbf{x}_i] = \lambda_i + \alpha\lambda_i^P.$$

The model already considered, the standard case, is their model Negbin 2, or NB2. An alternative form labeled Negbin 1 or NB1 is obtained by using $P = 1$. The density is obtained by replacing θ with $\theta\lambda_i$ in Prob($Y = y_i|\mathbf{x}_i$). This produces, after a bit of manipulation,

$$\mathrm{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(\theta\lambda_i)\Gamma(y_i + 1)} w_i^{\theta\lambda_i}(1 - w_i)^{y_i}$$

where $w_i = \theta / (\theta + 1)$

and $\theta = 1/\alpha$.

This is not a simple reparameterization of the model; it is a different model. An example given in Section E41.4.2 demonstrates. We also consider the fully general form of their negative binomial model, Negbin P.

The more general Negbin P model is obtained by replacing θ in

$$\mathrm{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i + 1)} u_i^{\theta}(1 - u_i)^{y_i}$$

where $u_i = \theta / (\theta + \lambda_i)$

with $\theta\lambda_i^{2-P}$. We have examined the cases of $P = 1$ and $P = 2$. For convenience, let $Q = 2 - P$. Then, the density is

$$\mathrm{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i^Q + y_i)}{\Gamma(\theta\lambda_i^Q)\Gamma(y_i + 1)} \left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q + \lambda_i}\right)^{\theta\lambda_i^Q} \left(\frac{\lambda}{\theta\lambda_i^Q + \lambda_i}\right)^{y_i}$$

Derivatives of $\log L_i$ for the general Negbin P model are tedious. We obtain them by writing the density as

$$\log L_i = \log\Gamma(y_i + g_i) - \log\Gamma(g_i) - \log\Gamma(y_i) + g_i \log w_i + y_i \log(1-w_i)$$

where $g_i = \theta\,\lambda_i^Q$ and $w_i = g_i / (g_i + \lambda_i)$.

Then, $\partial\log L_i / \partial\lambda_i = [\Psi(y_i + g_i) - \Psi(g_i) + \log w_i]\,\partial g_i/\partial\lambda_i + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial\lambda_i$

$$\partial\log L_i / \partial\theta = [\Psi(y_i + g_i) - \Psi(g_i) + \log w_i]\,\partial g_i/\partial\theta + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial\theta$$

$$\partial\log L_i / \partial Q = [\Psi(y_i + g_i) - \Psi(g_i) + \log w_i]\,\partial g_i/\partial Q + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial Q.$$

The inner parts are:

$$\partial g_i / \partial\lambda_i = \theta Q\lambda_i^{Q-1} = (Q/\lambda_i)g_i$$

$$\partial g_i / \partial\theta = \lambda_i^Q = (1/\theta)g_i$$

$$\partial g_i / \partial Q = \theta\lambda_i^Q \log\lambda_i = \log\lambda_i\, g_i$$

$$\partial w_i / \partial\lambda_i = [(Q-1)/\lambda_i]w_i(1-w_i)$$

$$\partial w_i / \partial\theta = (1/\theta)w_i(1-w_i)$$

$$\partial w_i / \partial Q = \log\lambda_i\, w_i(1-w_i)$$

Collecting terms, now, let $A_i = [\Psi(y_i + g_i) - \Psi(g_i) + \log w_i]$

$$B_i = [g_i(1 - w_i) - y_i w_i],$$

to obtain

$$\partial\log L_i/\partial\begin{pmatrix}\lambda_i\\ \theta\\ Q\end{pmatrix} = [A_i + B_i]\begin{pmatrix}Q/\lambda_i\\ 1/\theta\\ \log\lambda_i\end{pmatrix} - B_i\begin{pmatrix}1/\lambda_i\\ 0\\ 0\end{pmatrix}.$$

The final element needed is $\partial\log L_i/\partial\beta = \lambda_i\mathbf{x}_i$. We use these and the BHHH estimator to compute the maximum likelihood estimates and their asymptotic standard errors for the NBP model. When this model is estimated, there are three sets of iterations. The Poisson model is estimated first. These results are shown with the results. The Negbin 2 model is then estimated using the Poisson estimates as starting values. The NB2 results are not displayed, but you will observe this second set of iterations. These are used to improve the starting values for the NBP estimates. The starting values for NBP are the NB2 estimates, with $P = 2$ ($Q = 0$).

For all forms of the negative binomial model, when θ is allowed to be heteroscedastic, then $\theta_i = \theta\exp(\boldsymbol{\gamma}'\mathbf{z}_i) = \theta v_i$. To obtain the derivatives for the underlying parameters, let $\Delta_i = \partial\log L_i/\partial\theta_i$. Then, $\partial\log L_i/\partial\theta = \Delta_i v_i$ and $\partial\log L_i/\partial\boldsymbol{\gamma} = \Delta_i\theta_i\mathbf{z}_i$.

E41.5 Other Models for Count Data

There is a huge literature on variants of the Poisson model for counts. (See, e.g., Winkelmann (2003) or Hilbe (2011).) We have included estimators for several of them. In all cases, the models relax the equidispersion assumption of the Poisson model.

E41.5.1 Gamma Model with Under- or Overdispersion

The gamma model proposed by Winkelmann (1995) and also discussed in Cameron and Trivedi (1998) represents a significant innovation. The large majority of the extensions of the Poisson model that have been proposed have accommodated overdispersion – that is, variance greater than the mean. Underdispersion is a phenomenon which has been much less convenient to model directly – some extensions, such as various forms of the 'with zeros' models discussed in Chapter E43, can induce underdispersion, but otherwise involve more structure than desired. This extension provides a straightforward, easily implemented approach to a general model for counts that allows both under- and overdispersion. As this is a new technique (and, to our knowledge, the first implementation in a general econometrics package), a detailed presentation of the mathematical background is presented here.

The gamma (based) probability model for counts is

$$\text{Prob}[y_i = j] = G(\alpha j, \lambda_i) - G(\alpha j + \alpha, \lambda_i)$$

where $$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i) \text{ (as usual)}$$

and $$G(\alpha j, \lambda_i) = 1 \text{ if } j = 0, \text{ or } \frac{1}{\Gamma(\alpha j)}\int_0^{\lambda_i} u^{\alpha j-1}e^{-u}du \text{ if } j > 0, j = 1,...$$

The dispersion parameter is α; there is underdispersion if $\alpha > 1$, overdispersion if $\alpha < 1$, and equidispersion if $\alpha = 1$, which reduces the gamma probability to the Poisson model. The conditional mean function is

$$\text{E}[y_i|\mathbf{x}_i] = \sum_{j=1}^{\infty} jG(\alpha j, \lambda_i)$$

This has no closed form, but an approximation that we use is

$$\text{E}[y_i|\mathbf{x}_i] \approx \lambda_i / \alpha.$$

The Poisson case arises conveniently if $\alpha = 1$. With this approximation, the marginal effects are

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = \frac{\lambda_i}{\alpha}\boldsymbol{\beta}$$

The gamma model is requested with

POISSON ; ... as usual ... ; Model = gamma $

All other options available with the Poisson model are retained, including the fitted values, restrictions, optimization parameters, etc. Estimation is done via the BFGS algorithm. Since this model is quite complex, the algorithm parameters should not be changed. The only difference beyond the visible output, which is clearly marked, is the new scalar, *alpha*, which is retained by the estimator.

NOTE: The derivatives for this model are computed numerically, not analytically. The BHHH estimator is used to estimate the asymptotic covariance matrix of the MLE.

The gamma model for counts arises as follows: Assume that waiting times between occurrences of events (which are counted to produce the 'count variable') are distributed as a continuous, two parameter gamma variate, with shape parameter α and location parameter $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$ – we have skipped an introductory step and layered the regression model in at the outset. Then, the density for interarrival times is

$$f(t \mid \alpha, \lambda_i) \quad = \quad \frac{\lambda_i^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda_i t}, t \geq 0, \alpha > 0, \lambda_i > 0.$$

The arrival time of the jth event is

$$Q_j \quad = \quad t_1 + t_2 + \ldots + t_j.$$

The gamma distribution is 'reproductive;' the density of Q_j is

$$f(Q_j \mid \alpha, \lambda_i) \; = \; \frac{\lambda_i^{\alpha j}}{\Gamma(\alpha j)} Q_j^{\alpha j-1} e^{-\lambda_i Q_j}, Q_j \geq 0, \alpha > 0, \lambda_i > 0.$$

The cumulative distribution function is

$$\begin{aligned} F(T \mid \alpha, \lambda_i) \; &= \; \int_0^T \frac{\lambda_i^{\alpha j}}{\Gamma(\alpha j)} u^{\alpha j-1} e^{-\lambda_i u} du, \; \alpha > 0, \lambda_i > 0 \\ &= \; \frac{1}{\Gamma(\alpha j)} \int_0^{\lambda_i T} u^{\alpha j-1} e^{-u} du, \mathrm{j} = 0,1,\ldots \\ &= \; G(\alpha j, \lambda_i T). \end{aligned}$$

The integral is an incomplete gamma function (probability; $G(\alpha j, 0) = 0$ and $G(\alpha j, \infty) = 1$) which must be approximated numerically. Note that $G(0,\lambda_i T) = 1$. If j events occur in a period of length T or less, then equivalently, j or more events occur in the period of fixed length T. Thus,

$$G(\alpha j, \lambda_i T) \quad = \quad \text{Prob}[j \text{ events}] + \text{P}[j\text{+1 events}] \ldots \text{ in period of length } T$$

from which it follows that

$$G(\alpha(j+1), \lambda_i T) \; = \; \text{Prob}[j\text{+1 events}] \ldots$$

so that

$$\text{Prob}[j \text{ events}] \; = \; G(\alpha j, \lambda_i T) \; - \; G(\alpha(j+1), \lambda_i T).$$

We now normalize the period length to $T = 1$ to obtain the distribution for counts of events,

$$\text{Prob}[j \text{ events}] \; = \; G(\alpha j, \lambda_i) \; - \; G(\alpha j + \alpha, \lambda_i)$$

The mean/variance relationship is complicated, but it can be shown that the variance exceeds the mean if $\alpha < 1$, and is less than it if $\alpha > 1$.

E41.5.2 Generalized Poisson Models – GP1, GP2, GPP

The density for the generalized Poisson model suggested by Consul and Jain (1973) is

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \left(\frac{\lambda_i}{1+\theta\lambda_i}\right)^{y_i} \frac{(1+\theta y_i)}{y_i!} \exp\left(-\frac{\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\right), y_i = 0,1,2,...;\ \lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}.$$

The mean and variance of this random variable are

$$E[y_i \mid \mathbf{x}_i] = \lambda_i,\ \text{Var}[y_i \mid \mathbf{x}_i] = \lambda_i(1+\theta\lambda_i)^2.$$

Marginal effects are identical to those in the base Poisson model. The overdispersion, as in all the negative binomial models, is a 'mean preserving spread;' the mean is unchanged – mass is moved in both directions.

The parameter θ is unrestricted. The model provides for both over- and underdispersion, though the latter is likely to be the more empirically relevant case. Negative values can produce computational problems. However, small negative values consistent with underdispersion are, nonetheless admissible. The generalized Poisson model reverts to the familiar Poisson regression if θ = 0.

This model is requested with

POISSON ; <... the usual setup...>
; Model = GP $

The general options provided for Poisson models, including marginal effects, fitted values, constraints, weights, clustering, etc. are all available. However, truncation and censoring are not available for this model.

The 'nonPoissonness' of the distribution is embodied in the ancillary parameter θ. An extension of the model allows θ to be a linear function of any variables, using

; Hfn = list of variables

The list should include a constant term. (If you omit it, one is automatically inserted.)

As in the case of the negative binomial model, there are GP1, GP2 and GPP forms of the model. The same extension is used. The '*P*' form of the model is obtained by replacing θ with $\theta\lambda_i^{2-P}$ in the general form of the density. The default form is the GP2 model, which is obtained with

; Model = GP or GP2

The others are specified with

; Model = GP1
or **; Model = GPP**.

In all cases, the conditional mean function is still λ_i, so partial effects, the **PARTIALS** command and **SIMULATE** all work as they do for the Poisson and base case negative binomial model. (The GPP form of the model builds on Greene (2008) and was proposed explicitly by Ismail (2010).)

In the example below, we have fit the GPP model, then reported the partial effects for the base Poisson model for comparison. They are surprisingly different.

SAMPLE ; All $
NAMELIST ; x = one,age,hhninc,educ,female $
POISSON ; Lhs = docvis ; Rhs = x ; Model = GPP ; Partial Effects $
POISSON ; Lhs = sumy ; Rhs = x ; Partial Effects $

```
-----------------------------------------------------------------------------
Generalized Poisson (P) Model
Dependent variable              DOCVIS
Log likelihood function    -59915.33565
Restricted log likelihood -103923.54929
Chi squared [   1 d.f.]     88016.42728
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .37435***      .05994     6.25  .0000      .25688    .49182
     AGE|     .02009***      .00079    25.39  .0000      .01854    .02164
  HHNINC|    -.29243***      .04880    -5.99  .0000     -.38809   -.19678
    EDUC|    -.01970***      .00395    -4.99  .0000     -.02745   -.01196
  FEMALE|     .37817***      .01744    21.68  .0000      .34399    .41236
        |Dispersion parameter in generalized Poisson model
Constant|    1.43145***      .06229    22.98  .0000     1.30936   1.55354
        |Nesting Parameter for P form of Generalized Poisson
       P|    1.25803***      .03593    35.01  .0000     1.18760   1.32846
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   3.0160
Scale Factor for Marginal Effects  3.0160
(Generalized Poisson – P)
--------+--------------------------------------------------------------------
        |     Partial     Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .06060***      .00236    25.69  .0000      .05597    .06522
  HHNINC|    -.88196***      .14651    -6.02  .0000    -1.16912   -.59481
    EDUC|    -.05942***      .01188    -5.00  .0000     -.08270   -.03615
  FEMALE|    1.14055***      .05275    21.62  .0000     1.03716   1.24395   #
--------+--------------------------------------------------------------------
(Poisson)
--------+--------------------------------------------------------------------
        |     Partial     Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .06550***      .00100    65.62  .0000      .06354    .06745
  HHNINC|   -1.68266***      .06992   -24.06  .0000    -1.81971  -1.54562
    EDUC|    -.09132***      .00552   -16.54  .0000     -.10214   -.08050
  FEMALE|     .93023***      .02210    42.10  .0000      .88693    .97354   #
--------+--------------------------------------------------------------------
```

E41.5.3 Polya-Aeppli Model

The density for the Polya-Aeppli form of the Poisson model is

$$\text{Prob}[Y=0\,|\,\mathbf{x}_i]=\exp(-\lambda_i),$$
$$\text{Prob}[Y=y_i\,|\,\mathbf{x}_i]=\exp(-\lambda_i)(1-\theta)^{y_i}\sum_{j=1}^{y_i}\binom{y_i-1}{j-1}\frac{[\lambda_i(1-\theta)/\theta]^j}{j!},\ y_i=1,2,...$$
$$\lambda_i=e^{\boldsymbol{\beta}'\mathbf{x}_i}$$
$$0<\theta<1.$$

The mean and variance of this random variable are

$$E[y_i\,|\,\mathbf{x}_i]=\frac{\lambda_i}{(1-\theta)},\ \text{Var}[y_i\,|\,\mathbf{x}_i]=\frac{\lambda_i(1+\theta)}{(1-\theta)^2}$$

In order to impose the constraint on θ, *LIMDEP* estimates α then reports θ, where $\theta = \exp(\alpha)/[1+\exp(\alpha)]$. This model is primarily useful for data with excess zeros. Note that the density is of the same form as the Poisson model, in which

$$\text{Prob}[Y=0]=\exp(-\lambda_i),$$
$$\text{Prob}[Y=y_i\,|\,\mathbf{x}_i]=\exp(-\lambda_i)\frac{\lambda_i^{y_i}}{y_i!},\ y_i=1,2,...$$

The Polya-Aeppli model reverts to the Poisson model if $\theta = 0$. The model is requested with

```
POISSON        ; <... the usual setup...>
               ; Model = Polya $
```

The general options provided for Poisson models, including marginal effects, fitted values, constraints, weights, clustering, etc. are all available. However, truncation and censoring are not available for this model.

The Polya-Aeppli model is useful for modeling ‘excess zeros.’ (Note that the zero inflation models described in the next chapter are also specifically designed for that purpose.) Like other Poisson models, the framework is most compatible with values of the Lhs variable that truly are counts, generally of small to moderate numbers of events. Numbers of patents, visits to the medical establishment, visits to recreational sites, highway fatalities, etc. are applications that come to mind. Applications that stretch the definitions, such as one recently reported to us involving counts (in the millions) of financial transactions, are probably best applied in some other modeling platform. The danger here is that in models such as these, there is considerable risk of numerical overflow in attempting to compute the coefficients of the model.

For analysis of the Polya-Aeppli model, Johnson and Kotz (1993) provide some very useful recursions:

$$\text{Prob}[Y=j+1]=\frac{\lambda_i(1-\theta)}{(j+1)}\sum_{m=0}^{j}(j+1-m)\theta^{j-m}\text{Prob}[Y=j]$$

$$\frac{\partial \text{Prob}(Y=y_i)}{\partial \lambda_i}=(y_i/\lambda_i-1)\text{Prob}(Y=y_i)-\frac{(y_i-1)}{\lambda_i}\text{Prob}(Y=y_i-1),\text{ so that}$$

$$\frac{\partial \log \text{Prob}(Y=y_i)}{\partial \boldsymbol{\beta}}=\left[(y_i-\lambda_i)-(y_i-1)\frac{\text{Prob}(Y=y_i-1)}{\text{Prob}(Y=y_i)}\right]\mathbf{x}_i \text{ and}$$

$$\frac{\partial \log \text{Prob}(Y=y_i)}{\partial \theta}=\frac{1}{1-\theta}\left[(y_i-1)\frac{\text{Prob}(Y=y_i-1)}{P(y_i)}-y_i\right].$$

```
-----------------------------------------------------------------------------
Polya-Aeppli Model
Dependent variable               DOCVIS
Log likelihood function    -60999.55518
Restricted log likelihood -103923.54929
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.66892***      .05375   -12.44  .0000     -.77427   -.56356
     AGE|     .01713***      .00063    27.15  .0000      .01590    .01837
  HHNINC|    -.27029***      .04386    -6.16  .0000     -.35626   -.18432
    EDUC|    -.01733***      .00357    -4.85  .0000     -.02433   -.01034
  FEMALE|     .31139***      .01426    21.84  .0000      .28344    .33934
        |Shape parameter for Polya-Aeppli distribution
   Alpha|     .69409***      .00458   151.69  .0000      .68512    .70306
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1835
Scale Factor for Marginal Effects  3.1835
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .05455***      .00218    25.06  .0000      .05028    .05881
  HHNINC|    -.86049***      .13993    -6.15  .0000    -1.13475   -.58622
    EDUC|    -.05518***      .01139    -4.85  .0000     -.07750   -.03286
  FEMALE|     .99131***      .04821    20.56  .0000      .89682   1.08581   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E41.5.4 The Logarithmic Distribution

The logarithmic distribution for a *positive* count variable is

$$\text{Prob}(Y = y_i) = \frac{\alpha\theta^{y_i}}{y_i},\ y_i = 1, 2, \ldots \text{ and } 0 < \theta < 1$$

where $\alpha = -1 / \log(1 - \theta)$.

(See Winkelmann (2008).) We can produce a regression model in this context by the parameterization

$$\theta_i = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_i)}.$$

The mean is $\quad \text{E}[y|\mathbf{x}] = \dfrac{\alpha\theta}{1-\theta} = \alpha\exp(\boldsymbol{\beta}'\mathbf{x}) = \mu_i$.

After some tedious algebra, we obtain the partial effects:

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = \mu_i \alpha_i (1 - \alpha_i \theta_i)\boldsymbol{\beta}.$$

We implement the logarithmic model with a user written **MAXIMIZE** command and obtain the partial effects and simulation with **PARTIALS** and **SIMULATE**. The following template procedure can be used:

```
PROC = LogModel(y,x) $
CALC        ; k = Col(x) $
MAXIMIZE    ; Start = k_0 ; Labels = k_b
            ; Fcn = bxi = b1'x                       |
              thetai = Exp(bxi) / (1+Exp(bxi))       |
              ai = -1 / Log(1-thetai)                |
              Log(ai) + y*Log(thetai) - Log(y) $
PARTIALS    ; Parameters = b ; Labels = k_b
            ; Function = bxi = b1'x                  |
              thetai = Exp(bxi) / (1+Exp(bxi))       |
              ai = -1 / Log(1-thetai)                |
              ai*thetai / (1-thetai)
            ; Effects: x ; Summary $
ENDPROC
```

The model is fit using the 1991 data on doctor visits. For comparison, we have fit a Poisson model truncated at zero using the same data. The results are strikingly similar, which suggests that the difference in the functional forms is much less than it might appear at first.

SAMPLE ; All $
REJECT ; year # 1991 | docvis = 0 $
NAMELIST ; x = one,age,educ,hhninc,female $
EXECUTE ; Proc = LogModel(docvis,x) $
POISSON ; Lhs = docvis ; Rhs = one,x ; Truncation ; Limit = 0 ; Partial Effects $

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function     -6607.64511
Estimation based on N =   2932, K =   5
Inf.Cr.AIC  =13225.290 AIC/N =    4.511
Model estimated: Jul 28, 2011, 03:07:14
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    2.11282***      .25447     8.30  .0000     1.61407   2.61156
      B1|     .01915***      .00334     5.74  .0000      .01261    .02570
      B2|    -.05520***      .01872    -2.95  .0032     -.09188   -.01852
      B3|    -.48305**       .20597    -2.35  .0190     -.88674   -.07937
      B4|     .16321**       .07898     2.07  .0388      .00840    .31801
--------+--------------------------------------------------------------------
(Poisson coefficients)
--------+--------------------------------------------------------------------
Constant|    1.33639***      .07360    18.16  .0000     1.19213   1.48065
     AGE|     .01289***      .00082    15.68  .0000      .01127    .01450
    EDUC|    -.03860***      .00495    -7.80  .0000     -.04831   -.02890
  HHNINC|    -.39491***      .05694    -6.94  .0000     -.50650   -.28332
  FEMALE|     .09550***      .01958     4.88  .0000      .05711    .13388
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
---------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect       Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .04803      .00898    5.35       .03044       .06562
      EDUC          -.13845      .04807    2.88      -.23267      -.04422
      HHNINC       -1.21150      .51464    2.35     -2.22018      -.20281
   *  FEMALE         .40602      .19803    2.05       .01788       .79415
---------------------------------------------------------------------
(Poisson partial effects)
--------+--------------------------------------------------------------------
     AGE|     .04821***      .00310    15.54  .0000      .04212    .05429
    EDUC|    -.14443***      .01856    -7.78  .0000     -.18081   -.10804
  HHNINC|   -1.47742***      .21340    -6.92  .0000    -1.89567  -1.05917
  FEMALE|     .37794***      .07709     4.90  .0000      .22685    .52903   #
--------+--------------------------------------------------------------------
```

E41.5.5 NegBin X

The NBX model was proposed by Silva and Windmeijer (2001). (See Winkelmann (2008).) Let S be distributed as Poisson(λ_i). Let R_1, R_2, …, R_S be S draws from the logarithmic distribution described in the previous section. (Note, S may be zero.) Then, the random variable with NegBin X distribution is

$$Y = R_1 + R_2 + \ldots + R_S = \sum_{i=1}^{S} R_i \text{ , } S \sim \text{Poisson } (\lambda).$$

The parameter of each draw from the logarithmic distribution is

$$\theta_i = \frac{\exp(\boldsymbol{\gamma}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\gamma}'\mathbf{x}_i)}.$$

and the parameter of the Poisson distribution is

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i).$$

Winkelmann provides the density for y_i,

$$\text{Prob}(y_i|\mathbf{x}_i) = \frac{\Gamma(y_i+\mu_i)\exp(-\lambda_i)}{\Gamma(y_i+1)\Gamma(\mu_i)[1+\exp(-\boldsymbol{\gamma}'\mathbf{x}_i)]^{y_i}}, y_i = 0,1,\ldots$$

where $$\mu_i = E[y_i|\mathbf{x}_i] = \frac{\lambda_i \exp(\boldsymbol{\gamma}'\mathbf{x}_i)}{\log[1+\exp(\boldsymbol{\gamma}'\mathbf{x}_i)]}.$$

The model is constrained thus far in that the variables $\mathbf{x}_i$ appear in both the logarithmic part and the Poisson part. Winkelmann argues that this is the natural specification of the model. The implementation below allows the variables in the logarithmic part of the model to differ from those in the Poisson.

The essential part of the command for the NBX model is

```
POISSON   ; Lhs = dependent variable
          ; Rhs = independent variables
          ; Model = NBX $
```

To relax the assumption that the same variables appear in both parts of the model, use

; Rh2 = full set of variables in the logarithmic part of the model.

In the example below, the NBX model is fit first with the same regressors in both parts of the equation. The built in routine for computing partial effects is used first. Then **PARTIALS** is used to redo the computation. The results of the two routines arises because the built in routine computes the partial effects at the means whereas **PARTIALS** computes the sample average partial effects (APE).

The differences between the two sets of partial effects arises because of the use of data means in the first case and the average partial effects in the second. When **PARTIALS** computes the effects at the means, the same results are obtained. A further small difference in the standard errors is the result of using analytic derivatives for the Jacobian in computing the effects within the command and numerical derivatives by **PARTIALS**. The decomposition of the partial effects produces a part due to the Poisson part of the probability and the remainder due to the logarithmic model component.

```
SAMPLE      ; All $
REJECT      ; year # 1991 $
NAMELIST    ; x = one,age,educ,hhninc,female,hhkids,working $
POISSON     ; Lhs = docvis ; Rhs = x ; Model = NBX ; Partial Effects $
PARTIALS    ; Effects: x ; Summary $
```

```
-----------------------------------------------------------------------------
Negative Binomial - X Model
Dependent variable               DOCVIS
Log likelihood function     -7930.91193
Mean of LHS Variable =          3.78294
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson Probability
Constant|    -.17270         .15540    -1.11  .2664      -.47728    .13187
     AGE|     .00985***      .00196     5.03  .0000       .00602    .01369
    EDUC|    -.00973         .00921    -1.06  .2910      -.02778    .00833
  HHNINC|    -.18959*        .10078    -1.88  .0599      -.38712    .00794
  FEMALE|     .38516***      .04290     8.98  .0000       .30107    .46924
  HHKIDS|    -.20199***      .04516    -4.47  .0000      -.29050   -.11348
 WORKING|    -.05160         .04800    -1.07  .2825      -.14568    .04249
        |Parameters of Logarithmic Model in NB-X
Constant|    2.03291***      .27977     7.27  .0000      1.48457   2.58126
     AGE|     .00967***      .00371     2.60  .0092       .00239    .01694
    EDUC|    -.03456*        .01800    -1.92  .0549      -.06985    .00073
  HHNINC|    -.24824         .20553    -1.21  .2271      -.65108    .15460
  FEMALE|     .01335         .07443      .18  .8576      -.13253    .15924
  HHKIDS|     .07117         .07980      .89  .3725      -.08524    .22757
 WORKING|    -.20767**       .08484    -2.45  .0144      -.37395   -.04139
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the partial effects produced by the built in specification in the model command.

```
Analysis of Partial Effects in Two Part Negative Binomial Model
Expected Value of DOCVIS   at means of all variables =   2.5865
===============================================================
                      Effect  Standard Error    t ratio
---------------------------------------------------------------
[AGE     ]
Poisson Model         .02775      .00478          5.800
Logarithmic Model     .00662      .00353          1.875
Total Effect          .03437      .00480          7.162
---------------------------------------------------------------
```

```
[EDUC    ]
Poisson Model         -.01387      .02154           -.644
Logarithmic Model     -.07911      .01564          -5.059
Total Effect          -.09297      .02188          -4.248
----------------------------------------------------------------
[HHNINC  ]
Poisson Model         -.55969      .27401          -2.043
Logarithmic Model     -.38658      .20545          -1.882
Total Effect          -.94628      .27328          -3.463
----------------------------------------------------------------
[FEMALE  ]
Poisson Model          .88928      .10319           8.617
Logarithmic Model     -.19527      .06757          -2.890
Total Effect           .69400      .09877           7.027
----------------------------------------------------------------
[HHKIDS  ]
Poisson Model         -.29986      .11095          -2.703
Logarithmic Model     -.19147      .08012          -2.390
Total Effect          -.49133      .11145          -4.409
----------------------------------------------------------------
[WORKING ]
Poisson Model         -.15448      .11895          -1.299
Logarithmic Model     -.25740      .08315          -3.095
Total Effect          -.41188      .11481          -3.587
================================================================
```

These are the average partial effects produced by the **PARTIALS** command.

```
----------------------------------------------------------------------
Partial Effects for Negative Binomial Model (Type=X)
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
----------------------------------------------------------------------
      AGE           .03661     .00522    7.02       .02639      .04684
      EDUC        -.10010     .02373    4.22      -.14661     -.05358
      HHNINC     -1.01158     .29156    3.47     -1.58302     -.44013
   *  FEMALE        .72792     .10522    6.92       .52169      .93415
   *  HHKIDS       -.50494     .11090    4.55      -.72231     -.28757
   *  WORKING      -.45640     .13015    3.51      -.71149     -.20131
----------------------------------------------------------------------
```

These are computed by **PARTIALS** using **; Means.**

```
----------------------------------------------------------------------
      AGE           .03437     .00480    7.17       .02497      .04377
      EDUC         -.09297     .02205    4.22      -.13619     -.04976
      HHNINC       -.94628     .27228    3.48     -1.47993     -.41262
      FEMALE        .69400     .09898    7.01       .50001      .88800
      HHKIDS       -.49133     .11155    4.40      -.70996     -.27271
      WORKING      -.41188     .11481    3.59      -.63691     -.18685
----------------------------------------------------------------------
```

The NBX model may be fit with different or overlapping variables in the Poisson and logarithmic models by using **; Rh2** to specify the logarithmic model.

POISSON ; Lhs = docvis ; Rhs = x
; Rh2 = one,hhkids,working
; Model = NBX ; Partials Effects $

```
-----------------------------------------------------------------------------
Negative Binomial - X Model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson Probability
Constant|    -.19674         .14130    -1.39  .1638     -.47369    .08021
     AGE|     .01173***      .00174     6.74  .0000      .00832    .01514
    EDUC|    -.01614*        .00828    -1.95  .0512     -.03237    .00009
  HHNINC|    -.28243***      .10170    -2.78  .0055     -.48176   -.08311
  FEMALE|     .31017***      .03699     8.39  .0000      .23767    .38267
  HHKIDS|    -.10756**       .04269    -2.52  .0117     -.19123   -.02390
 WORKING|    -.05043         .04534    -1.11  .2661     -.13930    .03844
        |Parameters of Logarithmic Model in NB-X
Constant|    1.60729***      .04316    37.24  .0000     1.52270   1.69187
  HHKIDS|    -.21691***      .05461    -3.97  .0001     -.32394   -.10988
 WORKING|    -.28383***      .05400    -5.26  .0000     -.38966   -.17799
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Analysis of Partial Effects in Two Part Negative Binomial Model
Expected Value of DOCVIS   at means of all variables =   2.6055
===============================================================
                      Effect  Standard Error    t ratio
---------------------------------------------------------------
[AGE     ]
Poisson Model          .03057      .00450          6.786
---------------------------------------------------------------
[EDUC    ]
Poisson Model        -.04206      .02152         -1.954
---------------------------------------------------------------
[HHNINC  ]
Poisson Model        -.73589      .26449         -2.782
---------------------------------------------------------------
[FEMALE  ]
Poisson Model          .80816      .09621          8.400
---------------------------------------------------------------
[HHKIDS  ]
Poisson Model        -.28026      .11128         -2.518
Logarithmic Model    -.27854      .07036         -3.959
Total Effect         -.55880      .11113         -5.028
---------------------------------------------------------------
[WORKING ]
Poisson Model        -.13139      .11826         -1.111
Logarithmic Model    -.36446      .07039         -5.178
Total Effect         -.49585      .11361         -4.365
===============================================================
```

E41.5.6 Canonical Negative Binomial Regression Model

Hilbe (2011) recommends an alternative form of the negative binomial that he labels the 'canonical negative binomial' model. The signature feature of the model is that it applies to a discrete random variable with a formal negative binomial distribution – it is not obtained by integrating heterogeneity out of a mixed distribution. Hence the name 'canonical' – it derives from first principles. The conditional (on x_i) density of the random variable is

$$f(y_i \mid \mathbf{x}_i) = \frac{\Gamma(y_i + \theta)\lambda_i^{y_i}(1-\lambda_i)^{\theta}}{\Gamma(y_i+1)\Gamma(\theta)},\ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

The conditional mean function for this model is

$$E[y_i \mid \mathbf{x}_i] = -\theta \frac{\lambda_i}{\lambda_i - 1}.$$

The resemblance to the more familiar NB2 model is only superficial. It can be seen from the conditional mean function that the parameters in the models are very different. A more transparent way to examine the difference is to examine the partial effects. In the NB2 model,

$$\partial \text{E}[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i = \lambda_i \boldsymbol{\beta}.$$

In the CNB model,

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = -\theta \frac{-\lambda_i}{(\lambda_i - 1)^2}\boldsymbol{\beta} = -E[y_i \mid \mathbf{x}_i]\left(\frac{\lambda_i}{\lambda_i - 1}\right)\boldsymbol{\beta}.$$

This is a completely different scaling of the parameter vector. The implication seems likely to be that the parameters themselves from the two models will differ substantially if, as is common, the differences tend to even out in the partial effects. We will explore this in an example below.

The canonical NB model is not a built in procedure in *LIMDEP*. However, it is a very straightforward application of the maximize command to obtain the estimates followed by **PARTIALS** and **SIMULATE** to obtain the partial effects and model simulations. The program below is written in the form of a template that requires only the specification of the dependent variable and the namelist containing the regressors. A substantive complication for this estimator is the starting values. The ordinary NB estimates might seem natural, but as the analysis above suggests, the parameters in the NB2 and the CNB models are likely to be quite different. Hilbe suggests -1 for the constant term, zeros for the slopes, and 2 for θ (i.e., .5 for $\alpha = 1/\theta$).

The procedure is generic save for a single line that is modified for the specific application

```
PROC = CNBModel(y,x) $
CALC          ; k = Col(x) $
? MAXIMIZE estimates the model parameters
MAXIMIZE      ; Start = -1,k_0,2
              ; Labels = b0,k_b,theta
              ; Fcn = bx = b0 + b1'x       |
                lambdai = Exp(bx)          |
                y*bx + theta*Log(1-lambdai) + Lgm(y+theta) - Lgm(y+1) - Lgm(theta) $
? PARTIALS computes the partial effects for the variables in the namelist
PARTIALS      ; Parameters = b
              ; Labels = b0,k_b,theta
              ; Covariance = varb
              ; Function = bx = b0 + b1'x |
                lambdai = Exp(bx)          |
                -theta*lambda / (lambdai-1)
              ; Effects: x ; Summary  $
? We compare the results to the NB2 model.  Partials are comparable APEs
NEGBIN        ; Lhs = y ; Rhs = one,x $
PARTIALS      ; Effects: x ; Summary $
ENDPROC $
```

To execute the procedure, we use the health care data, and commands

```
SAMPLE        ; All $
NAMELIST      ; x = age,educ,hhninc,female$
EXECUTE       ; Proc = CNBModel(docvis, x) $
```

The results are as follows: Notice that they begin with several warnings about the computation of the function. Unlike other models that we have examined thus far, this model does involve a computation that is quite likely to produce this result. One of the terms in the log likelihood is log(1-λ_i). The implication is that λ_i must be between zero and one. Since λ_i = exp($\boldsymbol{\beta}'\mathbf{x}_i$), there is no constraint that can be placed on the parameters that will enforce this boundary. It is not unlikely that for some observations, this error will occur. The solver will draw the iterations on the parameters away from these values as it gets closer to a solution.

```
  Error   590: Obs.=     1 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=     1 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=     1 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=    10 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=    93 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=    96 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus
Normal exit:  19 iterations. Status=0, F=    60207.36
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable                Function
Log likelihood function    -60207.36401
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =********* AIC/N =    4.407
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.23013***      .00867   -26.54  .0000     -.24712   -.21313
      B1|     .00272***   .9967D-04    27.28  .0000      .00252    .00291
      B2|    -.00435***      .00063    -6.87  .0000     -.00559   -.00311
      B3|    -.06643***      .00747    -8.89  .0000     -.08107   -.05179
      B4|     .03904***      .00220    17.77  .0000      .03474    .04335
   THETA|     .52343***      .00545    96.04  .0000      .51275    .53411
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

---------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .06788      .00324   20.97       .06153      .07422
      EDUC          -.10864      .01603    6.78      -.14007     -.07722
      HHNINC       -1.65817      .19003    8.73     -2.03061    -1.28572
   *  FEMALE         .91204      .05373   16.97       .80672     1.01736
---------------------------------------------------------------------
 (Intermediate results for Poisson regression omitted)
-----------------------------------------------------------------------------
Normal exit:  10 iterations. Status=0, F=    60164.22
-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                  DOCVIS
Log likelihood function    -60164.22014
Restricted log likelihood -103923.54929
Chi squared [   1 d.f.]     87518.65830
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .62857***      .05457    11.52  .0000      .52162    .73553
     AGE|     .02042***      .00070    29.07  .0000      .01904    .02179
    EDUC|    -.03539***      .00378    -9.36  .0000     -.04281   -.02798
  HHNINC|    -.48779***      .04520   -10.79  .0000     -.57637   -.39921
  FEMALE|     .32673***      .01588    20.58  .0000      .29561    .35784
        |Dispersion parameter for count data model
   Alpha|    1.90309***      .01984    95.94  .0000     1.86421   1.94197
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
Partial Effects for Loglinear, Exponential Mean
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
      AGE          .06514     .00246   26.44       .06031      .06996
      EDUC        -.11290     .01213    9.31     -.13668     -.08913
      HHNINC     -1.55610     .14504   10.73    -1.84037    -1.27183
   *  FEMALE      1.03372     .05259   19.66       .93065     1.13680
-----------------------------------------------------------------------
Maximum repetitions of PROC
```

We note, finally, a possible extension of the model. In the NB1 and NB2 formulations, we allow for heterogeneity in the scale parameter, θ. In particular, the generalized model specifies

$$\theta_i = \theta \exp(\boldsymbol{\delta}'\mathbf{z}_i).$$

It is straightforward to incorporate the same extension in the canonical model, as shown in the revised procedure below:

```
PROC = CNBModel(y,x,z) $
CALC        ; k = Col(x) ; m = Col(z) $
? MAXIMIZE estimates the model parameters
MAXIMIZE    ; Start = -1,k_0,2, m_0
            ; Labels = b0,k_b,theta,m_d
            ; Fcn = bx = b0+b1'x          |
              lambdai = Exp(bx)           |
              vh  = Exp(d1’z)             |
              y*bx + theta*vh*Log(1-lambdai)
              + Lgm(y+theta*vh) - Lgm(y+1) - Lgm(theta*vh) $
? PARTIALS computes the partial effects for the variables in the namelist
NAMELIST    ; xz = x,z $
PARTIALS    ; Parameters = b
            ; Labels = b0,k_b,theta,m_d
            ; Covariance = varb
            ; Function = bx = b0+b1'x  |
              lambdai = Exp(bx)        |
              vh = Exp(d1’z)           |
              -theta*vh*lambdai/(lambdai-1)
            ; Effects: xz ; Summary $
ENDPROC $
SAMPLE      ; All $
NAMELIST    ; z = hhkids $
NAMELIST    ; x = age,educ,hhninc,female $
EXEC        ; Proc = CNBModel(docvis,x,z) $
```

The results of the computation of this extended model are shown below.

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function     -60147.26561
Estimation based on N =  27326, K =   7
Inf.Cr.AIC  =********* AIC/N =    4.403
Model estimated: Jul 27, 2011, 21:47:21
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.20971***      .00879   -23.86  .0000     -.22693   -.19248
      B1|     .00233***      .00010    22.58  .0000      .00213    .00253
      B2|    -.00459***      .00063    -7.23  .0000     -.00583   -.00335
      B3|    -.06695***      .00747    -8.96  .0000     -.08159   -.05231
      B4|     .03939***      .00220    17.91  .0000      .03508    .04370
   THETA|     .55914***      .00670    83.40  .0000      .54600    .57228
      D1|    -.17037***      .01569   -10.86  .0000     -.20112   -.13961
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
---------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .05720      .00306   18.71       .05121      .06320
      EDUC         -.11267      .01579    7.14      -.14362     -.08172
      HHNINC      -1.64338      .18613    8.83     -2.00818    -1.27858
   *  FEMALE         .90849      .05293   17.16       .80474     1.01224
   *  HHKIDS        -.52822      .04754   11.11      -.62139     -.43504
---------------------------------------------------------------------
```

E42: Censoring, Truncation and Heterogeneity in Count Models

E42.1 Introduction

This chapter details several extended models for count data. The basic formulation is the *Poisson regression model*. For a discrete random variable, Y observed over a period of length T_i, and observed frequencies, y_i, i=1,...,n, where y_i is a nonnegative integer count, and regressors $\mathbf{x}_i$,

$$\text{Prob}(Y = y_i|\mathbf{x}_i) \quad = \quad \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i\,!},\ y_i = 0,1,...;\ \ \log \lambda_i = \boldsymbol{\beta}'\mathbf{x}_i.$$

In this model, λ_i is both the mean and variance of y_i;

$$E[y_i|\mathbf{x}_i] \quad = \quad \lambda_i.$$

The partial effects in this nonlinear regression model are,

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} \quad = \quad \lambda_i\boldsymbol{\beta}\ .$$

The *negative binomial regression model* is an extension of the Poisson regression model which allows the variance of the process to differ from the mean. An interpretation which fits well with several of the extensions considered in this chapter is that the negative binomial model results from the introduction of unobserved individual heterogeneity into the Poisson model. The model arises as a modification of the Poisson model in which the mean is μ_i, respecified so that

$$\log \mu_i \ = \ \log\lambda_i \ + \ w_i \ = \ \boldsymbol{\beta}'\mathbf{x}_i + w_i,$$

where $\exp(w_i)$ has a gamma distribution with mean 1.0 and variance α. This random variable y_i then has

$$\text{Var}[y_i \,|\mathbf{x}_i] \ = \ \text{E}[y_i|\mathbf{x}_i]\{1 \ + \ \alpha\text{E}[y_i|\mathbf{x}_i]\}.$$

The models presented in this chapter are extensions of the model with heterogeneity. Different treatments of the source of w_i produce different model specifications. The Poisson model considered above assumes that there is no unobserved 'heterogeneity' across individuals save for that measured in the covariates. The Poisson and negative binomial models can both be extended to allow for unobserved heterogeneity in the conditional mean function, of the form

$$\log\text{E}[y_i|\mathbf{x}_i] \ = \ \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i.$$

Thus, the unobserved individual heterogeneity enters in the form of a normally distributed disturbance. This is the formulation which gives rise to the random effects model in the panel data case, but the models described here apply simply to cross section data. The extension is made for both the Poisson and negative binomial models. Another method of introducing heterogeneity into the Poisson and negative binomial models is to allow random variation in the parameters. This extension allows a large amount of flexibility in the functional form.

E42.2 Censoring and Truncation

The tobit model is a standard tool for accounting for censoring in the linear regression model. Censoring is also observed in count data. Terza (1985) analyzed a survey of shoppers in the Atlanta SMSA who were asked, 'How many times have you been to shopping area X in the past thirty days?' with possible responses 0, 1, 2, and 3 or more. This is a direct counterpart to the tobit model, though in this instance the censoring is at the right of the distribution rather than the left at zero, as is common in the regression case. (See, also, Greene (2011) for an analysis of extramarital affairs self reported in a survey and Gurmu (1991) for a study of health care facility utilization.) Grogger and Carson (1991) analyzed a survey of the number of recreational fishing trips taken by a sample of Alaskan fisherman in which the sample was choice based so as to eliminate any individuals who reported zero trips. In this case, the distribution is truncated, rather than censored. These two cases illustrate direct counterparts to the tobit and truncated regression models.

E42.2.1 Commands for Censoring and Truncation

The Poisson model with right censoring is the Poisson model described earlier with the modification that for some positive integer C, all values of y_i greater than or equal to C are reported as C. Negative binomial models with censoring are obtained analogously by changing the functional form of the probability. Either model may be estimated for a *truncated* distribution, instead of a censored one. Suppose, for the present, that truncation is from below, at a value C. Then, the distribution of *y*s applies only to values *above C*.

Specifying Censoring in the Data

The Poisson regression model with *right censoring* is obtained by adding

; Limit = C

to the command. The censoring limit, '*C*' must be a positive integer. This specification dictates that values of the original variable have been right censored. That is, the observed *y* is the *minimum* of C and a latent Poisson variable, Y_i*. As such, the *largest* value in your sample will be C.

You can specify *left censoring*, instead, with

; Limit = C ; Maximum

to indicate that your observed dependent variable is y_i = Max[Y_i* ,C]. Then, the *smallest* value in your sample will be C.

Specifying a Truncated Distribution

In the right censored regression model, all values at or above a certain value are given that value. Thus, if the censoring is at three, all values of the 'true' y_i at or above three take the value three in the observed sample. With truncation, values in the sample only take values strictly above or below a given limit value. Thus, suppose the distribution of y_i is *left truncated* at one. Then, the observed sample will only take values 2, 3, ... To specify left truncation, use

; Limit = C
; Truncation

where *C* is the lower truncation point. Note that for left truncation, the *smallest* value in your sample will be *C* + 1.

Alternatively, the distribution may be *right truncated* at some *upper* limit, such as four. Then, the sample will only contain the values 0, 1, 2, and 3. Upper (right) truncation is requested by adding

; Limit = C
; Truncation ; Upper

to the command. The *largest* y_i in your sample will be *C* - 1, so if you request upper truncation, *C* must be greater than one. For example,

```
POISSON    ; Lhs = y
           ; Rhs = ...
           ; Limit = 0
           ; Truncation $
```

estimates a model for y_i = 1,2,....

NOTE: In all these formulations for censoring and truncation, *C* may be a fixed integer or a variable. When the censoring limit is a variable, it is also a censoring indicator. That is, the only way to tell if an observation is censored is to compare it to the censoring variable. The implication is that, for example, two observations can have the same value, say 10, and one will be censored and the other will not. The upshot is that when you use a variable censoring limit, that value must be greater than (or less than) the dependent variable for right (left) censored data sets. Your data may also contain a mix of lower, upper or uncensored observations. To specify this, add a second Rhs variable which takes the value -1 for lower censored data, 0 for uncensored observations and +1 for upper censored observations

E42.2.2 Results for the Models with Censoring and Truncation

The models with censoring and truncation are otherwise the same as those for the unmodified Poisson and negative binomial models. All options are the same as well, including fitted values, optimization options, restrictions, hypothesis tests, and so on. The output will contain notations in a few places to indicate the censoring or truncation, as shown in the example below.

NOTE: The fitted values for the dependent variable do not account for censoring or truncation, so the result should be interpreted as applying to the underlying distribution. For relatively simple problems, it is possible to manipulate the results to obtain the mean of the censored or truncated distribution. We return to this issue below.

NOTE: The computation of partial effects by the model command (with **; Partial Effects**) accounts for censoring or truncation when the model is specified with one or the other. The **PARTIAL EFFECTS** (or just **PARTIALS**) command uses the structural conditional mean, λ_i, without accounting for censoring or truncation.

The following illustrates estimation of a model with right censoring. The doctor visits data contain a fairly large number of extremely large values. About 10% of the observations are larger than 10, with the maximum well over 100. A tail this long probably stretches what could be expected of a Poisson model. To accommodate this, we have censored the data at 10 visits and reestimated the model. (Note, the data themselves need not actually be censored. Where we have specified a censoring limit of 10, internally, the program converts all values larger than 10 to 10.) The first results are for a base case Poisson model. We then fit a negative binomial (type 2) model.

```
SAMPLE      ; All $
REJECT      ; _groupti < 7 $
POISSON     ; Lhs = docvis
            ; Rhs = one,age,hhninc,educ,female,hhkids,married
            ; Partial Effects $
POISSON     ; Lhs = docvis
            ; Limit = 10
            ; Rhs = one,age,hhninc,educ,female,hhkids,married
            ; Partial Effects $
NEGBIN      ; Lhs = docvis
            ; Limit = 10
            ; Rhs = one,age,hhninc,educ,female,hhkids,married
            ; Partial Effects $
```

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable              DOCVIS
Log likelihood function    -22965.36559
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .62879***      .07114     8.84  .0000       .48935     .76823
     AGE|     .02271***      .00095    23.98  .0000       .02086     .02457
  HHNINC|    -.26608***      .04901    -5.43  .0000      -.36213    -.17002
    EDUC|    -.05835***      .00434   -13.46  .0000      -.06684    -.04985
  FEMALE|     .35718***      .01517    23.54  .0000       .32744     .38691
  HHKIDS|    -.06041***      .01748    -3.45  .0006      -.09467    -.02614
 MARRIED|     .06023***      .02094     2.88  .0040       .01920     .10127
--------+--------------------------------------------------------------------
Log likelihood function    -16627.22612
RIGHT Censored  Data: Threshold = 10.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .47430***      .07675     6.18  .0000       .32386     .62473
     AGE|     .01842***      .00104    17.71  .0000       .01638     .02046
  HHNINC|    -.18180***      .05356    -3.39  .0007      -.28676    -.07683
    EDUC|    -.04525***      .00462    -9.79  .0000      -.05431    -.03619
  FEMALE|     .33313***      .01672    19.93  .0000       .30037     .36589
  HHKIDS|    -.08628***      .01922    -4.49  .0000      -.12396    -.04860
 MARRIED|     .08787***      .02341     3.75  .0002       .04200     .13375
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Scale Factor for Marginal Effects  3.1340
--------+--------------------------------------------------------------------
        |     Partial     Standard            Prob.      95% Confidence
  DOCVIS|      Effect       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .07118***      .00301    23.63  .0000       .06528     .07709
  HHNINC|    -.83388***      .15371    -5.42  .0000     -1.13516    -.53261
    EDUC|    -.18286***      .01365   -13.40  .0000      -.20961    -.15611
  FEMALE|    1.12583***      .04841    23.25  .0000      1.03094    1.22072   #
  HHKIDS|    -.18806***      .05409    -3.48  .0005      -.29408    -.08205   #
 MARRIED|     .18494***      .06297     2.94  .0033       .06151     .30837   #
--------+--------------------------------------------------------------------
Scale Factor for Marginal Effects  2.5946
RIGHT Censored  Data: Threshold = 10.
--------+--------------------------------------------------------------------
     AGE|     .04779***      .00272    17.57  .0000       .04246     .05313
  HHNINC|    -.47168***      .13899    -3.39  .0007      -.74410    -.19927
    EDUC|    -.11741***      .01203    -9.76  .0000      -.14099    -.09383
  FEMALE|     .87376***      .04448    19.64  .0000       .78658     .96094   #
  HHKIDS|    -.22257***      .04919    -4.52  .0000      -.31898    -.12615   #
 MARRIED|     .22187***      .05735     3.87  .0001       .10948     .33427   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function    -11966.78137
Restricted log likelihood  -16627.22612
Chi squared [   1 d.f.]      9320.88949
Significance level               .00000
McFadden Pseudo R-squared      .2802900
Estimation based on N =   6209, K =   8
Inf.Cr.AIC  =23949.563 AIC/N =    3.857
Model estimated: Jul 29, 2011, 06:36:58
RIGHT Censored  Data: Threshold = 10.
NegBin form 2; Psi(i) = theta
Tests of Model Restrictions on Neg.Bin.
Model               Logl ChiSquared[df]
Poisson(b=0)   -24176.44  ******** [**]
Poisson        -16627.23   15098.4 [ 6]
Negative Bin.  -11966.78    9320.9 [ 1]
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .59206***      .16825     3.52  .0004      .26230    .92183
     AGE|     .02214***      .00232     9.54  .0000      .01760    .02669
  HHNINC|    -.28891**       .11596    -2.49  .0127     -.51619   -.06164
    EDUC|    -.06254***      .01016    -6.16  .0000     -.08244   -.04263
  FEMALE|     .43292***      .04080    10.61  .0000      .35294    .51289
  HHKIDS|    -.10911**       .04359    -2.50  .0123     -.19455   -.02367
 MARRIED|     .14036**       .05662     2.48  .0132      .02939    .25133
        |Dispersion parameter for count data model
   Alpha|    1.75668***      .04666    37.65  .0000     1.66523   1.84812
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   2.5842
Scale Factor for Marginal Effects  1.8946
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .04195***      .00456     9.21  .0000      .03302    .05089
  HHNINC|    -.54736**       .21999    -2.49  .0128     -.97855   -.11618
    EDUC|    -.11848        3.49709     -.03  .9730    -6.97264   6.73568
  FEMALE|    1.33039        2.63995      .50  .6143    -3.84381   6.50459   #
  HHKIDS|    -.32999        1.31009     -.25  .8011    -2.89771   2.23774   #
 MARRIED|     .40972         .68777      .60  .5514     -.93829   1.75772   #
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E42.2.3 Technical Details on Censoring and Truncation

The Poisson model with right censoring is the Poisson model described earlier with the modification that for some integer $C > 0$, all values of y_i greater than or equal to C are reported as C. Define a 'latent' variable, Y^* which is the underlying Poisson variable;

$$\text{Prob}(Y_i^* = j) = \exp(-\lambda_i)\,\lambda_i^{\,j} / j!,\ \ \log \lambda_i = \boldsymbol{\beta}'\mathbf{x}_i.$$

The observed variable is

$$y_i = \text{Min}[Y_i^*, C].$$

Then, $\text{Prob}[y_i = j] = \text{Prob}[Y_i^* = j] = \text{ if } y_i < C$

and $\text{Prob}[y_i = C] = \text{Prob}[Y_i^* \geq C]$

$$= 1 - \text{Prob}[Y_i^* < C]$$

$$= 1 - \sum_{j=0}^{c-1} \text{Prob}[Y_i^* = j].$$

The model with left censoring is obtained by reversing the direction of the inequality in the preceding. Thus, with left censoring,

$$\text{Prob}[y_i = j] = \text{Prob}[Y_i^* = j] \text{ if } y_i > C$$

and $\text{Prob}[y_i = C] = \text{Prob}[Y_i^* \leq C]$

$$= \sum_{j=0}^{c} \text{Prob}[Y_i^* = j] \text{ otherwise.}$$

Negative binomial models with censoring are obtained analogously by changing the functional form of the probability.

For the censored distribution, the contribution of an observation to the log likelihood is

$$\log L_i = \delta_i \log \text{Prob}[Y_i^* = y_i] + (1 - \delta_i)\log\left\{1 - \sum_{j=0}^{C-1} \text{Prob}[Y_i = j]\right\}$$

if the observation is censored at the right and

$$\log L_i = \delta_i \log \text{Prob}[Y_i^* = y_i] + (1 - \delta_i)\log\left\{\sum_{j=0}^{C} \text{Prob}[Y_i = j]\right\}$$

if the observation is censored at the left,

where δ_i = 1 if the observation is not censored and 0 if it is.

To form the gradients of the log likelihood, we denote Prob[Y_i* = j] = Pj and make use of the result

$$\partial \text{Prob}[Y_i^* = j]/\partial\boldsymbol{\beta} = \text{Prob}[[Y_i^* = j] \times \partial \log \text{Prob}[[Y_i^* = j]/\partial\boldsymbol{\beta}$$

$$= P_j(j - \lambda_i)\mathbf{x}_i$$

Combining terms, then, for the Poisson model with right censoring

$$\frac{\partial \log L_i}{\partial\boldsymbol{\beta}} = \left[\delta_i(y_i - \lambda_i) - \frac{1-\delta_i}{1-\sum_{j=0}^{C-1} P_j} P_j(j-\lambda_i)\right]\mathbf{x}_i$$

For left censoring, the only change is the summation in the denominator in the second term. The BHHH estimator based on first derivatives is used for the estimator of the asymptotic covariance matrix. The analogous results for the negative binomial model are obtained by changing the individual term in the square brackets. The necessary results appear below.

If y is right censored (no values larger than C), then

$$E[y|\mathbf{x}] = \sum_{\text{j=0}}^{\text{C}-1} jP_j + \text{C}\left[1 - \sum_{\text{j=0}}^{\text{C}-1} P_j\right].$$

Another form of this which shows the effect of the censoring on the conditional mean is

$$E[y|\mathbf{x}] = \lambda - \sum_{j=C}^{\infty} (j-C)P_j.$$

Since this involves an infinite sum, it is not useable for computations as is. For computation of the conditional mean and the marginal effects, we use, instead,

$$E[y|\mathbf{x}] = \text{C} - \sum_{j=0}^{C-1} P_j\,(C - j).$$

(For convenience, the sum may run to C, as the last term is zero.) Then, the marginal effects are

$$\boldsymbol{\delta} = \frac{\partial E[y\,|\,\mathbf{x}]}{\partial\mathbf{x}} = \left[\sum_{j=0}^{C-1} P_j(j-C)\frac{\partial \log P_j}{\partial\mathbf{x}}\right]\boldsymbol{\beta}$$

$$= \left[\sum_{j=0}^{C-1} P_j(j-C)(j-\lambda)\right]\boldsymbol{\beta}$$

for the Poisson model, and

$$\boldsymbol{\delta} = \left[\sum_{j=0}^{C-1} P_j(j-C)(\theta(1-u) - ju)\right]\boldsymbol{\beta}$$

for the negative binomial model. Analytic derivatives of these expressions are used for the delta method to compute the standard errors for the marginal effects. Denote by $\boldsymbol{\gamma}'$ the parameter vector, either $\boldsymbol{\beta}'$ for the Poisson model or $[\boldsymbol{\beta}',\theta]$ for the negative binomial model. Then, the matrix of derivatives needed for the asymptotic covariance matrix is

$$\frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\gamma}'} = \left[\sum_{j=0}^{C-1} P_j(j-C)\frac{\partial \log P_j}{\partial \mathbf{x}}\right][\mathbf{I} \mid \mathbf{0}]$$

$$+\,\boldsymbol{\beta}\left[\sum_{j=0}^{C-1} P_j(j-C)\left(\frac{\partial \log P_j}{\partial \mathbf{x}}\frac{\partial \log P_j}{\partial \boldsymbol{\gamma}'}+\frac{\partial^2 \log P_j}{\partial \mathbf{x}\partial \boldsymbol{\gamma}'}\right)\right]$$

If y is censored at the left (no values smaller than C), then the conditional mean is just

$$E[y|\mathbf{x}] = \lambda + \sum_{j=0}^{C} (C-j)P_j.$$

so the marginal effects are

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \left[\lambda + \sum_{j=0}^{C} P_j(C-j)\frac{\partial \log P_j}{\partial \mathbf{x}}\right]\boldsymbol{\beta}$$

The remaining expressions used for analyzing the marginal effects are modified accordingly. With the vector of probabilities in hand, these are straightforward to compute, say at the means of a set of regressors.

Suppose, for the present, that truncation is from below, at a value C. Then, the distribution of ys applies only to values above C – it is left truncated. Thus,

$$\text{Prob}[y_i = j \mid y_i > C] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}/y_i!}{\text{Prob}[y_i > C]}, \text{ for } y_i = C+1, C+2, \ldots.$$

For computational purposes, we use $\text{Prob}[y_i > C] = 1 - \text{Prob}[y_i \le C]$, so the manipulable form of the Poisson distribution is

$$\text{Prob}[y_i = j \mid y_i > C] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}/y_i!}{1-\sum_{j=0}^{C}\exp(-\lambda_i)\lambda_i^{j}/j!}, \text{ for } y_i = C+1, C+2, \ldots.$$

The negative binomial model is formed likewise. Truncation may also be from above, in which case,

$$\text{Prob}[y_i = j \mid y_i < C] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}/y_i!}{\text{Prob}[Y < C]}$$

$$= \frac{\exp(-\lambda_i)\lambda_i^{y_i}/y_i!}{\sum_{j=0}^{C-1}\exp(-\lambda_i)\lambda_i^{j}/j!}, \text{ for } y_i = 0,1,\ldots,C-1$$

for the Poisson model and likewise for the negative binomial. The log likelihood function is the sum of the probabilities. For left truncation (that is, for Y greater than C)

$$\log L_i \;=\; \log \text{Prob}[Y_i = y_i] \;-\; \log\left(1-\sum\nolimits_{j=0}^{C} P_j\right)$$

For right truncation, (that is, if the distribution restricts Y to be less than C),

$$\log L_i \;=\; \log \text{Prob}[Y_i = y_i] \;-\; \log\left(\sum\nolimits_{j=0}^{C-1} P_j\right).$$

The conditional mean functions are considerably more involved in the truncation case. For left truncation,

$$E[y|\mathbf{x}, y > C] \;=\; \frac{\sum_{j=C+1}^{\infty} jP_j}{1-\sum_{j=0}^{C} P_j}$$

By adding and subtracting a term in the numerator, this can be written as

$$E[y|\mathbf{x}, y > C] \;=\; \frac{\lambda \;-\; \sum_{j=0}^{C} jP_j}{1-\sum_{j=0}^{C} P_j}$$

With some algebra (omitted), the marginal effects in this model can be written

$$\boldsymbol{\delta} \;=\; \frac{\partial E[y\,|\,\mathbf{x}, y \;>\; C]}{\partial \mathbf{x}} \;=\; \left(\frac{1 \;-\; \sum_{j=0}^{C}(j - E[y\,|\,\mathbf{x}, y > C])P_j \,\dfrac{\partial \log P_j}{\partial \lambda}}{1 \;-\; \sum_{j=0}^{C} \;P_j}\right)\lambda\,\boldsymbol{\beta}$$

For right truncation,

$$\boldsymbol{\delta} \;=\; \frac{\partial E[y\,|\,\mathbf{x}, y \;<\; C]}{\partial \mathbf{x}} \;=\; \left(\frac{\sum_{j=0}^{C-1}(j - E[y\,|\,\mathbf{x}, y < C])P_j \,\dfrac{\partial \log P_j}{\partial \lambda}}{\sum_{j=0}^{C-1} \;P_j}\right)\lambda\,\boldsymbol{\beta}$$

The functions differ between the Poisson and negative binomial models in the derivative term in the numerator of the scale,

$\partial \log P_j/\partial\lambda, \;=\; (j/\lambda - 1)$ for the Poisson model,

$\partial \log P_j/\partial\lambda \;=\; [ju/L \;-\; \theta(1-u)/\lambda]$ for the negative binomial model.

Standard errors based on the delta method are tedious but follow the same computations as shown earlier. For brevity, they are omitted here.

E42.3 Endogenous Truncation – On Site Sampling

Shaw (1988) examined the Poisson regression model in the context of onsite sampling. This is a type of truncation in that if the count is observed on site, it must equal at least one, and the truncation of the zero observations is a feature of the sampling mechanism. Shaw's important result on the density of the observed count under this assumption is

$$p(y_i \mid \mathbf{x}_i, y_i > 0) = \frac{\exp(-\lambda_i)\lambda_i^{y_i-1}}{(y_i-1)!},\ y_i = 1, 2, \ldots,\ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i).$$

This random variable has mean function $\lambda_i + 1$ and variance λ_i. It can be seen that the model can be estimated and analyzed as a Poisson model simply in terms of $w_i = y_i - 1$. The model continues to display (essentially) the equidispersion feature of the Poisson model. Englin and Shonkwiler (1992) proposed an extension of the Shaw model for the negative binomial distribution. The density for this random variable is

$$p(y_i \mid x_i, y_i > 0) = \frac{y_i\Gamma(y_i + 1/\alpha_i)\alpha_i^{y_i}\lambda_i^{y_i-1}[1+\alpha_i\lambda_i]^{-(y_i+1/\alpha_i)}}{\Gamma(y_i+1)\Gamma(1/\alpha_i)}.$$

The conditional mean and variance in the Englin and Shonkwiler's variant of the negative binomial model are

$$\mathrm{E}[y_i|\mathbf{x}_i] = \lambda_i + 1 + \alpha_i\lambda_i$$

and

$$\mathrm{Var}[y_i|\mathbf{x}_i] = \lambda_i + \alpha_i(1 + \lambda_i + \alpha_i\lambda_i).$$

We will allow for additional heterogeneity in the model by parameterizing α_i as

$$\alpha_i = \exp(\boldsymbol{\delta}'\mathbf{z}_i).$$

The model is requested by using the model command

```
NEGBIN    ; Lhs = dependent variable
          ; Rhs = independent variables, including one
          ; Model = NBE $
```

The extended specification for α_i is requested with

; Hfn = list of variables in z (not including one).

Results for this model are indicated by an identifier for the model, but appear otherwise as a variant of the negative binomial model shown in the previous chapter. An example below.

```
-----------------------------------------------------------------------------
NegBin with Endogenous Stratification
Dependent variable                DOCVIS
Log likelihood function      -5852.00081
Restricted log likelihood    -8676.55824
Chi squared [   1 d.f.]       5649.11487
Significance level                .00000
Tests of Model Restrictions on Neg.Bin.
Model                Logl ChiSquared[df]
Poisson(b=0)    -8913.52  ******** [**]
Poisson         -8676.56     473.9 [ 3]
Negative Bin.   -5852.00    5649.1 [ 1]
--------+--------------------------------------------------------------------
```

It should be noted, as in Shaw's case, this model is not the same as the truncated at zero version of the negative binomial model. If the data generating process is not consistent of the model, the results will suggest that. For example, the first result below show the NBE model applied to the positive observations in the 1994 wave of the health care panel show the value of α to be consistent with a Poisson model while the actual negative binomial model shown second is consistent with other results that have suggested the overdispersion in the data.

```
(NBE results)
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -4.61983       15.94205     -.29  .7720   -35.86567   26.62600
     AGE|     .01514***      .00178     8.49  .0000      .01164     .01864
    EDUC|    -.04057***      .00969    -4.19  .0000     -.05956    -.02159
  HHNINC|    -.36742***      .10950    -3.36  .0008     -.58203    -.15281
        |Dispersion parameter for count data model
   Alpha|    449.096       7165.590      .06  .9500  -13595.202  14493.394
--------+--------------------------------------------------------------------
(Truncated NB Results
Log likelihood function     -5848.91750
--------+--------------------------------------------------------------------
Constant|    1.44839***      .15690     9.23  .0000     1.14087    1.75591
     AGE|     .01538***      .00198     7.77  .0000      .01150     .01927
    EDUC|    -.04127***      .01066    -3.87  .0001     -.06216    -.02038
  HHNINC|    -.37279***      .12081    -3.09  .0020     -.60958    -.13601
        |Dispersion parameter for count data model
   Alpha|    1.17471***      .07414    15.84  .0000     1.02939    1.32002
--------+--------------------------------------------------------------------
```

E42.4 Unobserved Heterogeneity

This section will describe two models for unobserved heterogeneity in the Poisson and negative binomial regression models. The techniques are applied to an example in Greene (2003).

E42.4.1 Latent Heterogeneity in Poisson and Negative Binomial Models

The Poisson and negative binomial models are modified to allow individual heterogeneity. The Poisson model is modified so that

$$y|\mathbf{x},\varepsilon \sim \text{ Poisson with mean } \lambda|\mathbf{x},\varepsilon = \exp(\boldsymbol{\beta}'\mathbf{x} + \varepsilon) \text{ where } \varepsilon|\mathbf{x} \sim N[0,\sigma^2].$$

This is an alternative to the negative binomial model for unobserved heterogeneity in the count data model. The negative binomial model arises if ε has a log-gamma density, that is, $u = \exp(\varepsilon)$ has the gamma density with mean one. The unconditional variance of y can be obtained as

$$\text{Var}[y|\mathbf{x}] = E[\text{Var}[y|\mathbf{x},\varepsilon]] + \text{Var}[E[y|\mathbf{x},\varepsilon]].$$

Conditioned on ε, y has mean and variance equal to $\exp(\boldsymbol{\beta}'\mathbf{x})\times\exp(\varepsilon)$. The second term has a lognormal distribution. Using properties of the lognormal distribution, we find

$$\begin{aligned}
E[y|\mathbf{x}] &= \exp(\boldsymbol{\beta}'\mathbf{x}) \times \exp(\tfrac{1}{2}\sigma^2)\\
\text{Var}[y|\mathbf{x}] &= \exp(\boldsymbol{\beta}'\mathbf{x}) \times \exp(\tfrac{1}{2}\sigma^2) + [\exp(\boldsymbol{\beta}'\mathbf{x})]^2 \times [\exp(2\sigma^2) - \exp(\sigma^2)]\\
&= E[y] \times \{1 + E[y](\exp(\sigma^2) - 1)\}.
\end{aligned}$$

This does induce overdispersion, as might be expected. If $\sigma^2 \rightarrow 0$, $E[y|\mathbf{x}]$ reduces to the Poisson mean, and $\text{Var}[y|\mathbf{x}] = E[y|\mathbf{x}]$. Therefore a positive σ is the difference between this model and the Poisson. Essentially the same result is obtained if $E[\exp(\varepsilon)]$ is normalized to $-\tfrac{1}{2}\sigma^2$, so that $E[u] = 1$, as in the negative binomial case. In this case, the constant term in the regression must be adjusted. In principle, a model without heterogeneity can be obtained by setting σ to zero. But, this is not a well defined hypothesis for likelihood based tests, so we use a Vuong test, instead. The statistic and its implication are presented in the output.

Request this model with

POISSON ; Lhs = ... ; Rhs = ... ; Heterogeneity $

All other options for the Poisson model are available, including controls for the optimization, keeping residuals and predictions, and marginal effects and so on. Starting values for the iterations are the unconstrained Poisson estimates, with a moment estimator of σ. You may provide your own starting values with **; Start = ...** You may also impose constraints with **; Rst** and **; CML:**. Other options as usual are available, such as **; Par** for keeping ancillary parameters, etc.

Predicted values for this model are

$$
\begin{aligned}
E^*[y|\mathbf{x}] &= \int_{-\infty}^{\infty} E[y|\mathbf{x},\varepsilon]\, f(\varepsilon)\, d\varepsilon \\
&= E_{\varepsilon}\,[\, E[y|\mathbf{x},\varepsilon]] \\
&= \int_{-\infty}^{\infty} (1/\sigma)\phi(\varepsilon/\sigma)\exp(\boldsymbol{\beta}'\mathbf{x} + \varepsilon)d\varepsilon.
\end{aligned}
$$

These are requested with **; Keep = name** and **; List** as usual. Residuals kept with **; Res = name** are computed as y-E*[y]. Other saved results are matrices *b* and *varb* as usual, scalars *nreg*, *kreg*, *logl*, and *s* which contains the estimate of σ. The *Last Model* parameters are *b_name* for the Rhs variables.

The same heterogeneity model can be extended to the negative binomial regression. Since the negative binomial model can be interpreted as a Poisson model with gamma heterogeneity, this new variant is likely to be problematic, as it adds heterogeneity to a model which already accommodates heterogeneity – see the example below. Still, if the underlying population is believed to be negative binomial to start with, this model allows normal heterogeneity to be added on to that. The structure is

$$
\begin{aligned}
P(y|\mathbf{x},\varepsilon) &= \{\Gamma(\theta+y)/[\Gamma(\theta)y!]\}\,(u|\varepsilon)^{\theta}\,[1-(u|\varepsilon)]^{y} \\
u|\varepsilon &= \theta / (\theta + \lambda|\varepsilon) \\
\lambda|\mathbf{x},\varepsilon &= \exp(\,\boldsymbol{\beta}'\mathbf{x} + \varepsilon) \\
\varepsilon|\mathbf{x} &\sim N[0,\sigma^2].
\end{aligned}
$$

The unconditional distribution is difficult to derive, and is evaluated by Hermite quadrature, instead.

The command for this model is

NEGBIN ; Lhs = ... ; Rhs = ... ; Heterogeneity $

Other aspects of this model are the same as those for the Poisson model.

E42.4.2 Applications

The foregoing are applied to the health care, with both the Poisson and negative models. The base models are included for comparison. The first set of estimates is for the Poisson model. The starting value for the heterogeneity is shown with the initial estimates. The Poisson model with heterogeneity is an alternative to the negative binomial model shown below.

```
-----------------------------------------------------------------------------
Unrestricted Poisson Regression Start Value
Dependent variable                  DOCVIS
Log likelihood function     -31890.63195
Estimation based on N =   3377, K =   7
Inf.Cr.AIC  =63795.264 AIC/N =    18.891
Estd. s for heterogeneity =       .59862
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .92173***     .06876    13.41  .0000      .78697   1.05650
     AGE|     .01693***     .00090    18.90  .0000      .01518    .01869
  HHNINC|    -.40982***     .04784    -8.57  .0000     -.50359   -.31605
    EDUC|    -.02982***     .00435    -6.86  .0000     -.03834   -.02131
  FEMALE|     .40223***     .01821    22.09  .0000      .36654    .43792
  HHKIDS|    -.16138***     .02239    -7.21  .0000     -.20526   -.11750
 MARRIED|     .03198        .02304     1.39  .1651     -.01317    .07713
--------+--------------------------------------------------------------------
Line search at iteration 21 does not improve fn. Exiting optimization.
-----------------------------------------------------------------------------
Poisson Model with Normal Heterogeneity
Dependent variable                  DOCVIS
Log likelihood function      -8026.16404
Restricted log likelihood   -31890.63195
Chi squared [   1 d.f.]      47728.93581
Mean of LHS Variable =           3.78294
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson Probability
Constant|     .01745        .18133      .10  .9234     -.33795    .37284
     AGE|     .01799***     .00248     7.27  .0000      .01314    .02285
  HHNINC|    -.37714***     .13074    -2.88  .0039     -.63339   -.12090
    EDUC|    -.02716**      .01150    -2.36  .0182     -.04971   -.00461
  FEMALE|     .57068***     .05004    11.40  .0000      .47261    .66876
  HHKIDS|    -.27234***     .05940    -4.58  .0000     -.38876   -.15592
 MARRIED|     .07941        .06408     1.24  .2153     -.04618    .20500
        |Standard Deviation of Heterogeneity
   Sigma|    1.28080***     .02146    59.68  .0000     1.23874   1.32286
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
Neg.Bin. Model with Normal Heterogeneity
Dependent variable                  DOCVIS
Log likelihood function     -7955.35165
Restricted log likelihood  -31890.63195
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Parameters of Poisson Probability
Constant|     .82348***      .18292     4.50  .0000      .46496   1.18200
     AGE|     .01755***      .00247     7.12  .0000      .01272    .02239
  HHNINC|    -.38946***      .13449    -2.90  .0038    -.65306   -.12586
    EDUC|    -.03254***      .01177    -2.77  .0057    -.05560   -.00947
  FEMALE|     .46983***      .05271     8.91  .0000      .36652    .57314
  HHKIDS|    -.19711***      .05910    -3.33  .0009    -.31295   -.08127
 MARRIED|     .04231         .06508      .65  .5156    -.08524    .16986
        |Overdispersion parameter in NegBin
   Alpha|    1.53709***      .07616    20.18  .0000     1.38781   1.68637
        |Standard Deviation of Heterogeneity
   Sigma|     .35507***      .06128     5.79  .0000      .23496    .47519
--------+-------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Estimated value of E[y|x] computed at the
means is  4.22499.
--------+-------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
     AGE|     .07602***      .01050     7.24  .0000      .05545    .09660
  HHNINC|   -1.59343***      .55345    -2.88  .0040   -2.67816   -.50869
    EDUC|    -.11475**       .04844    -2.37  .0178    -.20969   -.01981
  FEMALE|    2.41113***      .21410    11.26  .0000     1.99150   2.83076
  HHKIDS|   -1.15063***      .25190    -4.57  .0000   -1.64436   -.65691
 MARRIED|     .33550         .27082     1.24  .2154    -.19530    .86630
--------+-------------------------------------------------------------------
 (Negative Binomial)
Estimated value of E[y|x] computed at the
means is  3.81318.
--------+-------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
     AGE|     .06694***      .01021     6.56  .0000      .04694    .08694
  HHNINC|   -1.48509***      .52004    -2.86  .0043   -2.50435   -.46582
    EDUC|    -.12407***      .04551    -2.73  .0064    -.21327   -.03486
  FEMALE|    1.79155***      .21395     8.37  .0000     1.37222   2.21088
  HHKIDS|    -.75161***      .22182    -3.39  .0007   -1.18637   -.31684
 MARRIED|     .16133         .24700      .65  .5137    -.32278    .64545
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

E42.4.3 Random Constant Poisson Regression

The next section describes estimation of a count data model with random parameters. The models of heterogeneity considered here can be treated and estimated as special cases. In particular, we specify

$$y_i|\mathbf{x}_i,\varepsilon_i \sim \text{Poisson } \lambda_i$$

$$E[y_i|\mathbf{x}_i,\varepsilon_i] = \lambda_i = \lambda|\mathbf{x},\varepsilon = \exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x})$$

$$\alpha_i = \alpha + \varepsilon_i \text{ where } \varepsilon|\mathbf{x} \sim N[0,\sigma^2].$$

This is a special case of the random parameters model, that is estimated by maximum simulated likelihood, rather than by quadrature as are the models above.

The following demonstrates the alternative methods of fitting the Poisson model with normally distributed heterogeneity. (The negative binomial model could be estimated this way as well.) The commands are discussed in more detail below.

```
POISSON    ; Lhs = docvis
           ; Rhs = one,age,hhninc,educ,female,hhkids,married
           ; Rpm ; Fcn = one(n)
           ; Pts = 125 ; Halton
           ; Partial Effects $
```

```
-----------------------------------------------------------------------------
Poisson  Regression Start Values for DOCVIS
Dependent variable                 DOCVIS
Log likelihood function    -31890.63195
Estimation based on N =   3377, K =   7
Inf.Cr.AIC  =63795.264 AIC/N =   18.891
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01693***      .00090    18.90  .0000      .01518    .01869
  HHNINC|   -.40982***      .04784    -8.57  .0000     -.50359   -.31605
    EDUC|   -.02982***      .00435    -6.86  .0000     -.03834   -.02131
  FEMALE|     .40223***      .01821    22.09  .0000      .36654    .43792
  HHKIDS|   -.16138***      .02239    -7.21  .0000     -.20526   -.11750
 MARRIED|     .03198         .02304     1.39  .1651     -.01317    .07713
Constant|     .92173***      .06876    13.41  .0000      .78697   1.05650
--------+--------------------------------------------------------------------
Normal exit:  33 iterations. Status=0, F=    8044.167

-----------------------------------------------------------------------------
Random Coefficients  Poisson  Model
Dependent variable                 DOCVIS
Log likelihood function     -8044.16703
Restricted log likelihood  -31890.63195
Sample is  1 pds and   3377 individuals
POISSON regression model
Simulation based on 125 Halton draws
```

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
AGE	.02116***	.00085	24.96	.0000	.01950	.02282
HHNINC	-.46239***	.04793	-9.65	.0000	-.55634	-.36845
EDUC	-.03983***	.00438	-9.10	.0000	-.04841	-.03126
FEMALE	.53115***	.01786	29.75	.0000	.49615	.56614
HHKIDS	-.10946***	.02039	-5.37	.0000	-.14942	-.06950
MARRIED	-.06366***	.02133	-2.98	.0028	-.10546	-.02186
	Means for random parameters					
Constant	.12623*	.06969	1.81	.0701	-.01036	.26283
	Scale parameters for dists. of random parameters					
Constant	1.37409***	.01077	127.60	.0000	1.35298	1.39519

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point 1.6866
Scale Factor for Marginal Effects 1.6866

DOCVIS	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence Interval	
AGE	.03569***	.90192	18.58	.0000	.03192	.03945
HHNINC	-.77989***	-.20566	-7.75	.0000	-.97725	-.58253
EDUC	-.06718***	-.45833	-5.96	.0000	-.08927	-.04510
FEMALE	.89585***	.24615	14.48	.0000	.77462	1.01708
HHKIDS	-.18462***	-.04243	-4.63	.0000	-.26281	-.10643
MARRIED	-.10737***	-.04515	-2.92	.0035	-.17945	-.03529

z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

(Poisson)

DOCVIS	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval		
AGE	.06406***	.00344	18.64	.0000	.05732	.07080	
HHNINC	-1.55034***	.18150	-8.54	.0000	-1.90608	-1.19460	
EDUC	-.11282***	.01647	-6.85	.0000	-.14510	-.08054	
FEMALE	1.51309***	.06861	22.05	.0000	1.37862	1.64756	#
HHKIDS	-.59474***	.08051	-7.39	.0000	-.75254	-.43694	#
MARRIED	.12011	.08592	1.40	.1621	-.04829	.28852	#

E42.4.4 Technical Details

Parameters of the heterogeneity model are estimated by maximum likelihood. The likelihood function is formed as follows: The conditional Poisson density for the observed y is

$$P(y|\varepsilon) = \exp[-E(y|\varepsilon)] \times [E(y|\varepsilon)]^y / y!$$

The unconditional probability is found by integrating ε out of this expression;

$$P(y) = E_\varepsilon[P(y|\varepsilon)] = \int_{-\infty}^{\infty} P(y|\varepsilon)\, f(\varepsilon)\, d\varepsilon$$

$$= \int_{-\infty}^{\infty} P(y\,|\,\varepsilon)\left(\frac{1}{\sigma}\right)\phi\left(\frac{\varepsilon}{\sigma}\right)d\sigma$$

where $\phi(.)$ is the standard normal PDF. This function and its derivatives are evaluated by Hermite quadrature to maximize the log likelihood, which is

$$\text{Log } L = \sum_{i=1}^{n} \log P(y_i).$$

Details on using Hermite quadrature to evaluate log likelihoods of this form are given in Section R23.3.1.

The hypothesis $\sigma = 0$, which would produce the model without heterogeneity, is not well defined – the restricted value is on the boundary of the parameter space. An alternative statistic that can be used for such a test is the Vuong statistic. Define

$$m_i = \log(P_i|H_0 \,/\, P_i|H_1)$$

Thus, m_i is the ratio of the logs of the fitted probabilities for the ith observation under the null and alternative hypotheses. (Minus twice $\overline{m}$ is the likelihood ratio statistic for testing H_0 against a broader alternative H_1. But, that is not the case here.) The test statistic is the standard measure for testing whether a mean is zero,

$$V = \sqrt{n}\;\; \overline{m} \,/\, s_m\,.$$

The limiting distribution of V is normal (0,1). Large values (greater than 2.0) favor H_0; small values (less than -2.0) favor H_1. The intermediate values are inconclusive. The Vuong statistic is reported for several models, including the heterogeneity models shown in the examples below.

E42.5 Heterogeneity in the Form of Random Parameters

The models described above are equivalent to the following formulation in which '*i*' indexes individuals:

$$\begin{aligned}
\beta_{1i} &= \beta_1 + v_{1i} \\
v_i &\sim N[0,\sigma^2] \\
\beta_{ki} &= \beta_k,\ k = 2,...,K \\
\boldsymbol{\beta}_i &= [\beta_{1i}, \beta_{2i}, ..., \beta_{Ki}]' \\
\mathbf{x}_i &= [1, x_{2i}, x_{3i},..., x_{Ki}]' \\
\lambda_i|v_i &= \exp(\boldsymbol{\beta}_i'\mathbf{x}_i) \\
P(y_i|v_i) &= \text{Poisson or negative binomial probability conditioned on } v_i.
\end{aligned}$$

This has reformulated the heterogeneity model as a model with a randomly distributed constant term. The correct log likelihood for this model is obtained by integrating out the heterogeneity term;

$$\log L = \sum_{i=1}^{n} \log \int_{v_{1i}} g(v_{1i}) P(y_i \mid \mathbf{x_i}, v_{1i}) dv_{1i}$$

The preceding applications have made this feasible by approximating the integration with Hermite quadrature (assuming that v_{1i} is normally distributed). (Note that in the negative binomial case with unit mean gamma heterogeneity, the integral has a closed form, which we treated in Section E41.4.5) The approximate log likelihood that was maximized in the previous section is

$$\log L_H = \sum_{i=1}^{n} \log \sum_{h=1}^{H} P(y_i \mid \mathbf{x}_i, v_{1h}) w_h$$

where v_{1h} and w_h are the nodes and weights for the Hermite quadrature.

An alternative approach to maximizing the log likelihood is integration of the simulated log likelihood – see Chapter R24. For the model examined here, the simulated log likelihood function is

$$\log L_S = \sum_{i=1}^{n} \log \frac{1}{R} \sum_{h=1}^{H} P(y_i \mid \mathbf{x}_i, v_{1ir})$$

where v_{1ir} is the *r*th of *R* simulated draws from the distribution of v_{1i}. With a sufficient number of draws, *R*, the estimator converges to the true MLE. This 'random constant term' approach is equivalent to the heterogeneity models in the previous section, though it uses a different method of approximating the log likelihood. Before proceeding to the more general formulation, we will illustrate this particular model with the data described below.

NOTE: Random parameter models are often associated with analysis of panel data. But, this is one of many models in *LIMDEP* that allow random parameter models in a cross section setting.

The random parameters model may be extended to the full parameter vector. We allow for a general model in which some parameters are random and others are not. Also, several extensions of the model are added at this point. The structure of the random parameters model is

$\boldsymbol{\beta}_{1i}$ = $\boldsymbol{\beta}_1$ (K_1 nonrandom parameters)

$\mathbf{x}_{1i}$ = variables multiplied by $\boldsymbol{\beta}_{1i}$

$\boldsymbol{\beta}_{2i}$ = $\boldsymbol{\beta}_2 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}v_{\mathrm{i}}$ (K_2 random parameters)

where

- $\boldsymbol{\beta}_2$ = the fixed means of the distributions for the random parameters
- $\mathbf{z}_i$ = a set of M observed variables which enter the means (optional)
- $\boldsymbol{\Delta}$ = coefficient matrix, $K_2 \times M$, which forms the observation specific term in the mean
- $\mathbf{v}_i$ = unobservable $K_2 \times 1$ latent random term in the ith observation in $\boldsymbol{\beta}_{2i}$. Each element of $\mathbf{v}_i$ has zero mean and variance one. Each element of $\mathbf{v}_i$ may be distributed as normal, uniform, or triangular. They need not be the same.
- $\boldsymbol{\Gamma}$ = lower triangular or diagonal matrix which produces the covariance matrix of the random parameters, $\boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$

$\mathbf{x}_{2i}$ = variables multiplied by $\boldsymbol{\beta}_{2i}$

$\boldsymbol{\beta}_i$ = $[\boldsymbol{\beta}_1', \boldsymbol{\beta}_{2\mathrm{i}}']'$

$\mathbf{x}_i$ = $[\mathbf{x}_{1i}', \mathbf{x}_{2i}']'$

$\lambda|\mathbf{v}_i$ = $\exp(\boldsymbol{\beta}_i'\mathbf{x}_i)$

$P(y_i|\mathbf{x}_i,\mathbf{v}_i)$ = Poisson or negative binomial probability given λ_i.

This formulation allows great flexibility in the specification of the model, and accommodates many special cases.

The command for the random parameters model is structured as follows:

```
    POISSON      ; Lhs  = dependent variable
    or NEGBIN    ; Rhs  = list of all variables in xi, including one if the model contains a
                          constant
                 ; Pts  = r (number of replications – this is optional)
                 ; RPM    (for random parameters model)
or               ; RPM  = list of variables in zi
                 ; Fcn  = specification of random parameters
                 ; Cor    (for correlated parameters – optional) $
```

The **; Fcn** list consists of a list of names of variables which appear in $\mathbf{x}_{2i}$, followed in parentheses by (n) for normally distributed, (u) for uniform, or (t) for triangular.

The following example refits the earlier model with three normally distributed coefficients

```
POISSON  ; Lhs = docvis
         ; Rhs = one,age,hhninc,educ,female,hhkids,married
         ; Rpm ; Fcn = one(n), hhninc(n), female(n) ; Correlated
         ; Pts = 125 ; Halton
         ; Partial Effects $
```

```
-----------------------------------------------------------------------------
Poisson  Regression Start Values for DOCVIS
Dependent variable              DOCVIS
Log likelihood function    -31890.63195
Estimation based on N =   3377, K =   7
Inf.Cr.AIC  =63795.264 AIC/N =   18.891
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01693***      .00090    18.90  .0000      .01518    .01869
    EDUC|    -.02982***      .00435    -6.86  .0000     -.03834   -.02131
  HHKIDS|    -.16138***      .02239    -7.21  .0000     -.20526   -.11750
 MARRIED|     .03198         .02304     1.39  .1651     -.01317    .07713
Constant|     .92173***      .06876    13.41  .0000      .78697   1.05650
  HHNINC|    -.40982***      .04784    -8.57  .0000     -.50359   -.31605
  FEMALE|     .40223***      .01821    22.09  .0000      .36654    .43792
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:  23 iterations. Status=0, F=    8048.005

-----------------------------------------------------------------------------
Random Coefficients  Poisson  Model
Dependent variable              DOCVIS
Log likelihood function     -8048.00464
Restricted log likelihood  -31890.63195
Chi squared [   6 d.f.]     47685.25460
Significance level               .00000
McFadden Pseudo R-squared      .7476373
Estimation based on N =   3377, K =  13
Inf.Cr.AIC  =16122.009 AIC/N =    4.774
Sample is  1 pds and   3377 individuals
POISSON regression model
Simulation based on 125 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|     .01220***      .00083    14.77  .0000      .01058    .01382
    EDUC|    -.04728***      .00420   -11.25  .0000     -.05551   -.03904
  HHKIDS|    -.41927***      .02094   -20.02  .0000     -.46031   -.37822
 MARRIED|     .25499***      .02173    11.73  .0000      .21240    .29758
        |Means for random parameters
Constant|     .51623***      .06881     7.50  .0000      .38137    .65109
  HHNINC|    -.59478***      .06467    -9.20  .0000     -.72154   -.46802
  FEMALE|     .62911***      .02711    23.21  .0000      .57598    .68223
```

```
        |Diagonal elements of Cholesky matrix
Constant|    1.12902***      .01800    62.71  .0000     1.09373   1.16431
  HHNINC|     .06537***      .02414     2.71  .0068      .01806    .11268
  FEMALE|     .53563***      .01231    43.50  .0000      .51150    .55977
        |Below diagonal elements of Cholesky matrix
lHHN_ONE|    -.41713***      .04503    -9.26  .0000     -.50539   -.32887
lFEM_ONE|    -.11997***      .01863    -6.44  .0000     -.15648   -.08346
lFEM_HHN|     .10061***      .01526     6.59  .0000      .07071    .13052
--------+----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-------------------------------------------------------------------------------

Implied covariance matrix of random parameters
Var_Beta|             1              2              3
--------+------------------------------------------
       1|       1.27468      -.470946      -.135446
       2|      -.470946       .178270      .0566190
       3|      -.135446      .0566190       .311417

Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       1.12902
       2|       .422220
       3|       .558048

Implied correlation matrix of random parameters
Cor_Beta|             1              2              3
--------+------------------------------------------
       1|       1.00000      -.987943      -.214978
       2|      -.987943       1.00000       .240299
       3|      -.214978       .240299       1.00000

-------------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   1.7120
Scale Factor for Marginal Effects  1.7120
--------+----------------------------------------------------------------------
        |     Partial                           Prob.      95% Confidence
  DOCVIS|      Effect    Elasticity      z     |z|>Z*         Interval
--------+----------------------------------------------------------------------
     AGE|     .02089***      .52009    14.82  .0000      .01813    .02365
    EDUC|    -.08094***     -.54398   -11.24  .0000     -.09505   -.06683
  HHKIDS|    -.71779***     -.16252   -18.92  .0000     -.79216   -.64342
 MARRIED|     .43655***      .18084    11.43  .0000      .36169    .51142
  HHNINC|   -1.01828***     -.26454    -9.07  .0000    -1.23842   -.79814
  FEMALE|    1.07704***      .29155    17.97  .0000      .95957   1.19452
--------+----------------------------------------------------------------------
```

Other options for the Poisson and negative binomial models are generally supported, including predictions and residuals, restrictions, controls on the optimization, display of output, marginal effects, and so on. Output includes the standard displays, as shown above. The matrices saved are *b* and *varb* as usual, as are scalars *nreg*, *kreg*, *logl*, and *exitcode*. An additional matrix, *sdrpm* is created. This is a column vector which contains the implied standard deviations of the random coefficients.

E43: Two Part Models for Count Data

E43.1 Introduction

This chapter describes several extended models for count data. The basic formulation is the *Poisson regression model*. For a discrete random variable, Y observed over a period of length T_i, and observed frequencies, y_i, $i = 1,...,n$, where y_i is a nonnegative integer count, and regressors $\mathbf{x}_i$,

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \; y_i = 0,1,...; \;\; \log \lambda_i = \boldsymbol{\beta}'\mathbf{x}_i.$$

In this model, λ_i is both the mean and variance of y_i;

$$E[y_i|\mathbf{x}_i] = \lambda_i.$$

The partial effects in this nonlinear regression model are,

$$\frac{\partial E[y_i \,|\, \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i\boldsymbol{\beta}\,.$$

The various extensions are also provided for the negative binomial model.

The two part models involve a behavioral specification of the count model generally involving a participation equation and an intensity equation. The hurdle model, for example, has been used in health care applications in which the individual decides whether or not to use the health care system (the first, binary outcome model equation) and then, given a decision to participation, how intensively to use the system (the second, count equation). The five models detailed in this chapter are

- sample selection
- endogenous treatment effects
- models for underreporting of counts
- zero inflation models (GPP allows this form **; Rh2)**
- hurdle models

These and other two part models are surveyed in Greene (2005a)

E43.2 Model for Sample Selection

The Poisson and negative binomial models can be fit with a Heckman style correction for sample selection. The method used here is maximum likelihood, however, not two step least squares. The formulation is similar to the linear selectivity model. The specification used here is as follows:

y_i = Poisson or negative binomial variable with conditional mean

$\log \lambda_i = \boldsymbol{\beta}' x_i + \varepsilon_i$;

z_i = a binary indicator of whether data on $[y_i, \mathbf{x}_i]$ are observed, with underlying latent structure, $z_i = 1(\boldsymbol{\gamma}' w_i + u_i) > 0$ (a probit model);

(ε_i, u_i) ~ bivariate standard normal with correlation ρ and $\text{Var}[\varepsilon_i] = \sigma^2$;

and $[y_i, \mathbf{x}_i]$ = observed only when $z_i = 1$.

Estimation of the selection model is by full information maximum likelihood.

The following is a counterpart to the sample selection model for linear regression. We present two approaches. The Poisson and negative binomial specifications are modified as follows: The selection indicator z_i is determined by

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1],$$

$$z_i = \mathbf{1}(z_i^* > 0).$$

Thus, a probit model applies to the indicator, z_i. For the observed count variable, in the population,

$$y_i \sim \text{Poisson } (\lambda_i) \text{ or negative binomial } (\lambda_i, \theta)$$

However, $y_i, \mathbf{x}_i$ are observed only when $z_i = 1$.

Then, $y_i | \mathbf{x}_i, (z_i = 1) \sim$ Poisson or negative binomial.

This leads to a Heckman style correction of the count data model. However, the last assumption is questionable. In the standard regression framework, the development proceeds by modeling the joint distribution of u_i and the disturbance in the regression model, which would correspond to

$$\varepsilon_i = y_i - E[y_i | \mathbf{x}_i].$$

The familiar Heckman model hinges on joint normality of $[u_i, \varepsilon_i]$, which is clearly untenable here – since y_i is discrete, its deviation from the conditional mean function would not be normally distributed. The approach taken is to join the selection approach with the heterogeneity model of the Section E42.4.1;

$$\lambda_i | \varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i)$$

Then, $[u_i, \varepsilon_i]$ ~ Bivariate normal with zero means, correlation ρ, and standard deviations 1 and σ.

Thus, $y|\varepsilon$ is distributed as Poisson with mean (and variance) $E[y|\varepsilon] = \exp(\boldsymbol{\beta}'\mathbf{x} + \varepsilon)$. The distribution in the selected population is nonPoisson, but this does preserve its discreteness. The force of the sample selection is exerted on the mean of the discrete variable (and its variance). The estimator is full information maximum likelihood. Discussion appears in Terza (1998, 2009) and Greene (2011).

E43.2.1 Full Information Maximum Likelihood Estimation

A full information maximum likelihood estimator for the sample selection model is requested for the **POISSON** or **NEGBIN** specifications with

PROBIT ; Lhs = ... ; Rhs = ... ; Hold $
POISSON ; Lhs = .. ; Rhs = ... ; Selection ; MLE $ (or NEGBIN)

The computations are based on the heterogeneity model. This must be preceded by the probit model in order to define the full set of variables in the model and to provide the starting values for the iterations. Partial effects are requested with

; Partial Effects.

All options, including

Optimization:	**; Maxit = n**	to set maximum restrictions
	; Alg = name	to select algorithm (you generally should not change this)
	; Tlf [= value]	to set tolerance for convergence criteria
	; Output = value	to control intermediate output
	; Hpt = n	to specify number of nodes for Hermite quadrature
Constraints:	**; Rst = list**	to specify fixed value and equality restrictions
	; CML: spec	to define a constrained maximum likelihood estimator
	; Test: spec	to define Wald tests
Output:	**; Covariance Matrix**	to display the estimated asymptotic covariance matrix, same as **; Printvc**.
	; List	to display predicted values
	; Keep = name	to retain fitted values
	; Res = name	to retain residuals
	; Parameters	to retain estimates of σ and ρ in *b* and *varb*

and so on for other program options are all supported. Output for this model will include the initial Poisson regression followed by the FIML results, then any optional output you have requested, such as a list of fitted values.

NOTE: This estimator reestimates the parameters of the probit model, and replaces the estimates that were initially retained with **; Hold** on the probit command. See the example below.

The results that are retained include

Matrices: *b* and *varb* include Poisson slopes followed by probit parameters σ then ρ with **; Parameters** option

Scalars:

logl = log likelihood
kreg = number of parameters in [β′, γ′, σ, ρ]′
nreg = number of observations, total, not just selected
s = estimate of σ
rho = estimate of ρ

Last Function: None

The elements of the partial effects are computed by quadrature with the model, during the estimation step. **PARTIALS** and **SIMULATE** are not enabled for this model.

This example bases the sample selection on the addon insurance variable in the health care data.

PROBIT ; Lhs = addon ; Rhs = one,age,hhninc,married,hhkids ; Hold $
POISSON ; Lhs = docvis ; Rhs = one,age,female,hsat
; Select ; MLE ; Partial Effects $

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                ADDON
Log likelihood function     -2545.01803
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
   ADDON|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|   -2.36091***      .08870   -26.62  .0000    -2.53475  -2.18706
     AGE|     .00447**       .00183     2.44  .0147      .00088    .00805
  FEMALE|     .05305         .03593     1.48  .1398     -.01738    .12348
 MARRIED|     .05665         .04745     1.19  .2325     -.03635    .14965
  HHKIDS|     .03654         .04251      .86  .3901     -.04679    .11986
--------+--------------------------------------------------------------------
Unrestricted Poisson Regression Start Value
Dependent variable               DOCVIS
Log likelihood function     -3424.32688
Sample size=   27326; selected      514
Estd sigma for heterogeneity =     .413
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2.09288***      .14137    14.80  .0000     1.81580   2.36996
     AGE|     .00748***      .00243     3.08  .0021      .00272    .01224
  FEMALE|     .29679***      .05149     5.76  .0000      .19587    .39770
    HSAT|    -.23627***      .01018   -23.20  .0000     -.25623   -.21631
--------+--------------------------------------------------------------------
Normal exit:  24 iterations. Status=0, F=    3646.965
-----------------------------------------------------------------------------
```

```
Poisson  Model with Sample Selection.
Dependent variable               DOCVIS
Log likelihood function     -3646.96549
Restricted log likelihood   -5969.34491
Chi squared [   2 d.f.]      4644.75884
Significance level               .00000
McFadden Pseudo R-squared      .3890510
Estimation based on N =  27326, K =  11
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit =    -2545.0180
LogL for initial Poisson=    -3424.3269
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson/Neg. Binomial Probability
Constant|    1.71261        3.79360       .45  .6517    -5.72271   9.14792
     AGE|     .01086         .00762      1.42  .1545    -.00409     .02580
  FEMALE|     .41822***      .13145      3.18  .0015     .16059     .67585
    HSAT|    -.24388***      .02512     -9.71  .0000    -.29312    -.19464
        |Parameters of Probit Selection Model
Constant|   -2.35926***      .09212    -25.61  .0000   -2.53980  -2.17872
     AGE|     .00446**       .00193      2.31  .0210     .00067     .00824
  FEMALE|     .05301         .03640      1.46  .1453    -.01833     .12434
 MARRIED|     .05560         .04984      1.12  .2647    -.04209     .15329
  HHKIDS|     .03565         .04357       .82  .4132    -.04974     .12104
        |Standard Deviation of Heterogeneity
   Sigma|     .93205***      .12290      7.58  .0000     .69117    1.17294
        |Correlation of Heterogeneity & Selection
     Rho|    -.09092        1.53194      -.06  .9527   -3.09347   2.91162
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the var-
iables.  Separate effects are shown first
followed by the sum of the two effects for
variables which appear in both Poisson and
Probit models.  Estimated value of E[y|D=1]
using sample mean =   .06401.
Note, std. errs. assume fixed rho & sigma.
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson/Neg. Binomial Probability
     AGE|     .00070         .00049      1.42  .1545    -.00026     .00165
  FEMALE|     .02677***      .00841      3.18  .0015     .01028     .04326
    HSAT|    -.01561***      .00161     -9.71  .0000    -.01876    -.01246
        |Parameters of Probit Selection Model
     AGE|     .03634**       .01647      2.21  .0273     .00406     .06861
  FEMALE|     .43230         .30056      1.44  .1503    -.15679    1.02138
 MARRIED|     .45340         .41237      1.10  .2715    -.35482    1.26163
  HHKIDS|     .29073         .35830       .81  .4171    -.41152     .99298
        |Combined effect of two terms
     AGE|     .03703**       .01632      2.27  .0232     .00505     .06902
  FEMALE|     .45907         .29990      1.53  .1258    -.12872    1.04686
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E43.2.2 Imposing Restrictions and Fixing ρ

The parameter vector is [β′, γ′, σ, ρ]′. Use this if you wish to impose constraints. For example, to fix the value of ρ at -.5, you could use the following:

```
NAMELIST  ; xp = Rhs variables in probit equation
          ; xr = Rhs variables in Poisson model $
CALC      ; kp = Col(xp) ; kr = Col(xr) $
PROBIT    ; Lhs = ...
          ; Rhs = xp
          ; Hold $
POISSON   ; Lhs = the dependent variable
          ; Rhs = xr
          ; Selection
          ; MLE
          ; Rst = kr_b, kp_c, sg, -.5 $
```

```
-----------------------------------------------------------------------------
Poisson  Model with Sample Selection.
Dependent variable                DOCVIS
Log likelihood function      -3647.22494
Restricted log likelihood    -5969.34491
Chi squared [   2 d.f.]       4644.23995
Significance level                .00000
McFadden Pseudo R-squared       .3890075
Estimation based on N =  27326, K =  10
Inf.Cr.AIC  =    7314.4 AIC/N =     .268
Model estimated: Jul 30, 2011, 11:06:12
Mean of LHS Variable =           3.12451
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit =     -2545.0180
LogL for initial Poisson=     -3424.3269
Means for Psn/Neg.Bin. use selected data.
Means for Probit based on all observations.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters of Poisson/Neg. Binomial Probability
Constant|    2.87547***      .30776     9.34  .0000     2.27228   3.47867
     AGE|     .00903*        .00523     1.73  .0841     -.00122    .01928
  FEMALE|     .39957***      .10993     3.63  .0003      .18411    .61504
    HSAT|    -.24353***      .02488    -9.79  .0000     -.29230   -.19477
        |Parameters of Probit Selection Model
Constant|   -2.34934***      .09188   -25.57  .0000    -2.52941  -2.16927
     AGE|     .00441**       .00191     2.31  .0209      .00067    .00815
  FEMALE|     .05250         .03633     1.45  .1484     -.01870    .12370
 MARRIED|     .04890         .04832     1.01  .3115     -.04580    .14361
  HHKIDS|     .03031         .04290      .71  .4799     -.05377    .11439
        |Standard Deviation of Heterogeneity
   Sigma|    1.05096***      .06327    16.61  .0000      .92696   1.17495
        |Correlation of Heterogeneity & Selection
     Rho|    -.50000     .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

You can use this device to test for a selectivity effect as well. The simple t and likelihood ratio tests can be carried out based on the value of ρ that is estimated. But, the t test requires estimation of the full model while the LR test requires assembling estimates of the pair of models and collecting three terms:

```
PROBIT    ; ... ; Hold $
POISSON   ; ... estimate full model by FIML $
CALC      ; lfiml = logl $
CALC      ; lprobit = logl $
REJECT    ; the Lhs variable for probit model = 0 $
POISSON   ; ... Poisson model without selection, on selected observations $
CALC      ; lpois = logl
          ; List
          ; lm = 2*(lfiml - lprobit - lpois)
          ; 1 - Chi(lm,1)  $
```

The LM test should be the simplest to carry out. In the earlier example, just change our -.5 to 0, and add **; Maxit = 0** to the command. An example appears below.

E43.2.3 Technical Details

The central result in estimation of the two part models by FIML as done here is the connection of the participation equation to the intensity equation through the correlation of the two disturbances. The participation equation is

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1],$$

$$z_i = \mathbf{1}(z_i^* > 0).$$

The intensity equation is based on a conditional mean function

$$\lambda_i|\varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i), \text{ where } \varepsilon_i \sim N[0,\sigma^2] \text{ and } \mathrm{corr}(\varepsilon_i,u_i) = \rho.$$

To obtain the log likelihood, we first project u_i on ε_i so

$$u_i = (\rho/\sigma)\varepsilon_i + \tau v_i \text{ where } \tau = \sqrt{1-\rho^2}\ .$$

Combining terms, the density for the observed $y_i|\varepsilon_i$ when $z_i = 1$ is the Poisson probability

$$P(y_i|z_i{=}1, \varepsilon_i) = P_i(y_i|\varepsilon_i).$$

The contribution to the likelihood is the density of the observed outcome. Conditioned on ε_i, when $z_i = 0$, this is

$$\mathrm{Prob}(z_i = 0|\varepsilon_i) = \Phi\left(\frac{-[\gamma' z_i + (\rho/\sigma)\varepsilon_i]}{\sqrt{1-\rho^2}}\right) = \Phi_i^0(\varepsilon_i)$$

The contribution to the likelihood when $z_i = 1$ is the joint density of z_i and y_i.

$$\text{Prob}(y_i,z_i{=}1|\varepsilon_i) = \text{Prob}(y_i|z_i{=}1, \varepsilon_i)\text{Prob}(z_i{=}1|\varepsilon_i)$$

$$= P_i(y_i|\varepsilon_i)\,\Phi\left(\frac{[\gamma' z_i + (\rho/\sigma)\varepsilon_i]}{\sqrt{1-\rho^2}}\right) = \Phi_i^1(\varepsilon_i)P(y_i \mid \varepsilon_i)\,.$$

To form the log likelihood function, it is necessary to integrate e_i out of the density. The unconditional contribution to the likelihood of observation i is

$$f(y_i, z_i) = \int_{\varepsilon_i} f(y_i, z_i \mid \varepsilon_i) f(\varepsilon_i) d\varepsilon_i.$$

Collecting terms once again, this is

$$f(y_i, z_i) = \int_{\varepsilon_i} \left[(1-z_i)\Phi_i^0(\varepsilon_i) + z_i\Phi_i^1(\varepsilon_i)P(y_i \mid \varepsilon_i)\right]\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)d\varepsilon_i.$$

The log likelihood function is then the sum of the logs of the terms. Parameters to be estimated are $\boldsymbol{\gamma}$ in the probit equation, $\boldsymbol{\beta}$ in the Poisson conditional mean function, ρ and σ. The integrals are computed using Gauss-Hermite quadrature. Further details on the method may be found in Section R26.7 and in Greene (2011). Partial effects are obtained as the derivatives of the expected conditional mean function,

$$E(y_i \mid \mathbf{x}_i) = \int_{\varepsilon_i} \left[\Phi_i^1(\varepsilon_i)\lambda(\varepsilon_i)\right]\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)d\varepsilon_i.$$

This function and its derivatives are also computed using Hermite quadrature.

E43.3 An Incidental Truncation Model

Winkelmann (1997, pp. 112-113) describes a model (attributed to Crepon and Duguet (1995)) which is labeled the 'incidental truncation' model. This is a case in which the binary variable is correlated with the Poisson outcome, and directly affects it, in a form similar to the ZIP models discussed below. In this model, the data are observed when $z_i = 0$, but $z_i = 0$ implies that $y_i = 0$. The difference between this and the ZIP model is the correlation between the two latent disturbances. The structure is actually a small modification of the model we have considered above.

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1],$$

$$z_i = \mathbf{1}(z_i^* > 0).$$

Thus, a probit model applies to the indicator, z_i. The following applies to the observed y_i:

y_i^* ~ Poisson $(\lambda_i|\varepsilon_i)$ is a latent variable distributed as Poisson

$\lambda_i|\varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i)$

$y_i = y_i^*$ and $\mathbf{x}_i$ are observed when $z_i = 1$.

$y_i = 0$ when $z_i = 0$, $\mathbf{x}_i$ is still observed when $z_i = 0$.

For the sample selection model, the joint density of the observed response variables y_i and z_i is of the form

$$\mathbf{1}(z_i = 1)\times\{\text{Prob}(z_i = 1) \times \text{Poisson probability}\} + \mathbf{1}(z_i = 0)\times\text{Prob}(z_i = 0)$$

while for the incidental truncation model, the joint density is of the form

$$\text{Prob}(z_i = 1) \times \text{Poisson probability} + \mathbf{1}(z_i = 0)\times\text{Prob}(z_i = 0)$$

This model is requested by adding

; All

to the **POISSON** command given earlier. All other aspects are the same.

E43.4 Endogenous Treatment Effect

The endogenous treatment is a modification of the selection model in which the 'selection' equation is replaced with a treatment equation,

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1],$$
$$z_i = \mathbf{1}(z_i^* > 0).$$

A probit model applies to the treatment indicator, z_i. The following applies to the observed y_i:

$$y_i^* \sim P(\lambda_i|\varepsilon_i) \text{ is a latent variable distributed as Poisson or negative binomial}$$
$$\lambda_i|\varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta z_i + \varepsilon_i), \text{ where } \varepsilon_i \sim N[0,\sigma^2] \text{ and } \text{corr}(\varepsilon_i,u_i) = \rho.$$

The treatment dummy variable appears in the conditional mean function of the count variable. The endogeneity of the treatment effect is induced by the correlation of ε_i and u_i.

This model is requested with

```
PROBIT       ; Lhs = ... ; Rhs = ... ; Hold $
POISSON      ; Lhs = ..  ; Rhs = ... , z
or NEGBIN    ; Selection ; MLE
             ; Treatment $
```

The command differs from the selection model by the appearance of z in the Rhs of the count model and in the addition of **; Treatment** in the command.

The example below continues the application of the selection model in Section E43.2.1.

```
PROBIT       ; Lhs = addon
             ; Rhs = one,age,female,married,hhkids ; Hold $
POISSON      ; Lhs = docvis
             ; Rhs = one,age,female,hsat,addon
             ; Selection ; MLE ; Treatment $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                ADDON
Log likelihood function     -2482.98865
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   ADDON|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|   -2.67747***      .09373   -28.56  .0000    -2.86118  -2.49375
     AGE|     .00521***      .00188     2.77  .0057      .00152    .00890
  HHNINC|     .93381***      .07746    12.06  .0000      .78199   1.08563
 MARRIED|    -.02112         .04875     -.43  .6649     -.11667    .07444
  HHKIDS|     .05577         .04276     1.30  .1921     -.02803    .13958
--------+--------------------------------------------------------------------
Unrestricted Poisson Regression Start Value
Dependent variable               DOCVIS
Log likelihood function   -212776.46611
Estd. sigma for heterogeneity =    .424
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.95339***      .01809   107.99  .0000     1.91794   1.98885
     AGE|     .01005***      .00031    32.30  .0000      .00944    .01066
  FEMALE|     .27823***      .00687    40.49  .0000      .26476    .29170
    HSAT|    -.22992***      .00132  -174.43  .0000     -.23251   -.22734
   ADDON|    -.00412         .02519     -.16  .8702     -.05349    .04526
--------+--------------------------------------------------------------------
Normal exit:  35 iterations. Status=0, F=    60497.98
-----------------------------------------------------------------------------
Poisson  Model with Endogenous Treatment
Dependent variable               DOCVIS
Log likelihood function    -60497.98492
Restricted log likelihood -215259.45475
Chi squared [   2 d.f.]    309522.93967
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit  =   -2482.9887
LogL for initial Poisson = -212776.4661
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Parameters of Poisson/Neg. Binomial Probability
Constant|    1.33582***      .04532    29.47  .0000     1.24698   1.42465
     AGE|     .01235***      .00078    15.84  .0000      .01082    .01388
  FEMALE|     .40651***      .01639    24.81  .0000      .37439    .43862
    HSAT|    -.25744***      .00323   -79.61  .0000     -.26378   -.25111
   ADDON|   -2.18277***      .07368   -29.63  .0000    -2.32717  -2.03836
        | Parameters of Probit Selection Model
Constant|   -2.75320***      .08502   -32.38  .0000    -2.91983  -2.58656
     AGE|     .00768***      .00183     4.20  .0000      .00410    .01126
  HHNINC|     .83750***      .06500    12.88  .0000      .71010    .96491
 MARRIED|    -.01517         .04396     -.35  .7300     -.10134    .07100
  HHKIDS|     .10323***      .03912     2.64  .0083      .02656    .17991
        | Standard Deviation of Heterogeneity
   Sigma|    1.24965***      .00820   152.44  .0000     1.23358   1.26571
        | Correlation of Heterogeneity & Selection
     Rho|     .81774***      .02025    40.38  .0000      .77804    .85743
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E43.5 Poisson Models with Underreporting

We consider two models for underreporting in count data. The basic formulation is as follows: The observed count y_i is assumed to be the sum of J indicators of whether an event that occurred was reported or not. Thus, suppose that an event occurs at instant j, and c_j is an indicator, 0 or 1, that the event is counted in the total. Thus, $y_i = \Sigma_j c_j$. The probability distribution associated with y_i is induced by the underlying probability that c_i is 1. We consider several models of underreporting, based on probit and logit models, and based on exogenous or endogenous reporting.

The Poisson model with underreporting is developed in Winkelmann (1996) and Winkelmann and Zimmermann (1993). The underlying logic is that the Poisson count, Y, is the result of recording of Y individual events, y_j. In the standard model, if c_i is an indicator that the ith event that happens is actually recorded, then, the probability that c_i equals one is 1.0. But, suppose that c_i is a binary variable determined by a binary process, such that

$$c_i^* = \boldsymbol{\gamma}'\mathbf{z}_i + u_i$$

$$c_i = 1 \text{ iff } c_i^* > 0,$$

where $\boldsymbol{\gamma}$ is a parameter vector, $\mathbf{z}_i$ is a covariate vector, and u_i is a disturbance. If u_i is normally distributed, this is a probit model. The authors show that with this form of underreporting in the Poisson model,

$$\text{E}[y_i \,|c_i\,] = P_i^* \text{ E}[y_i\,]$$

where $P_i = \text{Prob}[c_i = 1]$

and $E[y_i]$ = the mean of the underlying Poisson variable.

In the probit case, $P_i = \Phi(\boldsymbol{\gamma}'\mathbf{z}_i)$ where $\Phi(.)$ is the standard normal CDF.

We also allow a logistic model, with $P_i = \Lambda(\boldsymbol{\gamma}'\mathbf{z}_i)$.

The basic underreporting model, is a Poisson regression model with

$$\text{E}[y_i \mid \mathbf{x}_i,\mathbf{z}_i\,] = \exp[\,\boldsymbol{\beta}'\mathbf{x}_i\,]F(\boldsymbol{\gamma}'\mathbf{z}_i)$$

where $F(.)$ may be specified as either a probit or logit equation. The model commands for this model are

POISSON ; Lhs = y ; Rhs = x ; Rh2 = z $

and **POISSON ; Lhs = y ; Rhs = x ; Rh2 = z ; Logit $**

The other options for this model are the same as for the standard Poisson model, including **; Partial Effects**, the output controls for fitted values and residuals, display of technical output, controls of the optimization method, and restrictions, which can be imposed with

; Rst = specification

or **; CML: specification**

In both cases, the parameter vector being estimated is $[\boldsymbol{\beta}, \boldsymbol{\gamma}]$.

E43.5.1 Heterogeneity and Exogenous Underreporting

The model of the previous section can be converted to one with exogenous underreporting by adding unobserved heterogeneity to the Poisson model. The full structure becomes

$$c_i^* = \boldsymbol{\gamma}'\mathbf{z}_i + u_i$$

$$c_i = 1 \text{ iff } c_i^* > 0,$$

$$E[y_i \mid c_i, \varepsilon_i] = P_i^* \, E[y_i \mid \varepsilon_i]$$

where $P_i = \text{Prob}[c_i = 1]$

$E[y_i \mid \varepsilon_i]$ = the mean of the underlying Poisson variable

$$E[y_i \mid \mathbf{x}_i, \mathbf{z}_i, \varepsilon_i] = \exp[\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i] \, F(\boldsymbol{\gamma}'\mathbf{z}_i)$$

At this point we assume that the correlation between ε_i and u_i equals zero. This model is requested by the command

POISSON ; Lhs = y ; Rhs = x ; Het ; Rh2 = z ; No Correlation $

(As before, **; Logit** may be specified.) The last specification is provided to restrict the model to the exogeneity case – the more general model is presented just below.

An alternative model with exogenous underreporting is Winkelmann's Poisson/Logit model, which replaces the probit reporting equation with a logit reporting equation. The resulting model is

$$E^{**}[Y_i \mid \varepsilon_i] = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) \times 1 / \{1 + \exp[-\boldsymbol{\gamma}'\mathbf{z}_i]\}$$

This model is requested with

POISSON ; Lhs = y ; Rhs = x ; Heterogeneity ; Rh2 = z ; Logit $

This model is a modification of the heterogeneity model presented in Section E42.4. Estimation is done by the same quadrature method. This model merely changes the form of the conditional mean function.

E43.5.2 Endogenous Underreporting

The most general form of this class of models is the model with endogenous underreporting. This model is obtained by relaxing the restriction that the correlation between the heterogeneity and the latent effect in the binary choice model is zero. Winkelmann shows that the resulting distribution is a modification of our heterogeneity model in the previous section. We begin with the same specification:

$$c_i^* = \boldsymbol{\gamma}'\mathbf{z}_i + u_i.$$

Write $u_i = E[u_i \mid \varepsilon_i] + h_i = (\rho/\sigma)\varepsilon_i + h_i.$

Then, $\text{Var}[h_i] = (1 - \rho^2).$

The recording event is $c_i \quad = \; 1$ iff $c_i^* > 0$.

We require $\quad P_i|\varepsilon_i \; = \; \text{Prob}[c_i = 1 \mid \varepsilon_i]$

$$= \; \text{Prob}[\boldsymbol{\gamma}'\mathbf{z}_i/\sqrt{1-\rho^2} \;+\; \{(\rho\sigma)/\sqrt{1-\rho^2}\,\}\varepsilon_i \;+\; v_i \;>\; 0]$$

where $v_i = h_i \,/\, \sqrt{1-\rho^2}$ has standard normal distribution. Let $\theta = (\rho\sigma)/\sqrt{1-\rho^2}$ and $\boldsymbol{\delta} \;=\; \boldsymbol{\gamma}/\sqrt{1-\rho^2}$. The conditional probability is now the usual for a probit model,

$$P_i \mid \varepsilon_i \;=\; \Phi[\boldsymbol{\delta}'\mathbf{z}_i + \theta\varepsilon_i].$$

Returning to the Poisson model,

$$\text{E}[y_i \mid c_i, \varepsilon_i] \;=\; P_i|\varepsilon_\text{i} \times \text{E}[y_i \mid \varepsilon_i]$$

where we have denoted

$$\text{E}[y_i \mid \varepsilon_i] \;=\; \text{the mean of the underlying Poisson variable.}$$

Combining terms, we have

$$\text{Prob}[Y_i = j \mid \text{x}_i, \mathbf{z}_i, \varepsilon_i] \;=\; \exp(-E^*[Y_i \mid \varepsilon_i\,]) \times \{\text{E}^*[\,Y_i \mid \varepsilon_i]\}^j \,/\, j!,$$

where $\quad \text{E}^{**}[\,Y_i \mid \varepsilon_i] \;=\; \exp[\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i] \times \Phi[\boldsymbol{\delta}'\mathbf{z}_i + \theta\varepsilon_i],$

$$\boldsymbol{\delta} \;=\; \boldsymbol{\gamma} \,/\, (1 - \rho^2)^{\,1/2} \;\text{ and }\; \theta \;=\; \rho\sigma \,/\, \sqrt{1-\rho^2}\;.$$

and ρ is the correlation between ε_i and u_i. That is, the endogenous underreporting changes the mean of the Poisson distribution. As before, estimation is carried out by integrating u_i out of the conditional distribution. A nonzero value of ρ produces the endogeneity of the reporting.

This model is requested simply by adding

; Rh2 = variables in z(i) ; Het

to the Poisson model, so the full command is

POISSON ; Lhs = y ; Rhs = ... x ; Rh2 = ... z ; Het $

The additional parameters estimated are scalar coefficient ρ and coefficient vector $\boldsymbol{\gamma}$. For purposes of starting values and fixed value/equality restrictions, the coefficient vector in this model is $[\boldsymbol{\beta} \quad \boldsymbol{\gamma} \quad \rho \quad \sigma]$. Thus, you should use

; Start = list of values for $\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \sigma$,
and/or **; Rst** = list of specifications for $\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \sigma$

Other results for this model are the same as for the heterogeneity model, including parameter estimates, predicted values, marginal effects, etc. An application below illustrates.

E43.6 Zero Inflation Models for Counts

In some settings, the zero outcome of the count data model represents a sort of partial observability. Consider, for example, one's answer to a survey question about utilization of a sport fishing site within a recent period. The answer, 'zero' could arise from two underlying responses. If the individual is not a participant in this sport, they would always answer zero. If they are, however, then the zero may be just the number of times they used the site *in the particular period,* and the response might be some positive number in another period.

- $Y = 0$ happens to be the number of times the individual used that facility in the survey period. At some other time, the same individual might choose $Y = j > 0$.

- $Y = 0$ occurs because the individual would never use the facility, regardless of the characteristics that appear in the model.

If so, then fitting a simple Poisson model (or negative binomial) to these data would overstate ('inflate') the theoretical probability of zero in the Poisson model. The Poisson model may not accurately assign probability to the outcome $Y = 0$, if a separate process is simultaneously at work influencing this outcome. An alternative formulation for these data that might be more appropriate is the 'Zero Inflated Poisson' (ZIP) model:

$$z = 0 \text{ if the response would always be 0, 1 if a Poisson model applies,}$$

$$y = \text{the response from the Poisson model,}$$

$$zy = \text{the observed response.}$$

Then, the probabilities of the various outcomes are

$$\text{Prob}[y = 0] = \text{Prob}[z = 0] + \text{Prob}[z = 1]\times\text{Prob}[y = 0 \mid \text{Poisson}]$$

$$\text{Prob}[y = \text{j} > 0] = \text{Prob}[z = 1] \times \text{Prob}[y = j \mid \text{Poisson}].$$

Another clearly defined example is provided by Lambert (1992) in which the observed outcome is the number of defective items produced by a production process. If the process is 'in control,' the number will be zero, by definition. If the process is 'not in control,' the sampled count might be zero or some positive value, depending partly on sampling variability. As one more alternative, consider the number of children reported by a survey respondent. The response 0 *yet* is different from the response 0 *and none planned*.

The ZIP model is based on construction of a model for z, such as the probit model, which is then integrated into the count data settings (Poisson and negative binomial) discussed above. We allow several different formulations for the model. (See Lambert (1992) and Greene (1994).) The ZIP model is also extended to some of the other variants of the model, including the underreporting model and the semiparametric random effects model.

The ZIP model is, using our own notation (not Lambert's),

$$Y_i = 0 \text{ with probability } q_i$$

$$Y_i \sim \text{Poisson}(\lambda_i) \text{ with probability } 1 - q_i$$

so that

$$\text{Prob}[Y_i = 0] = q_i + [1 - q_i]R_i(0)$$

$$\text{Prob}[Y_i = j > 0] = [1 - q_i]R_i(j)$$

where

$$R_i(y) = \text{the Poisson probability} = e^{-\lambda i}\, \lambda_i^{\,yi} \;/\; y_i!$$

and

$$\lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}i}.$$

We allow four formulations of the ancillary, state probability, q_i,

$$q_i \sim \text{Logistic}[v_i]$$

and

$$q_i \sim \text{Normal}[v_i].$$

Let $F[v_i]$ denote either the normal or logistic CDF. Then, v_i may be defined in two ways. First,

$$v_i = \tau\log[\lambda_i] = \tau\boldsymbol{\beta}'\mathbf{x}_i$$

which defines a single new parameter (which may be positive or negative). This is labeled the ZIP(τ) model in the following. The alternative model is

$$v_i = \boldsymbol{\gamma}'\mathbf{z}_i$$

for a parameter vector $\boldsymbol{\gamma}$ and set of variables $\mathbf{z}_i$ which may or may not share variables with $\mathbf{x}_i$. Note that if $\mathbf{z}_i$ equals $\mathbf{x}_i$, the two models are still not the same since even if so, the second allows a full set of new parameters. An excellent reference on this model, albeit with a somewhat narrower focus on the choice aspects of the model than we consider, is Lambert (1992). Another related source is Mullahy (1986). Finally, Greene (1994) presents the theory upon which the ZIP estimator and selectivity estimator described earlier are based.

The same formulations are provided for the negative binomial model, which has

$$R_i(j) = \Gamma(\theta+y_i)/[y_i!\Gamma(\theta)]\; u_i^{\theta}\; [1 - u_i]^{\,yi}$$

$$\theta = 1/\alpha, \text{ where } \alpha \text{ is the overdispersion parameter}$$

$$u_i = \theta / [\theta + \lambda_i].$$

and the gamma count model,

$$R_i(j) = G(\alpha j, \lambda_i) - G(\alpha j + \alpha, \lambda_i)$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i) \text{ (as usual)}$$

$$G(\alpha j, \lambda_i) = 1 \text{ if } j = 0, \text{ or } \frac{1}{\Gamma(\alpha j)}\int_0^{\lambda_i} u^{\alpha j-1}e^{-u}du \text{ if } j > 0,\ j = 0,1,...$$

The ZIP model extends the Poisson model in a few different directions. First, of course, the altering of the zero probability will be useful in some settings. In addition, the changed zero probability induces a divergence between the mean and variance of the distribution. For the Poisson model, define the indicator $d_i = 1$ if state 1, y_i always = 0 and $d_i = 0$ if state 2, y_i produced by the Poisson process. Thus, $q_i = \text{Prob}[d_i = 1]$. Then,

$$E[y_i] = q_i 0 + (1 - q_i)\lambda_i = (1 - q_i)\lambda_i$$

and

$$\begin{aligned}\text{Var}[y_i] &= E_{di}[\text{Var}[y_i \mid d_i]] + \text{Var}_{di}[E[y_i \mid d_i]] \\ &= [q_i 0 + (1 - q_i)\lambda_i] + [q_i(0 - (1 - q_i)\lambda_i)^2 + (1 - q_i)(\lambda_i - (1 - q_i)\lambda_i)^2] \\ &= \lambda_i(1 - q_i)[1 + \lambda_i q_i] > E[y_i].\end{aligned}$$

As such, the ZIP specification induces overdispersion, though it arises from a different source than is assumed in the familiar treatments in the literature.

It will be useful to test for the overdispersion. However, there will be a problem distinguishing the ZIP model from an underlying negative binomial specification as the source of the overdispersion. In particular, recall for the negative binomial model that

$$\text{Var}[y_i]/E[y_i] = 1 + \alpha E[y_i].$$

In the ZIP model, we have that

$$\text{Var}[y_i]/E[y_i] = 1 + [q_i/(1-q_i)]E[y_i],$$

which is quite similar. The testing procedure is complicated by the fact that the ZIP model is not nested within either the Poisson or the negative binomial models. That is, the restriction which produces the simpler model is $q_i = 0$, which is not a simple parametric restriction. In order to make $q_i \to 0$, it is necessary for some parameter to $\to +\infty$ or $-\infty$. Vuong (1989) has proposed a test statistic for nonnested models which appears to have some power to distinguish between non-Poissonness due to the overdispersion of the negative binomial model and the force of the splitting mechanism in the ZIP part of the model. The statistic is

$$V = \sqrt{n}\,\overline{m}/s_m$$

where

$$m_i = \log[f_1(y_i)/f_2(y_i)],$$

$f_1(\bullet)$ and $f_2(\bullet)$ are densities for the competing models, and $\overline{m}$ and s_m are the sample mean and standard deviation for the sample of m_is. Asymptotically, the statistic is distributed as standard normal, so its value may be compared to the critical value from the standard normal distribution, e.g., 1.96. The test is directional; large positive values favor f_1 while large negative values favor f_2. The Vuong test is included in the standard output for the ZIP models. (See the application below.)

E43.6.1 Commands for the ZIP Models

The basic model command for the Poisson/ZIP(τ) is

POISSON ; Lhs = ... ; Rhs = ...
or NEGBIN ; ZIP $

This uses the logistic splitting distribution. For the model, with normally distributed splitting rule, add

; ZIP = normal

to the command. Once again, this requests the τ form of the model, $z_i = \tau\boldsymbol{\beta}'\mathbf{x}_i$. To request the second form, with an independent model for the regime split, simply add to the command

; Rh2 = the variables in z

Note that the presence of **; ZIP** in the command is the essential switch. If this is omitted, the default Poisson model is estimated. The **NEGBIN ; ZIP ; …** combination is often called the ZINB or ZINB(τ) in the recent literature. Other model frameworks available for this model are

; Model = Gamma

for the gamma model and

; Model = GP1 or GP2 or GPP

for the generalized Poisson model. The generalized Poisson model does not provide the τ format. If you do not include **; Rh2 = list** for the generalized Poisson model, the program substitutes **; Rh2 = one**. The preceding provides 14 different specifications for the zero inflation model (four models, τ or not, logistic or normal, lead to the τ forms for the generalized Poisson).

NOTE: Censoring, truncation and sample selection are not supported in this model.

If you wish to give starting values and/or fixed value and equality constraints, specify them for the ZIP model with parameter vector [$\boldsymbol{\beta}$,τ] for the ZIP(τ) model or [$\boldsymbol{\beta},\boldsymbol{\gamma}$] for the ZIP model. For the ZINB(τ), ZINB, and the gamma models, the extra parameter, θ for the negative binomial model and P for the gamma model follows $\boldsymbol{\beta}$ in the list. The generalized Poisson has yet another parameter, the P = 1 or 2 or free. If you choose the ZIP form, $\boldsymbol{\gamma}$ will be the last set of parameters in the list (and in the displayed output). Note, once again, you provide θ, not α for the negative binomial model. The options

; Start = list of starting values
and **; Rst = constraints**
or **; CML: specification of constraints**

may be used freely with these models. Do note that because of a difference in the order of magnitude, equality constraints forcing elements of $\boldsymbol{\gamma}$ to equal elements of $\boldsymbol{\beta}$ will probably produce very poor results.

All other options for the Poisson and negative binomial models are available, including

; Keep = name

to save the conditional mean, $E[y_i] = (1 - q_i)\lambda_i$,

; Res = name

to save the residual, and

; Prob = name

to save the probability associated with the observed outcome,

$$\text{Prob}[Y_i = 0] = q_i + [1 - q_i]R_i(0)$$
$$\text{Prob}[Y_i = j > 0] = [1 - q_i]R_i(j).$$

All other options for nonlinear optimization listed earlier are also supported. You may also request

; Partial Effects

as usual. The computed marginal effects are the derivatives of the conditional mean function shown above, which are computed using

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial \mathbf{x}_i} = (1 - q_i)\lambda_i \boldsymbol{\beta} - \lambda_i \frac{\partial q_i}{\partial \mathbf{x}_i}$$

The second part of the marginal effect will vary depending on the model. There are four forms for the normal or logistic models and the linear form, $\boldsymbol{\gamma}'\mathbf{z}_i$ and the τ form $\tau\boldsymbol{\beta}'\mathbf{x}_i$. For convenience, assume that $\mathbf{z}_i = \mathbf{x}_i$. (Otherwise, this applies to any variables that $\mathbf{z}_i$ and $\mathbf{x}_i$ have in common.). The splitting probability is either the logistic CDF, $\Lambda(.)$, which has density $q_i'(.) = \Lambda(.)[1-\Lambda(.)]$, or the normal CDF, $\Phi(.)$ which has density $q_i'(.) = \phi(.)$. Assembling the parts gives

$$\frac{\partial q_i}{\partial \mathbf{x}_i} = q_i' \times \boldsymbol{\delta}$$

where $\boldsymbol{\delta} = \boldsymbol{\gamma}$ for the linear form or $\boldsymbol{\delta} = \tau\boldsymbol{\beta}$ for the τ form. The computation of the marginal effects accounts for the splitting effect and for any overlap between the variables in the splitting model and in the base Poisson or negative binomial model. An example appears below.

NOTE: Because the terms in the marginal effect enter with different signs, it is possible for the marginal effect of a variable to have the opposite sign from the corresponding coefficient in the Poisson regression. This is likely to occur in samples which have a very large proportion of zeros, since in this case, the q_i is likely to be quite large.

E43.6.2 ZIP Models with Latent Heterogeneity

The Zero Inflated Poisson model is extended to allow normal heterogeneity in the regression. Thus, in the model,

$$\text{Prob}[y_i = 0] = [1 - q_i]\text{R}[\,\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, 0\,] + q_i,$$

$$\text{Prob}[y_i = j] = [1 - q_i]\text{R}[\,\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, j\,],\ j > 0$$

where R[.] is the Poisson probability with mean function exp[$\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$,] and q_i is the regime splitting model described above. The same four forms of the model for q_i are available, logistic and normal (probit) for the distribution and linear or τ for the argument of the regime probability.

The basic model command is

POISSON ; Lhs = y ; Rhs = ... ; ZIP ; Het ; ... $

The different forms of the splitting model are the ZIP(τ) form, in which

$$q_i = F[\,\tau\,\boldsymbol{\gamma}'\mathbf{z}_i\,]$$

and the ZIP form $q_i = F[\boldsymbol{\gamma}'\mathbf{z}_i\,]$

and F(.) may be either a probit or logit equation. The ZIP(τ) forms are requested with

POISSON ; Lhs = y ; Rhs = x ; Heterogeneity ; ZIP ; ...$
and **POISSON ; Lhs = y ; Rhs = x ; Heterogeneity ; ZIP ; Logit ; ...$**

The default form is the probit model, and **; Logit** requests the logit model instead. (Note that this reverses the default in the model without heterogeneity, where logit is the default.) The ZIP forms are requested with the **; Rh2 = ...** specification,

POISSON ; Lhs = y ; Rhs = x ; Rh2 = z ; Heterogeneity ; ZIP $
and **POISSON ; Lhs = y ; Rhs = x ; Rh2 = z ; Heterogeneity ; ZIP ; Logit $**

NOTE: This model does not extend to the negative binomial, gamma or GP models. Each of these already accounts for heterogeneity and overdispersion, so adding this feature to a model which already has heterogeneity and excess zeros greatly overspecifies the model. The estimation process will break down in these instances.

E43.6.3 A ZIP Model with Endogenous Zero Inflation

Finally, a model with an endogenously determined splitting model would be

$$q_i = \Phi[\,\boldsymbol{\gamma}'\mathbf{z}_i + u_i]$$

$$\text{Prob}[y_i = 0] = [1 - q_i]\text{R}[\,\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, 0\,] + q_i,$$

$$\text{Prob}[y_i = j] = [1 - q_i]\text{R}[\,\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, j\,],\ j > 0$$

and u_i is correlated with ε_i. This model is requested with

POISSON ; Lhs = y ; Rhs = x ; Rh2 = z
; Het ; ZIP ; Correlation $

There is only one form for this model, the **POISSON ; Zip = normal**, index model with latent heterogeneity correlated with the latent effect in the splitting equation.

E43.6.4 Output for the ZIP Models

The ZIP models begin with the full set of output for the underlying model, Poisson or negative binomial. If you request it with **; OLS**, this will include the initial OLS results, the Poisson regression model, if requested, the negative binomial model, then, finally, the results for the ZIP model. The initial Poisson and, if requested, negative binomial model(s) will be fit ignoring any constraints in order to obtain starting values. The final output presents the table of statistical results for the estimated coefficients. The Poisson or negative binomial parameters will be followed by the parameter(s) for the splitting model. The leading diagnostics table contains a number of related statistics. A table gives the following comparison of the original model and the zero altered model:

Estimates of prob[$Y = 0 \mid \mathbf{x}$ = sample means],
Number of zero observations, actual and predicted using $N\times$Prob[Y=0],
Log likelihood.

Note that although in most cases, the log likelihood function for the zero inflated model will exceed that for the unaltered model, the two values are not comparable because the base model is not obtainable from the ZIP model by restricting the coefficients in the latter. Finally, Vuong's statistic is presented.

The saved results from the ZIP model are:

Matrices:	*b* and *varb*	
	zaptau =	the parameters of the splitting model, τ or $\boldsymbol{\gamma}$
Scalars:	*ybar* and *sy*	for the Lhs variable
	nreg and *kreg*	for the estimated model
	exitcode	
	alpha and *varalpha*	for an estimated negative binomial model
	tau and *vartau*	if you fit the ZIP(τ) model
Last Model:	*b_variable*	for the Poisson or negative binomial model
	alpha	if you fit the negative binomial model
	c_variable	for the ZIP model, the coefficients on the Rh2 variables

Last Function: Conditional mean, $(1 - q_i)\lambda_i$.

E43.6.5 Application

To illustrate the model, we have fit the least elaborate Poisson specification, then the same model using the generalized Poisson (P) format.

```
SAMPLE      ; All $
POISSON     ; Lhs = docvis ; Rhs = one,age,hhninc
            ; Rh2 = one,age,female,married,hhkids
            ; ZIP ; Partial Effects  $
PARTIALS    ; Effects: age/female ; Summary $
POISSON     ; Model = GPP
            ; Lhs = docvis ; Rhs = one,age,hhninc
            ; Rh2 = one,age,female,married,hhkids
            ; ZIP ; Partial Effects  $
```

```
-----------------------------------------------------------------------------
Zero Altered Poisson      Regression Model
Logistic distribution used for splitting model.
ZAP term in probability is F[tau x Z(i)     ]
Comparison of estimated models
            Pr[0|means]       Number of zeros        Log-likelihood
Poisson          .04703   Act.= 10135 Prd.=  1285.2   -105125.23124
Z.I.Poisson      .36394   Act.= 10135 Prd.=  9945.0    -83907.65103
Note, the ZIP log-likelihood is not directly comparable.
ZIP model with nonzero Q does not encompass the others.
Vuong statistic for testing ZIP vs. unaltered model is     46.9219
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Poisson/NB/Gamma regression model
Constant|    1.19474***      .00623   191.72  .0000     1.18253   1.20695
     AGE|     .01296***      .00011   114.08  .0000      .01274    .01318
  HHNINC|    -.53791***      .00834   -64.47  .0000     -.55427   -.52156
        |Zero inflation model
Constant|     .53149***      .06009     8.84  .0000      .41371    .64926
     AGE|    -.01907***      .00133   -14.38  .0000     -.02167   -.01647
  FEMALE|    -.61374***      .02635   -23.29  .0000     -.66539   -.56209
 MARRIED|    -.11832***      .03361    -3.52  .0004     -.18420   -.05244
  HHKIDS|     .25065***      .03017     8.31  .0000      .19152    .30977
--------+--------------------------------------------------------------------
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1475
Scale Factor for Marginal Effects  3.1475
Effects of common variables in two part
models are added to obtain partial effect.
```

```
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect         Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index Function in Count Probability
     AGE|     .06119***      .00151    40.59  .0000      .05824    .06414
  HHNINC|   -1.69308***      .02709   -62.49  .0000    -1.74618  -1.63998
        |Zero Inflation Probability
     AGE|     .06119***      .00151    40.59  .0000      .05824    .06414
  FEMALE|     .65679***      .02720    24.14  .0000      .60347    .71011
 MARRIED|     .12662***      .03596     3.52  .0004      .05614    .19709
  HHKIDS|    -.26823***      .03213    -8.35  .0000     -.33119   -.20526
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
---------------------------------------------------------------------
Partial Effects for Zero Inflation Model for Counts
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .06119     .00146   42.05      .05834      .06404
   *  FEMALE         .66395     .02809   23.64      .60890      .71901
---------------------------------------------------------------------
-----------------------------------------------------------------------------
Generalized Poisson (P) Model
Dependent variable                DOCVIS
Log likelihood function    -59787.37107
Restricted log likelihood -105125.23124
Chi squared [   1 d.f.]     90675.72034
Significance level               .00000
McFadden Pseudo R-squared      .4312748
Estimation based on N =  27326, K =  10
Inf.Cr.AIC  = 119594.7 AIC/N =    4.377
Wald test for dispersion      17.6 [ 1]
Zero Inflated Generalized Poisson Model
Logit Zero Inflation Probability Model
Zeros in Sample: Actual           10135
Zeros in Sample: Predicted        10150
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .71230***      .04894    14.55  .0000      .61637    .80823
     AGE|     .01798***      .00092    19.53  .0000      .01618    .01978
  HHNINC|    -.56393***      .05016   -11.24  .0000     -.66224   -.46563
        |Zero Inflation Logit Probability = Logit(c*z)
Constant|    -.40029**       .16205    -2.47  .0135     -.71791   -.08267
     AGE|    -.01569***      .00374    -4.20  .0000     -.02301   -.00836
  FEMALE|   -1.55244***      .12005   -12.93  .0000    -1.78774  -1.31714
 MARRIED|    -.39331***      .07790    -5.05  .0000     -.54600   -.24063
  HHKIDS|     .56189***      .07054     7.97  .0000      .42364    .70014
        |Dispersion parameter in generalized Poisson model
Constant|     .77613***      .05728    13.55  .0000      .66385    .88840
        |Nesting Parameter for P form of Generalized Poisson
       P|    1.57807***      .05272    29.93  .0000     1.47474   1.68139
--------+--------------------------------------------------------------------
```

```
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1806
Scale Factor for Marginal Effects  3.1806
Effects of common variables in two part
models are added to obtain partial effect.
--------+-------------------------------------------------------------
        |     Partial      Standard             Prob.      95% Confidence
  DOCVIS|      Effect         Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------
        |Index Function in Count Probability
     AGE|     .06416***      .00418    15.36  .0000      .05597    .07235
  HHNINC|   -1.79366***      .16039   -11.18  .0000    -2.10801  -1.47931
        |Zero Inflation Probability
     AGE|     .06416***      .00418    15.36  .0000      .05597    .07235
  FEMALE|     .69047***      .03782    18.26  .0000      .61634    .76459
 MARRIED|     .17493***      .03507     4.99  .0000      .10619    .24367
  HHKIDS|    -.24991***      .02950    -8.47  .0000     -.30772   -.19209
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

E43.6.6 Technical Details

To formulate the log likelihood and gradient for the ZIP models, let

$$q_i = F(\boldsymbol{\gamma}'\mathbf{z}_i) \text{ for the ZIP models}$$

and

$$q_i = F(\tau\boldsymbol{\beta}'\mathbf{x}_i) \text{ for the ZIP}(\tau) \text{ model,}$$

where $F(t)$ is either the cumulative normal probability, $\Phi(t)$, for the probit model or the cumulative logistic probability, $\Lambda(t)$ for the logit model. Let $f(t)$ denote either the Poisson(λ_i), the negative binomial (λ_i,θ) or the gamma (λ_i,P) probability density function. (This produces 12 possible models.) Then, the probability density function for the observed random variable, y_i, is

$$p(y_i) = p_i = (1 - q_i)f(y_i) + \mathbf{1}(y_i = 0)q_i,$$

so, the log likelihood is simply

$$\log L = \Sigma_i \log p(y_i).$$

To obtain the gradient, let $\boldsymbol{\beta}^*$ equal either $\boldsymbol{\beta}$ for the Poisson model, ($\boldsymbol{\beta}$,θ) for the negative binomial model or ($\boldsymbol{\beta}$,P) for the gamma model. Then, each term in

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}*} = \Sigma_i (\partial \log p_i/\partial \boldsymbol{\beta}^*)$$

is

$$\frac{\partial \log p_i}{\partial \boldsymbol{\beta}*} = (1/p_i)[(1 - q_i)f(y_i)\{\partial \log f(y_i)/\partial \boldsymbol{\beta}^*\} + \{\mathbf{1}(y_i = 0) - f(y_i)\}\{\partial q_i/\partial \boldsymbol{\beta}^*\}].$$

The derivatives of log$f$(y_i) were given earlier.

NOTE: These derivatives are approximated numerically for the gamma model.

The cross derivatives, $\partial q_i/\partial \boldsymbol{\beta}^*$ will equal **0** in the ZIP model, or $\tau \mathbf{x}_i q_i'$ for the ZIP(τ) model with a trailing zero for θ or P if $f(y_i)$ is the negative binomial or gamma model, since these parameters do not enter q_i. (The inner derivative, q_i', is either the standard normal density, ϕ_i for the probit model, or $\Lambda_i(1-\Lambda_i)$ for the logit model.) Finally, the parameters of the ZIP model are either $\boldsymbol{\gamma}$, a vector, in the ZIP model or τ, a scalar, in the ZIP(τ) model. Denoting these generically as $\boldsymbol{\gamma}$, we have

$$\partial \log p_i/\partial \boldsymbol{\gamma} \;=\; [\mathbf{1}(y_i = 0)](q_i'/p_i)(\boldsymbol{\beta}'\mathbf{x}_i)$$

for the ZIP(τ) model. For the ZIP model, $\boldsymbol{\beta}'\mathbf{x}_i$ is replaced with $\mathbf{z}_i$, the vector of covariates. The second derivatives are fairly complicated, but in *LIMDEP*'s implementation, the BHHH estimator is used, instead, as a convenient expedient.

For the ZIP specification, a natural set of starting values for the parameters is provided by the probit or logit and independent Poisson or negative binomial estimates. (The Poisson values are used to start the gamma model.) In the ZIP(τ) case, the Poisson or negative binomial model can be used for the regression parameters. One could then choose a value for τ which would produce approximately the correct probability for zero. An alternative possibility would be to estimate τ by fitting a probit or logit model to the binary indicator $\mathbf{1}(y_i = 0)$ with the single covariate equal to the Poisson estimates of $\boldsymbol{\beta}'\boldsymbol{x}_i$ (only to get the right sign and approximately the right magnitude on τ; this is not a consistent estimator). Save for a few badly identified cases found by experimentation in which no solution could be found, convergence of the DFP or Broyden algorithms appears to be routine for these models.

E43.7 Hurdle Models

A case related to the ZIP model is known as the hurdle model. Hurdle models arise when the 'zero or positive' decision is different from the count decision. One can think in terms of two decisions – the health care utilization data examined in the previous chapter provides a good example. One decision is whether to 'be a participant.' This is equivalent to a decision as to whether the count will be zero or positive. The second decision is how many, given that the count will be positive. In the health care utilization case, we can consider two types of individuals, those who do not intend to visit a doctor or the hospital and those who do. For the latter, the observed count is the number of visits, given that the number of visits will be positive. The formal model is

z = 0 if the response will be zero, 1 if the response will be positive

y = the count of occurrences given that the count will be positive.

This model consists of two parts, which may be dependent or independent:

Prob[z = 0 or 1|$\mathbf{x}_i$] = a probit or logit model

Prob[y = j | y > 0|$\mathbf{x}_i$] = a count data model with truncation at zero.

An alternative approach to the excess zeros case is the hurdle model presented by Mullahy (1986) and Creel and Loomis (1990). (A complete description may be found in Winkelmann (2000).) Logically, the model arises from two simultaneous process. The first is a 'hurdle,' in which the individual 'decides' whether y will equal zero or some value greater than zero. The second is a conditional count model in which the number of occurrences is conditional on that number being positive. The hurdle model is very similar to the ZIP model:

$$\begin{aligned}\text{Prob}(Y_i = 0) &= f_i(0)\\ \text{Prob}[Y_i = j] &= \frac{1-f_i(0)}{1-P_i(0)}P_i(j),\ j = 1,2,...\\ &= A_i(0)P_i(j)\end{aligned}$$

where $f_i(0)$ is the probability of the zero outcome, $P_i(j)$ is the probability of the nonzero outcomes conditioned on the outcome being greater than zero, and the subscript i indicates dependence on covariates $\mathbf{z}_i$ for f_i and $\mathbf{x}_i$ for P_i. The combination of the two produces the unconditional distribution above. The hurdle model can be assembled from any desired binary choice model and count model.

Let d_i denote a binary indicator of whether the observed count is zero or positive. The log likelihood function separates the probabilities into two simple parts:

$$\begin{aligned}\log L &= \sum\nolimits_{d=0} \log f_i(0)\\ &+ \sum\nolimits_{d=1} \log[1 - f_i(0)] - \log[1 - P_i(0)] + \log P_i(j).\end{aligned}$$

The four terms of the log likelihood partition into two log likelihoods,

$$\begin{aligned}\log L &= \sum\nolimits_{d=0} \log f_i(0) + \sum\nolimits_{d=1} \log[1 - f_i(0)]\\ &+ \sum\nolimits_{d=1} \log P_i(j) - \log[1 - P_i(0)].\end{aligned}$$

The first term is the log likelihood for a binary choice model – probit, logit, complementary log log, etc. (See Chapter E27.) The second part is the log likelihood for a count distribution that is truncated at zero. We have already presented this model in Section E42.2.3. The end result is that the hurdle model can be fit in two simple parts using models already presented. A natural formulation would be the logit binary choice model coupled with the Poisson model for the positive counts. (Mullahy's original presentation of this model suggested an $f_i(0)$ that was a constant – no covariates. This could be obtained in our formulation simply by specifying **; Rhs = one** in the binary choice model.

The hurdle model induces over- or underdispersion in the distribution, but in a nonconstant fashion. Winkelmann presents the following convenient result:

$$\text{Var}_i[Y_i] = \text{E}_i[Y_i] + \frac{1-A_i(0)}{A_i(0)}\{\text{E}_i[Y_i]\}^2$$

where now, the subscript indicates dependence on both $\mathbf{z}_i$ and $\mathbf{x}_i$. Since $A_i(0)$ can exceed one, this model can induce underdispersion. (The mean function and $A_i(0)$ are functionally dependent in such a way that no combination of parameters produces a negative variance.) Underdispersion occurs if zeros are less frequent than the parent (Poisson or negative binomial) model would predict.

The conditional mean in this model can be obtained by making convenient use of the fact that the sums from one to infinity are the same if the zero outcome is included. This produces the conditional mean function

$$\mathrm{E}_i[Y_i] \quad = \quad \frac{1 - f_i(0)}{1 - P_i(0)}\lambda_i\,,$$

where $\lambda_i \quad = \quad \exp(\boldsymbol{\beta}'\mathbf{x}_i)$

and $P_i(0) \quad = \quad \exp(-\lambda_i)$ for the Poisson model and

$= \; [\theta/(\theta + \lambda_i)]^\theta$ for the negative binomial model.

The marginal effects obtained by differentiation, after a bit of algebra and collecting terms, are

$$\frac{\partial E_i[Y_i]}{\partial \mathbf{x}_i} \quad = \quad \frac{1 - f_i(0)}{1 - P_i(0)}\lambda_i\,\boldsymbol{\beta} \times \left(1 + \frac{\lambda_i}{1 - P_i(0)}\frac{\partial P_i(0)}{\partial \lambda_i}\right).$$

The latter derivatives are

$$\frac{\partial P_i(0)}{\partial \lambda_i} \quad = \quad -P_i(0) \;\text{ for the Poisson model and}$$

$$= \; -\,[P_i(0)]^{(\theta+1)/\theta}.$$

(The Poisson model results when $\theta \rightarrow \infty$ so the results are consistent.) Finally,

$$\frac{\partial E_i[Y_i]}{\partial \mathbf{z}_i} \quad = \quad \frac{-1}{1 - P_i(0)}\lambda_i \times \frac{\partial f_i(0)}{\partial \mathbf{z}_i}.$$

The three cases supported here are

logit: $f_i(0) \; = \; \Lambda(-\boldsymbol{\delta}'\mathbf{z}_i)$, $\quad \dfrac{\partial f_i(0)}{\partial \mathbf{z}_i} \; = \; -\Lambda(-\boldsymbol{\delta}'\mathbf{z}_i)[1 - \Lambda(-\boldsymbol{\delta}'\mathbf{z}_i)]\,\boldsymbol{\delta}$

probit: $f_i(0) \; = \; \Phi(-\boldsymbol{\delta}'\mathbf{z}_i)$, $\quad \dfrac{\partial f_i(0)}{\partial \mathbf{z}_i} \; = \; -\phi(-\boldsymbol{\delta}'\mathbf{z}_i)\,\boldsymbol{\delta}$

complementary log log: $f_i(0) \; = \; \exp(-\exp(\boldsymbol{\delta}'\mathbf{z}_i)) \; = \; \exp(-\gamma_i)$, $\quad \dfrac{\partial f_i(0)}{\partial \mathbf{z}_i} \; = \; -\gamma_i f_i(0)\,\boldsymbol{\delta}$

(The third of these, suggested by Mullahy (1986) is convenient as it allows a straightforward test of hurdle effects against the Poisson null, which is nested. In the other two models, the Poisson model is not nested, so the test is less convenient. See below for details) Finally, when the regime model and the count model have variables in common, the two effects are added.

Hurdle models can be fit with a single instruction rather by fitting the parts separately. The command is

POISSON ; Lhs = dependent variable
; Rhs = independent variables
; Hurdle $

In this base case

- the count model is assumed to be Poisson,
- the hurdle equation is assumed to be logistic,
- variables that enter the hurdle equation are the same as in the Poisson equation.

Several alternative specifications may be chosen: Use

; Model = Negbin

(or change the command to **NEGBIN**) for the negative binomial count model

Use **; Normal** for the probit model
; Cloglog for the complementary log log model. (See below.)

Use **; Rh2 = list of variables** to specify the variables in **z** explicitly.

Other specifications, including

; Keep = name	to retain fitted values (using conditional mean)
; Res = name	to retain residuals
; Prob = name	to retain predicted probabilities for observed outcomes
; List	to display predictions, residuals, etc.
; Output = value	to control technical output during iterations
; Partial Effects	
; Test: spec	to define Wald tests
; Rst = list	to specify fixed value and equality restrictions
; CML: spec	to define a constrained maximum likelihood estimator

and so on, are all available as in other count models already discussed.

The estimation results saved by this estimator are as usual:

Matrices: *b* = coefficient vector – this will be all parameters, including θ if the negative binomial model is fit
varb = asymptotic covariance matrix

Scalars: *logl* = scalar log likelihood
nreg = number of observations
kreg = number of variables in *x* - does not include *z*.

Last Function: hurdle conditional mean function

E43.7.1 Testing for Hurdle Effects

There are two ways to test for hurdle effects in this setting. Under the assumption of a Poisson count model, complementary log log hurdle model, and identical variables in the two equations, the probability which enters the log likelihood becomes

$$\text{Prob}(Y_i = 0) = \exp(-\gamma_i)$$

$$\text{Prob}[Y_i = j] = \frac{1-\exp(-\gamma_i)}{1-\exp(-\lambda_i)}\frac{\exp(-\lambda_i)\lambda_i^{yi}}{y_i!}, j = 1,2,...$$

where $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$

and $\gamma_i = \exp(\boldsymbol{\delta}'\mathbf{x}_i)$.

In this case, the hurdle model becomes the Poisson model if $\boldsymbol{\delta} = \boldsymbol{\beta}$. Since this is a simple parametric restriction, and the models are nested, a Wald, likelihood ratio, or LM test could be used. Here is an approach:

```
NAMELIST    ; x = the Rhs variables $
CALC        ; k = Col(x) $
POISSON     ; Lhs = ... ; Rhs = x  $
MATRIX      ; bp = b $
CALC        ; lp = logl $
POISSON     ; Lhs = ... ; Rhs = x
            ; Hurdle ; Cloglog $
CALC        ; lh = logl $
MATRIX      ; bh = b ; vh = varb $
```

The next three commands carry out the LM, LR and Wald tests, respectively.

```
POISSON     ; Lhs = ... ; Rhs = x
            ; Hurdle ; Cloglog
            ; Start = bp,bp ; Maxit = 0 $
CALC        ; List ; lrtest = 2*(lh - lp) ; 1 - Chi(lrtest,k) ; Wald = 0 $
MATRIX      ; d = i2’ * bh ; ik = Iden(k) ; ikm = -ik ; i2 = [ik/ikm]
            ; Wald = d’*< i2’[vh]i2>*d $
CALC        ; List ; Wald ; 1 - Chi(Wald,k) $
```

When other forms are used, the models are nonnested. In these cases, the classical testing procedures no longer have the limiting chi squared distributions, and are no longer useable. One possibility is the Vuong statistic for testing the nonnested models that was suggested in the previous section. The statistic is

$$V = \sqrt{n}\,\overline{m} \,/\, s_m$$

where $m_i = \log[f_h(y_i)/fp(y_i)]$,

$f_h(\bullet)$ and $fp(\bullet)$ are densities for the hurdle and Poisson models, respectively, and $\overline{\mathrm{m}}$ and s_m are the sample mean and standard deviation for the sample of m_is. In principle, asymptotically, the statistic is distributed as standard normal, so its value may be compared to the critical value from the standard normal distribution, e.g., 1.96. The test is directional; large positive values favor *fh* while large negative values favor *fp*. We do note that the hurdle model involves an extension of the Poisson model with the addition of additional parameters. As such, intuition suggests that it is unlikely that the test would ever favor the Poisson model. If the intuition is right, then the asymptotic behavior of the statistic may not be as assumed here, and the validity of the test becomes questionable. With that caveat, the following procedure could be used for this test – it is not part of the standard output for the hurdle model:

```
NAMELIST    ; x = the Rhs variables $
POISSON     ; Lhs = ... ; Rhs = x
            ; Prob = fp $
POISSON     ; Lhs = ... ; Rhs = x
            ; Prob = fh
            ; Hurdle ; ... (whatever other specification) $
CREATE      ; mi = Log(fh/fp) $
CALC        ; List ; v = Sqr(n) * Xbr(mi) / Xdv(mi) $
```

E43.7.2 Heterogeneity and Endogeneity

Heterogeneity may be entered into the conditional mean of the count model in the same fashion as the ZIP models. With

; Heterogeneity

The mean function becomes

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i)$$

As before, the log likelihood is now maximized using Hermite quadrature. Results are essentially the same as before, with the additional results related to the distribution of ε_i. The hurdle effect becomes endogenous if ε_i is correlated with u_i in the hurdle equation. Once again, this model parallels the zero inflation model. The model with endogenous hurdle effects is requested with

; Heterogeneity ; Correlated.

E43.7.3 Applications

Like the zero inflation model, there are many different combinations of hurdle equation, count model, heterogeneity and endogeneity. Testing procedures for distinguishing some of them statistically were suggested earlier. The following illustrates two specifications, the base case Poisson with exogenous hurdle effect and the Poisson model with endogenous hurdle effect.

```
SAMPLE    ; All $
POISSON   ; Lhs = docvis ; Rhs = one,age,hhninc
          ; Rh2 = one,age,female,married,hhkids
          ; Hurdle ; Partial Effects $
POISSON   ; Lhs = docvis ; Rhs = one,age,hhninc
          ; Rh2 = one,age,female,married,hhkids
          ; Hurdle ; Partial Effects ; Heterogeneity ; Correlated $
```

```
-----------------------------------------------------------------------------
Poisson hurdle model for counts
Dependent variable               DOCVIS
Log likelihood function    -84270.74039
Restricted log likelihood -105125.23124
LOGIT  hurdle equation
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        | Parameters of count model equation
Constant|    1.21788***      .00594   205.10  .0000     1.20624   1.22952
     AGE|     .01300***      .00011   119.36  .0000      .01279    .01322
  HHNINC|    -.58459***      .00800   -73.07  .0000     -.60027   -.56891
        | Parameters of binary hurdle equation
Constant|    -.60129***      .05889   -10.21  .0000     -.71671   -.48586
     AGE|     .02019***      .00130    15.50  .0000      .01764    .02275
  FEMALE|     .60428***      .02586    23.37  .0000      .55359    .65496
 MARRIED|     .11115***      .03305     3.36  .0008      .04637    .17593
  HHKIDS|    -.24273***      .02960    -8.20  .0000     -.30075   -.18471
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point    .0130
Scale Factor for Marginal Effects  3.0552
Effects of common variables in two part
models are added to obtain partial effect.
--------+--------------------------------------------------------------------
        |     Partial    Standard            Prob.      95% Confidence
  DOCVIS|      Effect       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        | Effects in Count Model Equation
     AGE|     .03972         .03436     1.16  .2476     -.02762    .10707
  HHNINC|   -1.78606        1.54499    -1.16  .2477    -4.81419   1.24207
        | Effects in Binary Hurdle Equation
     AGE|     .02216***      .00143    15.50  .0000      .01935    .02496
  FEMALE|     .66299***      .02837    23.37  .0000      .60739    .71860
 MARRIED|     .12195***      .03626     3.36  .0008      .05087    .19302
  HHKIDS|    -.26631***      .03248    -8.20  .0000     -.32997   -.20265
     AGE|     .06188*        .03448     1.79  .0727     -.00571    .12946
```

```
--------+-------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------

----------------------------------------------------------------------
Hurdle Poisson with Normal Hetero.
Dependent variable               DOCVIS
Log likelihood function    -59634.85591
Restricted log likelihood -227531.30639
Vuong Stat. vs. Poisson =      37.38614
Vuong test favors extended model.
Hurdle model determines truncation (PROBIT)
Endogenous censoring model. See RHO below.
Predicted zeros: Poisson= 27325, ZIP=     0
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS| Coefficient       Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------
        | Parameters of Poisson Probability
Constant|    1.07365***      .04952    21.68  .0000      .97659   1.17071
     AGE|     .01036***      .00083    12.54  .0000      .00874    .01198
  HHNINC|    -.47608***      .04909    -9.70  .0000     -.57230   -.37986
        | Parameters of Probit/Logit ZIP/Hurdle Equation
Constant|     .36125***      .03641     9.92  .0000      .28989    .43262
     AGE|    -.01237***      .00081   -15.35  .0000     -.01395   -.01079
  FEMALE|    -.38250***      .01708   -22.39  .0000     -.41599   -.34902
 MARRIED|    -.05677***      .01955    -2.90  .0037     -.09509   -.01846
  HHKIDS|     .15521***      .01790     8.67  .0000      .12012    .19030
        | Correlation between hurdle and count eqns.
     Rho|     .46429         .42738     1.09  .2773     -.37335   1.30193
        | Standard Deviation of Heterogeneity
   Sigma|     .99609***      .00921   108.15  .0000      .97804   1.01414
--------+-------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics
computed at the means of the variables.
Separate effects are shown first followed
by the sum of the two effects for variables
which are in both Poisson and Probit models
Estimated value of E[y|x] computed at the
means is  3.12934.
--------+-------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------
        | Parameters of Poisson Probability
     AGE|     .03243         .06571      .49  .6216     -.09635    .16122
  HHNINC|   -1.48981***      .31414    -4.74  .0000    -2.10551   -.87411
        | Parameters of Probit/Logit ZIP/Hurdle Equation
     AGE|     .00270         .00741      .36  .7162     -.01184    .01723
  FEMALE|     .08331         .22963      .36  .7167     -.36675    .53338
 MARRIED|     .01237         .03432      .36  .7186     -.05490    .07963
  HHKIDS|    -.03381         .09321     -.36  .7168     -.21649    .14888
        | Combined effect of two terms
     AGE|     .03513         .06640      .53  .5968     -.09501    .16527
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

E43.7.4 Technical Details

Let d_i denote a binary indicator of whether the observed count is zero or positive. The log likelihood function separates the probabilities into two simple parts:

$$\log L = \sum_{d=0} \log f_i(0)$$

$$+ \sum_{d=1} \log[1 - f_i(0)] - \log[1 - P_i(0)] + \log P_i(j)$$

The four terms of the log likelihood partition into two log likelihoods,

$$\log L = \sum_{d=0} \log f_i(0) + \sum_{d=1} \log[1 - f_i(0)]$$

$$+ \sum_{d=1} \log P_i(j) - \log[1 - P_i(0)]$$

The derivatives for the hurdle model are quite simple. Six mixtures of models are supported. For the hurdle equation, let

$$w_i = \boldsymbol{\delta}'\mathbf{z}_i.$$

Then,

$$f_i(0) = \text{logit model} = \Lambda(-w_i)$$

$$= \text{probit model} = \Phi(-w_i)$$

$$= \text{complementary log log model} = \exp(-\exp(w_i))$$

For the count model,

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

$$P_i(j) = \text{Poisson model } (\lambda_i)$$

$$= \text{negative binomial model } (\lambda_i).$$

The necessary terms for differentiation of these functions appear elsewhere in this chapter. The BHHH estimator is used for the estimator of the asymptotic covariance matrix of the MLE. As noted, this model can be estimated in parts, by fitting a binary choice model to the dependent variable obtained as **1**(count > 0) and a truncated (at zero) count model to the observations with nonzero counts. The identical parameter estimates will be obtained if you do so. The advantage here, aside from the simplicity of the combined command, is the ability to impose various restrictions and use different procedures for testing hypotheses.

E44: Panel Data Models for Counts

E44.1 Introduction

This chapter describes estimators for models for counts based on panel data. The basic formulation, once again, is the Poisson regression model. For a discrete random variable, Y, observed frequencies, y_i, $i = 1,...,n$, where y_i is a nonnegative integer count, and regressors $\mathbf{x}_i$,

$$\text{Prob}(Y = y_i) = \frac{\exp(\lambda_i)\lambda_i^{y_i}}{y_i!},\ y_i = 0,1,...;\ \log \lambda_i = \boldsymbol{\beta}'\mathbf{x}_i.$$

In this model, λ_i is both the mean and variance of y_i that is $E[y_i|\mathbf{x}_i] = \text{Var}[y_i|\mathbf{x}_i] = \lambda_i$. The partial effects in this nonlinear regression model are

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i \boldsymbol{\beta}.$$

The negative binomial regression model is an extension of the Poisson regression model that results from the introduction of a certain kind of unobserved individual heterogeneity into the Poisson model; the negative binomial model arises as a modification of the Poisson model in which the mean is μ_i, respecified so that

$$\log \mu_i = \log\lambda_i + \varepsilon_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

where $\exp(\varepsilon_i)$ has a gamma distribution with mean 1.0 and variance α. The resulting unconditional distribution (derived in Section E41.4.5) is

$$\text{Prob}[Y = y_i] = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i + 1)} u_i^{\theta}(1-u_i)^{y_i},\ u_i = \theta / (\theta + \lambda_i).$$

This model has an additional parameter, $\alpha = 1/\theta$, such that $\text{Var}[y_i] = E[y_i]\{1 + \alpha E[y_i]\}$.

This chapter presents *LIMDEP*'s implementation of panel data models for Poisson and negative binomial regressions. The topics described in this chapter are

- Panel data models
- Commands for estimating panel data models
- Fixed effects models
- Random effects models
- Random parameters models
- Latent class models
- GMM estimators for count models with panel data

E44.2 Panel Data Models for Count Data

The full range of *LIMDEP*'s panel data estimation routines are provided for the Poisson, and negative binomial regression models and for some of the formulations of the zero inflation models.

NOTE: Save for the exceptions explicitly noted below, the panel data treatments are not supported for the gamma model or for the extensions of the count data models, including the sample selection models, and the underreporting models. Limited cases for the heterogeneity models are noted below. The ZIP/Logit and ZINB/Logit models are supported, but the (τ) forms are not.

E44.2.1 Fixed Effects

For the fixed effects case,

$$\log\lambda_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} \;\; (+ \varepsilon_{it} \text{ for the negative binomial model }).$$

The difference here is that the model cannot be fit by least squares using deviations from group means. Two approaches are used instead. One possibility is to use a conditional maximum likelihood approach – the model conditioned on the sum of the observations is free of the fixed effects and has a closed form. This is provided for both Poisson and negative binomial models. A second approach is direct, brute force estimation of the full model including the fixed effects. Neglecting the latent log gamma heterogeneity in the negative binomial model, write the fixed effects model as

$$\log\lambda_{it} = \alpha_i d_{it} + \boldsymbol{\beta}'\mathbf{x}_{it},\; i = 1,...,N,\; t = 1,...,T_i,$$

where α_i is the coefficient on a binary variable, d_i, which indicates membership in the ith group. The panel is assumed to consist of N groups with T_i observations in the ith group. The panel need not be balanced; T_i may vary across groups. This model is estimated in two ways. The *conditional* estimators are obtained by using the conditional joint distribution, $f(y_{i1},y_{i2},...,y_{iT}|\Sigma_t y_{it})$. (See Griliches, Hall, and Hausman (1984).) The resulting density is a function of $\boldsymbol{\beta}$ alone, which is then estimated by (conditional) maximum likelihood. The *unconditional* estimator is obtained by a direct maximization of the full log likelihood function and estimating all parameters including the group specific constants.

E44.2.2 Random Effects

The random effects model is

$$\log \lambda_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + u_i.$$

Once again, the approach used for the linear model, in this case, FGLS, is not useable. The approach is to integrate out the random effect and estimate by maximum likelihood the parameters of the resulting distribution (which, it turns out, is the negative binomial model when the kernel is Poisson). Both fixed and random effects models are provided for the Poisson and negative binomial (gamma mixture) formulations. The bulk of the received literature on random effects is in the Poisson model. We also present models for random effects based on the normal distribution.

The random effects model for the count data framework is

$$\log\lambda_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + u_i,\ i = 1,...,N,\ t = 1,...,T_i,$$

where u_i is a random effect for the ith group such that $\exp(u_i)$ has a gamma distribution with parameters(θ,θ). Thus, $E[\exp(u_i)]$ has mean 1 and variance $1/\theta = \alpha$. This is the framework which gave rise to the negative binomial model earlier, so that, with minor modifications, this is the estimating framework for the Poisson model with random effects. For the negative binomial model, Hausman, et al. proposed the following approach: We begin with the Poisson model with the random effects specification shown above. The random term, u_i is distributed as gamma with parameters(θ_i,θ_i), which produces the negative binomial model with a parameter that varies across groups. Then, it is assumed that $\theta_i/(1+\theta_i)$ is distributed as beta(a_n,b_n), which layers the random group effect onto the negative binomial model. The random effect is added to the negative binomial model by assuming that the overdispersion parameter is randomly distributed across groups.

The two random effects models discussed above may be modified to use the normal distribution for the random effect instead of the gamma, with $u_i \sim N[0,\sigma^2]$. For the Poisson model, this is an alternative to the log-gamma model which gives rise to the negative binomial model. It is also essentially the same as the model of latent heterogeneity discussed in Section E42.4.1. The negative binomial model is much more involved than this, and the normal model is a considerably simpler alternative.

E44.2.3 Random Parameters

We provide a full random parameters formulation for both Poisson and negative binomial models,

$$\log \lambda_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + u_i \text{ (} + \varepsilon_{it} \text{ for the negative binomial model),}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i,\ \mathbf{z}_i = \text{ a set of observed covariates,}$$

$$\mathbf{v}_i \sim \text{ joint standard normal, uniform, or triangular.}$$

The random parameters models are described in detail below and in Chapter R24.

E44.2.4 Latent Class Models

The Poisson model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is

$$\text{Prob}[Y_{it} = y_{it}|\ \lambda_{it}] \;=\; \exp[-\lambda_{it}] \times \lambda_{it}^{yit} \;/\; y_{it}! \;=\; P(i,t)$$

where $\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{\text{it}})$

is the conditional [on $\mathbf{x}_{it}$] mean, as usual. Henceforth, we use the term ‘group’ to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. The following extends to the negative binomial model as well, but for the moment, we focus on the Poisson model.

Unobserved heterogeneity in the distribution of Y_{it} is assumed to impact the mean (and variance) λ_{it}. The continuous distribution of the heterogeneity is approximated by using a finite number of ‘points of support.’ The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J ‘classes’ with a model which allows for heterogeneity as follows: The probability of observing y_{it} given that regime j applies is

$$P(i,t|j) \;=\; \text{Prob}[Y_{it} = y_{it}|\lambda_{it},j]$$

where the mean $\lambda_{it}|j$ is specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$\lambda_{it}|j \;=\; \exp[\,\boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \delta_j].$$

We formulate this approximation more generally as,

$$\lambda_{it}|j \;=\; \exp[\,\boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \boldsymbol{\delta}_j'\mathbf{x}_{it}].$$

In this formulation, each group has its own parameter vector, $\boldsymbol{\beta}'_j = \boldsymbol{\beta} + \boldsymbol{\delta}_j$, though the variables that enter the mean are assumed to be the same. (This could be changed by imposing restrictions on the full parameter vector, as described below.) This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters. We denote the mass, or probability of membership in class j as F_j, $j = 1,...,J$, such that $F_1 + F_2 + ... + F_J = 1$. Then, the posterior probability of an observed sequence of observations is

$$P(i) \;=\; \sum\nolimits_{j=1}^{J} F_j P(i|j)$$

where $$F_j \;=\; \frac{\exp(\theta_j)}{\sum\nolimits_{m=1}^{J}\exp(\theta_m)},\; \theta_J \;=\; 0,\; \Sigma_j\, F_j = 1.$$

The model is fit by maximizing the likelihood for the observed data with respect to all parameters including θ_j, $j = 1,...,J$.

E44.3 Commands for Panel Data Models

The command structure for the panel data models is built by adding specifications to the common command. The panel data set is declared first.

SETPANEL ; Group = id variable ; Pds = group count variable $

Then, **POISSON ; Lhs = dependent variable ; Rhs = regressors**
or NEGBIN ; ... any model specific specifications ...
; Panel $

As always, panels may be unbalanced. The **; Panel** may be replaced with**; Pds** specification which gives either the fixed number of periods or the variable which gives the group count. The zero inflation models must be of the form

; ZIP [; Rh2 = list of variables in zero inflation probability]

Only the logit splitting form is supported for this model. If you omit the **; Rh2** list, then this probability will be a constant. Otherwise, it will be of the form

$$\text{Prob[regime 0]} = \frac{\exp(\delta'\mathbf{z}_{it})}{1+\exp(\delta'\mathbf{z}_{it})}$$

(The (τ) form is not supported.) The variables in $\mathbf{z}_{it}$ may vary across time, or, for the random parameters model and latent class model, may be the same in every period.

NOTE: The panel data estimators automatically bypass missing values, and keep all valid observations in a group. Thus, you should not use **SKIP** or **REJECT** to bypass missing values with these estimators. An important implication of this is that in the actual data set used to fit the model, the actual group sizes may be smaller than specified by the **SETPANEL** command or the**; Pds** variable.

Other options for these models are the same as in other settings, including

; Start = list	to give starting values
; Keep = name	to retain fitted values
; Res = name	to retain residuals
; Prob = name	to retain fitted probabilities for observed outcome
; List	to display predicted values
; Partial Effects	
; Rst = list	to specify fixed value and equality restrictions
; CML: spec	to define a constrained maximum likelihood estimator

and the various options for output and control during the iterations.

E44.4 Fixed Effects Models

The fixed effects Poisson and negative binomial models may be estimated two ways. The conditional estimator is the one presented in Hausman, Hall, and Griliches (1984). The unconditional estimator is computed by maximizing the log likelihood directly for all parameters, including the dummy variable coefficients, as described in Chapter R23. We consider them in turn. Only the second method is available for the zero inflation models.

E44.4.1 Conditional Estimation of Poisson and Negative Binomial Models

The conditional estimators for the Poisson and negative binomial models are based on the conditional log likelihood,

$$\log L_c = \sum_{i=1}^{n} \log P\left(y_{i,1}, y_{i,2}, \ldots, y_{i,T_i} \mid \sum_{t=1}^{T_i} y_{i,t}\right).$$

For the negative binomial, this is a different log likelihood and produces different results from the unconditional estimator given in the next section. For the Poisson model, it turns out to be algebraically and numerically identical. The command structure for the conditional fixed effects estimator is

```
POISSON      ; Lhs = y
or NEGBIN    ; Rhs = independent variables
             ; Panel $
```

That is, these are the default panel data estimators for these models. Other options such as marginal effects, fitted values, controls on output, starting values, constraints, and so on are all available. The model in this framework has

$$E[y_{it} \mid \mathbf{x}_{it}] = \exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it}) = \lambda_{it}$$

so the marginal effects would be

$$\partial E[y_{it} \mid \mathbf{x}_{it}]/\partial \mathbf{x}_{it} = \lambda_{it}\boldsymbol{\beta}.$$

In order to compute this quantity, it would be necessary to have an estimate of α_i in hand. But, the estimator is conditioned on the sum, so α_i is conditioned out of the log likelihood, and not estimated. In order to provide information about scaling in the model, we compute marginal effects by using $\bar{y}$ as an estimate of the scale factor at the means. This is an approximation that will estimate the marginal effects reasonably well. The standard errors are questionable, as the variance of the mean is ignored – as a consequence, the t ratios for the marginal effects will be the same as for the corresponding coefficients.

NOTE: The fixed effects model for the Poisson distribution does not allow an overall constant. Surprisingly, an overall constant term is identified in the conditional distribution of the negative binomial model. Your Rhs list may include one in the negative binomial model (but not the Poisson model.) Allison (2001) shows that the reason this occurs is that HHG did not formulate a true fixed effects model in the mean of the random variable. Their formulation layers the fixed effect into the heterogeneity model, not the conditional mean. This is then conditioned out of the distribution to produce the model that is estimated. Thus, in the HHG formulation for the negative binomial model, we do not have $\log\lambda_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it}$. The conditional mean is still $\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it})$, however. The unconditional estimator below (also advocated by Allison) produces this formulation for $\log\lambda_{it}$. A fuller discussion of the two different treatments appears in the technical details in Section E44.4.5.

E44.4.2 Unconditional Estimation of Count Data Models

The unconditional estimator for the Poisson model is obtained by direct maximization of the log likelihood

$$\log L = \sum_{i=1}^{n} \log\left[\prod_{t=1}^{T_i} \frac{\exp(-\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it}))[\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it})]^{y_{it}}}{y_{it}!}\right]$$

with respect to all K+N parameters, where N may be up to 100,000. (Details on the mechanics of estimating 100,000+K parameters are given below and in Chapter R23.) This estimator is also available for the negative binomial model, which has a similar log likelihood with the Poisson density replaced by the negative binomial counterpart;

$$p(y_{it}|\mathbf{x}_{it}) = \text{Prob}[Y = y_{it}] = \frac{\Gamma(\theta + y_{it})}{\Gamma(\theta)\Gamma(y_{it}+1)} u_{it}^{\theta}(1-u_{it})^{y_{it}}, \; u_{it} = \theta / (\theta + \lambda_{it}).$$

NOTE: Full estimation of the fixed effects negative binomial model in this fashion generally encounters the 'incidental parameters' problem. This does not affect the Poisson model, however. The incidental parameters problem is discussed in detail in Chapter R23. The specific relationship to the Poisson model is discussed in Section E44.4.5 below.

This estimator is obtained by adding

; FEM

to the **POISSON** or **NEGBIN** command. (The default estimator is the conditional one. The optional estimator is the unconditional one.)

The unconditional estimator allows for truncation (not censoring) at zero. The model specification is

; TPM

with no other specifications. This is for the conditional distribution $y_i | y_i > 0$, as appears in hurdle models.

You may also estimate the zero inflation models with fixed effects. The full specification of the zero inflation model with this modification is

$$Y_{it} = 0 \text{ with probability } q_{it}$$

$$Y_{it} \sim \text{Poisson}(\lambda_{it}) \text{ or negative binomial with probability } 1 - q_{it}$$

so that

$$\text{Prob}[Y_{it} = 0] = q_{it} + [1 - q_{it}]R_{it}(0)$$

$$\text{Prob}[Y_{it} = j > 0] = [1 - q_{it}]R_{it}(j)$$

where

$$R_{it}(y) = \text{the Poisson probability} = e^{-\lambda it}\lambda_{it}^{yit} / y_{it}!$$

and

$$\lambda_{it} = e^{\alpha i + \boldsymbol{\beta}'\mathbf{x}it}$$

or the negative binomial probability with overdispersion parameter $\alpha = 1/\theta$,

$$R_{it}(y_{it}) = \Gamma(\theta + y_{it})/[y_{it}!\Gamma(\theta)]\, u_{it}^{\theta}\, [1 - u_{it}]^{yit},\ u_i = \theta / [\theta + \lambda_{it}].$$

The state probability, q_i has

$$\text{Prob}[q_{it} = 1] = \Lambda(\boldsymbol{\gamma}'\mathbf{z}_{it}).$$

You must provide a set of starting values for this model (unlike the other two). Do this simply by fitting the model without fixed effects before fitting the model with the fixed effects. The command structure would be as follows:

```
POISSON      ; Lhs   = dependent variable
or NEGBIN    ; Rhs   = independent variables
             ; Zip   ; Rh2 = variables in z (optional) $
POISSON      ; Lhs   = dependent variable
or NEGBIN    ; Rhs   = independent variables
             ; ZIP   ; Rh2 = variables in z (optional)
             ; FEM   ; Pds = panel specification $
```

Note the Rh2 list is optional. If you do not include it, then the regime model will contain only a constant term; i.e., q_{it} will be a constant.

NOTE: In specifying the ZIP model, include *one* in both the Rhs and Rh2 lists in both model commands, even if the splitting probability is constant. That is, in the first command above, you should have **; Rh2 = one**. If you do not have this, then the first model fit will be the ZIP(τ) or ZINB(τ), which will provide inappropriate starting values for the second model. The starting values are very important for this model.

NOTE: In the zero inflation model, the individual effect enters the mean of the probability model, not the splitting probability.

In this setting, a few of the optional estimation features are restricted. The options generally available are

; Keep = name	to retain the fitted value, λ_{it}
; Res = name	to retain residuals
; Prob = name	to retain p_{it}
; Cprob = name	to retain the group probability, $\Pi_t\, p_{it}$
; List	to produce a list of actual and predicted outcomes and probabilities
; Covariance Matrix	to display the covariance matrix for the slopes only, same as **; Printvc**
; Partial Effects	to produce the marginal effects computed at the data means

The restrictions specifications, **; Rst** and **; CML:** are unavailable, but you may use

; Test: spec	to specify a Wald test based on the coefficient vector not including α_i
; Start = list	to give starting values for $\boldsymbol{\beta}$(and θ for **NEGBIN**).

You may also provide one common value for the α_is, but not a full set, regardless of N, and **; Maxit = value**, for example **; Maxit = 0** to carry out LM tests. Estimation is only by Newton's method, so **; Alg = method** is not available. But, you may set the convergence rules as usual.

NOTE: Though the fixed effects estimators are computed, the asymptotic covariance matrix is not. As such, the only hypotheses related to the fixed effects which may be tested will rely on the likelihood function, not the individual coefficients.

E44.4.3 Two Way Unconditional Fixed Effects Estimator

The unconditional estimator can also produce a two way fixed effects model,

$$E[y_{it}\,|\mathbf{x}_{it}] = \exp(\alpha_i + \delta_t + \boldsymbol{\beta}'\mathbf{x}_{it}) = \lambda_{it}$$

There will now be MaxT_i-1 additional coefficients in the model. You can request this estimator by adding

; Time = ti

where the variable *ti* tells, for each observation, in which period the observation occurred. This variable must take the values 1,2,...,MaxT_i. That is, it must be coded with 't,' the index number of the period. A date will not work – it will be flagged as identifying too many coefficients. Observations may be made at different periods in the different groups. For example, if you have a panel with three observations in the first group and seven in the second, the first three observations could have been made at t = 2, t = 4, and t = 7. The program assumes that MaxT_i is equal to the largest group size in the model. (That way, it is assured that there are no holes in the sequence of observations.) Thus, the largest group in the sample must have this variable coded with the complete set of integers, 1,2,...,*Tmax*.

NOTE: If you have a balanced panel with **; Pds = T** where *T* is a fixed value, then you can specify the time effects with **; Time = one** as there can be no variation in the coding of the period in a balanced panel.

NOTE: Our experience has been that the fixed effects model produces considerable instability in the negative binomial, though it works nicely in the Poisson model. The reason may be that as in the normal heterogeneity case, there is heterogeneity already embodied in the model.

The fixed effects model with time effects is estimated by actually creating the time specific dummy variables. You will see a complete set of time effects in the output. As such, however, if you have a large group size in your panel, this extension may create an extremely large model.

E44.4.4 Applications

To illustrate the panel data estimators, we will return to the German health care data used in the preceding chapters. This is an unbalanced panel with 7,293 individuals observed from one to seven times. The following fits a few of the basic fixed effects models. For these data, the ZIP model with fixed effects appears to be badly specified. The unconditional fixed effects estimator for the negative binomial model is also inestimable with these data.

```
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = age,educ,hhninc,newhsat $
CREATE      ; date = year - 1983 $
CREATE      ; If(date = 8)date = 6 $
CREATE      ; If(date = 11)date = 7 $
```

The base case is the Poisson model with no fixed effects.

```
POISSON     ; Lhs = docvis ; Rhs = x,one ; Partial Effects $
```

This is the Poisson conditional fixed effects estimator.

```
POISSON     ; Lhs = docvis ; Rhs = x,one
            ; Panel ; Partial Effects $
```

This is the Poisson unconditional fixed effects estimator. The coefficient estimates are identical, but the marginal effects are computed differently.

```
POISSON     ; Lhs = docvis ; Rhs = x
            ; Panel ; Partial Effects ; FEM $
```

This is the Hausman et al. negative binomial conditional fixed effects estimator. The unconditional fixed effects is computed as well. As expected, since they are different models, the estimates are noticeably different.

```
NEGBIN      ; Lhs = docvis ; Rhs = x
            ; Panel ; Partial Effects $
NEGBIN      ; Lhs = docvis ; Rhs = x,one $
NEGBIN      ; Lhs = docvis ; Rhs = x,one ; Panel ; FEM $
```

These are two different fixed effects estimators. The first is the conditional estimator; the second is the unconditional estimator. The parameter estimates and standard errors are identical. The log likelihood values are not because the conditional and unconditional log likelihoods are different functions.

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                DOCVIS
Log likelihood function    -90999.58348
Restricted log likelihood -108662.13583
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01005***      .00031    32.08  .0000      .00944    .01067
    EDUC|    -.01936***      .00170   -11.41  .0000     -.02269   -.01603
  HHNINC|    -.27193***      .02150   -12.65  .0000     -.31407   -.22979
 NEWHSAT|    -.22841***      .00133  -171.90  .0000     -.23102   -.22581
Constant|    2.39944***      .02640    90.90  .0000     2.34771   2.45118
--------+--------------------------------------------------------------------
(Conditional Poisson FE identical to unconditional Poisson FE)
-----------------------------------------------------------------------------
Panel Model with Group Effects
Dependent variable                DOCVIS
Log likelihood function    -45515.17412
Unbalanced panel has   7293 individuals
Missing or sumY=0, Skipped  1153 groups
Poisson      Regression - Fixed Effects
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .02230***      .00143    15.55  .0000      .01949    .02511
    EDUC|    -.04858***      .01732    -2.81  .0050     -.08252   -.01464
  HHNINC|    -.18627***      .04159    -4.48  .0000     -.26778   -.10476
 NEWHSAT|    -.14569***      .00222   -65.63  .0000     -.15004   -.14134
--------+--------------------------------------------------------------------
(Negative binomial conditional fixed effects)
-----------------------------------------------------------------------------
Panel Model with Group Effects
Dependent variable                DOCVIS
Log likelihood function    -33473.82946
Neg.Binomial Regression - Fixed Effects
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01569***      .00086    18.17  .0000      .01400    .01738
    EDUC|     .02196***      .00417     5.27  .0000      .01379    .03012
  HHNINC|     .09409*        .05487     1.71  .0864     -.01346    .20164
 NEWHSAT|    -.13569***      .00333   -40.70  .0000     -.14222   -.12915
--------+--------------------------------------------------------------------
(Negative binomial unconditional fixed effects)
-----------------------------------------------------------------------------
FIXED EFFECTS NegBin Model
Dependent variable                DOCVIS
Log likelihood function    -48797.32676
Skipped 1153 groups with inestimable ai
Negative binomial regression model
```

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Index function for probability					
AGE	.03065***	.00264	11.62	.0000	.02548	.03582
EDUC	-.04438	.02865	-1.55	.1214	-.10053	.01178
HHNINC	-.12178*	.07070	-1.72	.0850	-.26035	.01678
NEWHSAT	-.16121***	.00434	-37.13	.0000	-.16972	-.15270
	Overdispersion parameter					
Alpha	2.16113***	.03520	61.40	.0000	2.09214	2.23012

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial Effects
(Poisson)

DOCVIS	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
AGE	.03200***	.00100	31.90	.0000	.03003	.03397
EDUC	-.06163***	.00541	-11.40	.0000	-.07223	-.05103
HHNINC	-.86570***	.06851	-12.64	.0000	-.99998	-.73142
NEWHSAT	-.72716***	.00490	-148.52	.0000	-.73676	-.71756

(Poisson Conditional FE)

AGE	.07099***	.00457	15.55	.0000	.06204	.07995
EDUC	-.15465***	.05513	-2.81	.0050	-.26271	-.04660
HHNINC	-.59298***	.13239	-4.48	.0000	-.85247	-.33350
NEWHSAT	-.46380***	.00707	-65.63	.0000	-.47765	-.44995

(Poisson unconditional FE)

AGE	.12776***	.97983	4.51	.0000	.07222	.18330
EDUC	-.27831***	-.54699	-5.61	.0000	-.37548	-.18114
HHNINC	-1.06713***	-.06527	-3.19	.0014	-1.72232	-.41193
NEWHSAT	-.83465***	-.97090	-5.05	.0000	-1.15839	-.51091

(Negative binomial conditional fixed effects)

AGE	.04996***	.00275	18.17	.0000	.04457	.05534
EDUC	.06990***	.01327	5.27	.0000	.04390	.09590
HHNINC	.29954*	.17469	1.71	.0864	-.04285	.64194
NEWHSAT	-.43197***	.01061	-40.70	.0000	-.45277	-.41117

(Negative binomial unconditional fixed effects)

AGE	.17304***	1.34672	2.75	.0059	.04989	.29620
EDUC	-.25054***	-.49968	-2.83	.0047	-.42420	-.07687
HHNINC	-.68754	-.04267	-1.43	.1530	-1.63045	.25537
NEWHSAT	-.91015***	-1.07438	-3.03	.0024	-1.49836	-.32194

E44.4.5 Technical Details for Fixed Effects Models

For the conditional approaches, the fixed effects models are transformed to an estimable form by obtaining the conditional density, $p(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it})$. This removes the fixed effect from the resulting distribution. (Derivations may be found in Hausman, Hall, and Griliches (1984).) For the Poisson distribution,

$$p(y_{i1},y_{i2},...y_{iT}|\Sigma_t y_{it}) = \left[\frac{\left(\sum_t y_{it}\right)!}{\prod_t y_{it}!}\right]\prod_t p_{it}{}^{y_{it}}$$

where $$p_{it} = \frac{\lambda_{it}}{\sum_t \lambda_{it}} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{it})}{\sum_{t=1}^{T_i} \exp(\boldsymbol{\beta}'\mathbf{x}_{it})}.$$

Note that $p_{it} = \lambda_{it}/\Sigma_t\lambda_{it}$, where $\lambda_{it} = e^{\alpha_i}e^{\boldsymbol{\beta}'\mathbf{x}_{it}}$, but the fixed effects fall out of the result. The contribution to the log likelihood, gradient and Hessian for the ith group is

$$\log L_i = \log p\big(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it}\big)$$

$$\partial \log p\big(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it}\big)/\partial\boldsymbol{\beta} = \Sigma_t y_{it}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$$

$$\bar{\mathbf{x}}_i = \Sigma_t p_{it}\mathbf{x}_{it}.$$

$$\partial^2 \log p\big(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it}\big)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -(\Sigma_t y_{it})\,[\Sigma_t p_{it}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'].$$

The negative inverse of the Hessian is used to estimate the asymptotic covariance of the estimator for the Poisson model. Though it might not be obvious from the preceding, this result is algebraically identical to the solution that is obtained by using the unconditional, brute force approach described below.

The contribution of the ith group to the conditional log likelihood function in the Poisson fixed effects model is of the form

$$\log L_i = \sum_{t=1}^{T_i} y_{it} \log p_{it}.$$

Suppose for the moment that β were known. The likelihood equation for the ith fixed effect coefficient provides the implicit solution

$$\alpha_i = \log\left[\left(\sum_{t=1}^{T_i} y_{it}\right)/\left(\sum_{t=1}^{T_i}\exp(\boldsymbol{\beta}'\mathbf{x}_{it})\right)\right].$$

If the dependent variable takes value zero for every observation in group i, then the sum equals zero, and group i does not contribute to the log likelihood. Such observation groups are dropped from the analysis. The results from conditional estimation of the Poisson model will contain a statement such as the following which appears in our application:

```
Skipped 1153 groups with inestimable ai
```

Note that unlike the binary choice models, it is not necessary for the y_{it} to vary within the group, so long as it is nonzero. In essence, the estimator of α_i is based on log $\overline{y}_i$, so any nonzero value will suffice, whether or not there is within group variation. The same consideration arises in the unconditional estimator described below, in this case for the negative binomial model as well. (This would be expected, since for the Poisson model, the conditional and unconditional estimators are expected.)

For the negative binomial model, the treatment is quite different in the conditional and unconditional formulations. Winkelmann (2008) provides a useful summary: We begin with the NB2 assumption,

$$\text{Prob}(Y = y_{it}) = \frac{\Gamma(\theta + y_{it})}{\Gamma(\theta)\Gamma(y_{it}+1)}\left(\frac{\theta}{\theta+\lambda_{it}}\right)^{\theta}\left(\frac{\lambda_{it}}{\theta+\lambda_{it}}\right)^{y_{it}},$$

where $\Gamma(.)$ is the gamma function. In order to obtain the contagion result needed to derive the distribution of the sum of negative binomials, it is necessary to assume that the NB1 form applies to the individual observation. The NB1 form can obtained by replacing θ with $\theta\lambda_{it}$ in each appearance in the NB2 form. To avoid a step, we suppose as well that the overdispersion parameter plays the role of the fixed effect, θ_i. We now also absorb the effect in λ_{it}, and write $\phi_{it} = \theta_i\lambda_{it}$. The density for NB1 becomes

$$\text{Prob}(Y = y_{it}) = \frac{\Gamma(\phi_{it} + y_{it})}{\Gamma(\phi_{it})\Gamma(y_{it}+1)}\left(\frac{\theta}{\theta+1}\right)^{\phi_{it}}\left(\frac{1}{\theta+1}\right)^{y_{it}}.$$

In this form, $E[y_{it}] = \theta_i\lambda_{it} = \phi_{it}$

and $\text{Var}[y_{it}] = \phi_{it}(1 + \theta_i)$.

This is a fixed effects model of a sort, since

$$E[y_i] = \exp(\beta'x_{it} + \log\theta_i) = \exp(\beta'x_{it} + \alpha_i)$$

but note that θ_i(or α_i) is playing more than just the role of a fixed effect here. It is also changing the variance. It cannot be interpreted the way that we are accustomed to interpreting fixed effects. We do note, this makes clear the source of a frequently observed peculiarity of the model. No problem is caused by the presence of an overall constant, or, indeed, other time invariant variables in $\mathbf{x}_{it}$. Notwithstanding this ambiguity of the model, this is the formulation that was devised by Hausman, Hall and Griliches (1984) and that has been widely used since then. The conditional distribution that corresponds to the NB1 model with that type of fixed effect is

$$p\big(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it}\big) = \prod_t\left(\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})y_{it}!}\right)\frac{\Gamma(\Sigma_t\lambda_{it})[(\Sigma_t y_{it}!)]}{\Gamma(\Sigma_t\lambda_{it} + \Sigma_t y_{it}]}.$$

The log likelihood is, again,

$$\text{Log } L = \Sigma_i \log p\big(y_{i1},y_{i2},...,y_{i,Ti}|\Sigma_t y_{it}\big).$$

The gradient for the *i*th observation group is

$$\partial\, p\big(y_{i1},y_{i2},\ldots,y_{i,Ti}|\Sigma_t y_{it}\big)/\partial\boldsymbol{\beta} \;=\; \Sigma_t\lambda_{it}e_{it}\mathbf{x}_{it},$$

$$e_{it} \;=\; [\Psi(\lambda_{it}+y_{it}) - \Psi(\lambda_{it})] - [\Psi(\Sigma_t\,\lambda_{it}+\Sigma_t\,y_{it}) - \Psi(\Sigma_t\,\lambda_{it})],$$

$$\Psi(t) \;=\; \Gamma'(t)/\Gamma(t) \;=\; \text{the digamma function.}$$

The asymptotic covariance matrix for the negative binomial estimator is computed with the BHHH estimator. We do note an aspect of the negative binomial model. The Hausman et al. conditional estimator is numerically quite stable, in spite of its questionable theoretical pedigree. (The model is overspecified – it essentially is a Poisson model with two heterogeneity effects and in spite of this, it is not really an ‘effects’ model.) In contrast, the true fixed effects model fit with the unconditional estimator, while more theoretically orthodox is, in our experience, quite numerically unstable. We have frequently observed serious numerical problems such as overflows.

The unconditional log likelihood for both models is maximized by using Newton’s method. A full discussion of the method is given in Chapter R23; for convenience, only a short sketch of the result is given here, for the Poisson model. (The results for the negative binomial (NB2) model are similar.) The log likelihood is

$$\log L \;=\; \sum_{i=1}^{n}\;\log\left[\prod_{t=1}^{T_i}\;\frac{\exp(-\exp(\alpha_i+\boldsymbol{\beta}'\mathbf{x}_{it}))[\exp(\alpha_i+\boldsymbol{\beta}'\mathbf{x}_{it})]^{y_{it}}}{y_{it}\,!}\right]$$

Let p_{it}, y_{it}, $\mathbf{x}_{it}$ and λ_{it} denote the obvious components of this function. Then,

$$\frac{\partial\log L}{\partial\boldsymbol{\beta}} \;=\; \sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it}-\lambda_{it})\mathbf{x}_{it} \;=\; \mathbf{g}_{\beta}$$

$$\frac{\partial\log L}{\partial\alpha_\iota} \;=\; \sum_{t=1}^{T_i}(y_{it}-\lambda_{it}) \;=\; g_i$$

$$\frac{\partial^2\log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \;=\; -\sum_{i=1}^{N}\sum_{t=1}^{T_i}\lambda_{it}\mathbf{x}_{it}\mathbf{x}_{it}' \;=\; \mathbf{H}_{\beta\beta'}$$

$$\frac{\partial^2\log L}{\partial\alpha_i^{\,2}} \;=\; -\sum_{t=1}^{T_i}\lambda_{it} \;=\; h_{ii}$$

$$\frac{\partial^2\log L}{\partial\boldsymbol{\beta}\partial\alpha_i} \;=\; -\sum_{t=1}^{T_i}\lambda_{it}\mathbf{x}_{it} \;=\; \mathbf{h}_{\beta i}$$

The results for the Poisson model (not the negative binomial model) are identical to the conditional estimator. This has an important implication: Unlike most other models, the incidental parameters issue does not apply to the Poisson model. The unconditional fixed effects estimator is consistent.

E44.5 Random Effects Models

The random effects model for the Poisson framework is

$$\begin{aligned}\log\lambda_{it}^* &= \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_i,\ i = 1,...,N,\ t = 1,...,T_i,\\ &= \log[\lambda_{it}\exp(\varepsilon_i)]\end{aligned}$$

where ε_i is a random effect for the *i*th group, the same in every period, such that $\exp(\varepsilon_i)$ has a gamma distribution with parameters(θ,θ). Thus, $E[\exp(\varepsilon_i)]$ has mean 1 and variance $1/\theta = \alpha$. This is the framework which gave rise to the negative binomial model earlier, so that, with the minor modifications, this is the estimating framework for the Poisson model with random effects. For the negative binomial model, Hausman, et al. proposed the following approach: We begin with the Poisson model with the random effects specification shown above. The random term, ε_i is distributed as gamma with parameters(θ_i,θ_i), which produces the negative binomial model with a parameter that varies across groups. Then, it is assumed that $\theta_i/(1+\theta_i)$ is distributed as beta(a_n,b_n), which layers the random group effect onto the negative binomial model. Details on the resulting distribution are given below. In sum, then, the random effect is added to the negative binomial model by assuming that the overdispersion parameter is randomly distributed across groups. Use

; Random

to request the random effects model.

There is another useful interpretation of the random effects model. Rewrite the model as

$$\log\lambda_{it}^* = \alpha_i + \boldsymbol{\beta}_1'\mathbf{x}_{it},\ i = 1,...,N,\ t = 1,...,T_i,$$

where $$\alpha_i = \alpha + \varepsilon_i$$

This is a trivial modification of the essential structure of the model. However, as written, we can reinterpret the model as an ordinary count model with a random constant term. This suggests an alternative approach to estimation. The random parameters models discussed in Section R24.3 gives further details, including this special case. The two random effects models discussed above may be modified to use the normal distribution for the random effect instead of the gamma, $\varepsilon_i \sim N[0,\sigma^2]$. For the Poisson model, this is an alternative to the log-gamma model which gives rise to the negative binomial. The negative binomial model is much more involved than this, and the normal model is a considerably simpler alternative. To request the random effects models with normally distributed heterogeneity, use

```
POISSON      ; Lhs = ... ; Rhs = ... ; Pds = ...
or NEGBIN    ; Random Effects
             ; Normal distribution $
```

The parameters estimated by these models are as follows:

	Log-gamma	Lognormal
Poisson	$\boldsymbol{\beta},\alpha$	$\boldsymbol{\beta},\sigma^2$
Negative Binomial	$\boldsymbol{\beta},a_n,b_n$	$\boldsymbol{\beta},\alpha,\sigma^2$

NOTE: The negative binomial model might be somewhat overparameterized by this extension. The random effect essentially adds a heterogeneity term to a model that is obtained by adding a heterogeneity term to a lower level (the Poisson) model. As such, it will be common that attempts to fit the negative binomial model with random effects will be unsuccessful.

You may use

	; Start = list	to give starting values
and	**; Rst = list**	to impose restrictions

for any of the four models. The default algorithm is BFGS, which you should use unless there is some definite reason to use some other. Other controls, such as **; Maxit [= 0]** are available as usual.

The random effects models are computed using the Butler and Moffitt method, with Gauss-Hermite integration. They can also be fit as random parameters, i.e., random constant models, using the RP estimator described in Section E44.6.

The panel data models produce a full set of results for the base model before estimation of the random effects model. Thus, the Poisson models produce three sets of estimates while the negative binomial model will produce four sets of results:

- OLS results (if requested with **; OLS**)
- Poisson regression, ignoring the group effects
- negative binomial model ignoring the group effects
- negative binomial model including the fixed or random effects.

The retrievable results are

Matrices: *b* and *varb*

Scalars: *kreg*, *nreg*, *logl*, *s* (when σ is estimated)

Last Model: *b_variable*, *a* (if you fit a negative binomial model)

E44.5.1 Application

To illustrate the random effects estimators, we will reestimate two of the models fit earlier with fixed effects. Several variants of the random effects models are suggested. We first fit the Poisson model with no effects, then with log gamma then normally distributed random effects. The random parameters model is an alternative estimator for the model with normally distributed effects fit by the Butler and Moffitt method. The three random effects Poisson models can also be fig using the negative binomial specification.

```
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = age,educ,hhninc,newhsat $
POISSON     ; Lhs = docvis ; Rhs = x,one ; Panel ; Partial Effects ; Random Effects $
POISSON     ; Lhs = docvis ; Rhs = x,one ; Panel ; Partial Effects ; Random Effects
            ; Normal $
? These are alternative random effects specifications.
NEGBIN      ; Lhs = docvis ; Rhs = x,one ; Pds = ni ; Partial Effects ; Random Effects $
NEGBIN      ; Lhs = docvis ; Rhs = x,one ; Pds = ni ; Partial Effects ; Random Effects
            ; Normal $
```

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable              DOCVIS
Log likelihood function    -90999.58348
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01005***      .00031    32.08  .0000      .00944    .01067
    EDUC|    -.01936***      .00170   -11.41  .0000     -.02269   -.01603
  HHNINC|    -.27193***      .02150   -12.65  .0000     -.31407   -.22979
 NEWHSAT|    -.22841***      .00133  -171.90  .0000     -.23102   -.22581
Constant|    2.39944***      .02640    90.90  .0000     2.34771   2.45118
--------+--------------------------------------------------------------------
Panel Model with Group Effects
Log likelihood function    -68895.20568
Unbalanced panel has   7293 individuals
Poisson      Regression - Random Effects
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01432***      .00047    30.66  .0000      .01340    .01523
    EDUC|    -.02658***      .00424    -6.27  .0000     -.03489   -.01828
  HHNINC|    -.14117***      .01668    -8.46  .0000     -.17386   -.10848
 NEWHSAT|    -.15979***      .00072  -220.90  .0000     -.16121   -.15838
Constant|    1.82709***      .05312    34.39  .0000     1.72297   1.93121
   Alpha|     .92313***      .01640    56.27  .0000      .89098    .95528
--------+--------------------------------------------------------------------
Panel Model with Group Effects
Log likelihood function    -69020.27550
Poisson      Regression - Random Effects
Normally distributed random effect
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AGE|     .01613***      .00044    36.59  .0000      .01527    .01700
    EDUC|    -.02613***      .00389    -6.73  .0000     -.03375   -.01852
  HHNINC|    -.16793***      .01665   -10.08  .0000     -.20057   -.13529
 NEWHSAT|    -.16041***      .00071  -225.23  .0000     -.16181   -.15902
Constant|    1.32711***      .04795    27.68  .0000     1.23312   1.42109
   Sigma|     .97697***      .00699   139.72  .0000      .96327    .99068
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1835
Scale Factor for Marginal Effects  3.1835
```

DOCVIS	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
AGE	.04558***	.00149	30.66	.0000	.04266	.04849
EDUC	-.08463***	.01349	-6.27	.0000	-.11108	-.05818
HHNINC	-.44941***	.05310	-8.46	.0000	-.55348	-.34535
NEWHSAT	-.50871***	.00230	-220.90	.0000	-.51322	-.50419
AGE	.05136***	.00140	36.59	.0000	.04861	.05411
EDUC	-.08320***	.01237	-6.73	.0000	-.10744	-.05896
HHNINC	-.53461***	.05302	-10.08	.0000	-.63852	-.43069
NEWHSAT	-.51068***	.00227	-225.23	.0000	-.51512	-.50623

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

E44.5.2 Technical Details for Random Effects Models

The random effects models are obtained by integrating $\exp(\varepsilon_i)$ out of

$$p(y_{i1},y_{i2},\ldots,y_{iT},\exp(\varepsilon_i)) \;=\; p(y_{i1},y_{i2},\ldots,y_{iT}|\exp(\varepsilon_i))g(\exp(\varepsilon_i)).$$

For the Poisson model, the random effect is assumed to enter multiplicatively through λ_{it}, the same as in the earlier derivation of the negative binomial model. Conditioned on the heterogeneity, the T_i observations $y_{it}|\exp(\varepsilon_i)$ are distributed as independent Poisson variates each with parameter $\exp(\varepsilon_i)\lambda_{it}$. Then,

$$p(y_{i1},y_{i2},\ldots,y_{iT}|\exp(\varepsilon_i)) \;=\; \Pi_t p(y_{it}|\exp(\varepsilon_i)).$$

We assume that $g(\exp(\varepsilon_i))$ is the gamma distribution with parameters (θ,θ) with $\theta = 1/\alpha$, so that $E[\exp(\varepsilon_i)]$ equals 1. The density that results when u_i is integrated out is

$$p(y_{i1},y_{i2},\ldots,y_{iT},) \;=\; \frac{\left(\prod_t \lambda_{it}^{\,y_{it}}\right)\Gamma\left(\theta+\sum_t y_{it}\right)}{\left(\prod_t y_{it}!\right)\Gamma(\theta)\left(\sum_t y_{it}\right)!\left(\sum_t \lambda_{it}\right)^{\sum_t y_{it}}}u_i^{\theta}(1-u_i)^{\sum_t y_{it}}$$

where $u_i = \theta / (\theta + \Sigma_t y_{it})$. As usual, $\log L = \Sigma_i \log p(\ldots)$. The gradient for the ith term is

$$\partial \log L_i/\partial \boldsymbol{\beta} \;=\; \Sigma_t w_{it}\mathbf{x}_{it}$$

where $$w_{it} \;=\; \lambda_{it}u_i(A_i - 1) \;+\; \lambda_{it}(y_{it} - A_i)$$

$$A_i \;=\; \Sigma_t y_{it} / \Sigma_t \lambda_{it}$$

and $$\partial \log L_i/\partial \theta \;=\; \Psi(\theta + \Sigma_t y_{it}) - \Psi(\theta) + \log u_i + (1 - u_i) - (u_i/\theta)\Sigma_t y_{it}.$$

Construction of the density for the random effects negative binomial model is described above. We first build the heterogeneity in to the distribution of ε_i by letting θ_i carry the random effect. (Note this is similar to the handling of the fixed effects negative binomial earlier.) It is assumed that $\theta_i/(1+\theta_i)$ is distributed as beta(a_n,b_n), which layers the random group effect onto the negative binomial model. The resulting model is

$$p(y_{i1},y_{i2},\ldots,y_{iT},) = \frac{\Gamma(a_n+b_n)\Gamma\left(a_n+\sum_{t=1}^{T_i}\lambda_{it}\right)\Gamma\left(b_n+\sum_{t=1}^{T_i}y_{it}\right)}{\Gamma(a_n)\Gamma(b_n)\Gamma\left(a_n+b_n+\sum_{t=1}^{T_i}\lambda_{it}+\sum_{t=1}^{T_i}y_{it}\right)}\prod_{t=1}^{T_i}\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}$$

The derivatives are

$$\begin{aligned}
\partial\log L_i/\partial\boldsymbol{\beta} &= \Sigma_t\lambda_{it}e_{it}\mathbf{x}_{it}\\
\partial\log L_i/\partial a_n &= \Sigma_t\lambda_{it}e_{it} - \Psi(a_n) + \Psi(a_n + \Sigma_t\lambda_{it})\\
\partial\log L_i/\partial b_n &= \Sigma_t\lambda_{it}e_{it} - \Psi(b_n) + \Psi(b_n + \Sigma_t y_{it})\\
e_{it} &= \Psi(a_n + \Sigma_t\lambda_{it}) - \Psi(a_n + b_n + \Sigma_t\lambda_{it} + \Sigma_t y_{it}) + \Psi(\lambda_{it} + y_{it}) - \Psi(\lambda_{it})\\
\Psi(z) &= d\log\Gamma(z)/dz \ \text{ (the `digamma' function).}
\end{aligned}$$

For the random effects model with normally distributed group effects, we form the likelihood function in the same fashion as in Section E44.5.2 for the cross section case – the derivation is identical with a small change in the notation. For group *i*, conditioned on ε_i, the T_i observations are independent. The unconditional density for the observed data is formed by integrating ε_i out of the joint density. Thus,

$$p(y_1,y_2,\ldots,y_{T_i}) = \int_{-\infty}^{\infty}\prod_{t=1}^{T_i}\frac{\exp(-\lambda_{it}\,|\,\varepsilon_i)(\lambda_{it}\,|\,\varepsilon_i)^{y_{it}}}{y_{it}!}\left(\frac{1}{\sigma}\right)\phi\left(\frac{\varepsilon_i}{\sigma}\right)d\varepsilon_i$$

where $\lambda_{it}|\varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_i)$ is the mean of y_{it} conditioned on the group effect. The log likelihood, its derivatives with respect to $\boldsymbol{\beta}$ and σ, and the estimate of the Hessian are computed as discussed in Section E44.5.2. In all cases, the estimator of the covariance matrix for the estimated coefficients is the BHHH estimator obtained by summing the outer products of the gradients for the *N* observations.

As before, there are two ways to handle the normally distributed effect, with quadrature using the Butler and Moffitt method, or as a random constant model, using maximum simulated likelihood, as considered in the next section.

E44.6 Random Parameters Models

The random parameters model is described in detail in Chapter R24. The general form of the model for the Poisson and negative binomial regressions is

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{it}$$

where $\boldsymbol{\beta}$ = the fixed means of the distributions for the random parameters.

$\mathbf{z}_i$ = a set of M observed variables which do not vary over time and which enter the means (optional).

$\boldsymbol{\Delta}$ = coefficient matrix, $K \times M$, which forms the observation specific term in the mean.

$\mathbf{v}_{it}$ = unobservable $K\times1$ latent random term in the ith observation in $\boldsymbol{\beta}_i$. Each element of $\mathbf{v}_{it}$ has zero mean and variance one. Each element of $\mathbf{v}_{it}$ may be distributed as normal, uniform, or triangular. They need not be the same.

$$\lambda|\mathbf{v}_{it} = \exp(\boldsymbol{\beta}_i'\mathbf{x}_{it})$$

$$P(y_i|\mathbf{x}_{it},\mathbf{v}_{it}) = \text{Poisson or negative binomial probability given } \lambda_i.$$

Several extensions and narrower details for the model are given in Section R24.3. The Poisson and negative binomial models are standard applications of the results given there.

The command for the random parameters model is structured as follows:

	POISSON	**; Lhs**	**= dependent variable**
	or NEGBIN	**; Rhs**	**= list of all variables in $\mathbf{x}_i$, including one if the model contains a constant**
		; Panel	
		; RPM	**(random parameters model)**
or		**; RPM**	**= list of variables in $\mathbf{z}_i$**
		; Fcn	**= specification of random parameters**
		; Pts	**= r (number of replications)**
		; Cor	**(for correlated parameters) $**

The last two specifications are optional. The remainder are mandatory parts of the command. The **; Fcn** list consists of a list of names of variables which appear in $\mathbf{x}_i$, followed in parentheses by (n) for normally distributed, (u) for uniform, or (t) for triangular. Other options for the Poisson and negative binomial model are specified as usual. These include:

; Par	to keep individual specific parameter estimates.
; Keep = name	to retain fitted values
; Res = name	to retain residuals
; Prob = name	to retain fitted probabilities for observed outcome
; Partial Effects	
; List	to display predicted values (only available if T_i is $\leq$ 10 for all i)
; Maxit= n	to set maximum iterations

and so on. The optional specifications are described in the technical details below.

Here is an example command for the model estimated in the previous section:

```
POISSON      ; Lhs = docvis ; Rhs = x,one ; Pds = ti
             ; RPM ; Fcn = one(n),income(n) ; Correlation ; Pts = 50
             ; Partial Effects $
```

This command specifies two correlated random and three fixed parameters, and 50 replications for the simulations.

The random parameters estimator allows for truncation (not censoring) at zero. The model specification is

```
             ; TPM
```

with no other specifications. This is for the conditional distribution $y_i|y_i > 0$, as appears in hurdle models.

The random parameters model with only a random constant term is equivalent to the random effects model in the previous section. However, the estimates obtained will be different for two reasons. First, the model is estimated by simulation, not by analytical maximum likelihood. Second, the distribution of the random term is assumed to be normal here, whereas in the previous models it is assumed to be log-gamma (though you can specify a normally distributed term as well). A comparison appears below. The random parameters model is also extended to the Poisson and negative binomial ZIP models. This estimator is described at the end of this section.

E44.6.1 Application

In order to replicate results with this estimator, you must either reset the seed for the random number generator to the same value every time, or use Halton sequences as we have below. Otherwise, results produced by identical commands will differ slightly. The command **CALC ; Ran(your value)** placed before each command will remove this source of variation. We have continued the illustrations in the previous sections with a slightly different specification. For this illustration, we have also restricted the sample to those 886 individuals observed in all seven periods of the panel. The commands are as follows:

```
REJECT       ; ti < 7 $
NAMELIST     ; x = one,female,hhninc,educ $
POISSON      ; Lhs  = docvis ; Rhs = x ; Partial Effects $
POISSON      ; Lhs  = docvis ; Rhs = x ; Partial Effects
             ; RPM  ; Fcn = female(n),hhninc(n),educ(n)
             ; Panel ; Pts = 25 ; Halton ; Correlated
             ; Partial Effects $
POISSON      ; Lhs  = docvis ; Rhs = x ; Partial Effects
             ; RPM  ; Fcn = female(n),hhninc(n),educ(n)
             ; Panel ; Pts = 25 ; Halton ; Correlated ; AR1 $
NEGBIN       ; Lhs  = docvis ; Rhs = x ; Partial Effects $
NEGBIN       ; Lhs  = docvis ; Rhs = x ; Partial Effects
             ; RPM  ; Fcn = female(n),hhninc(n),educ(n)
             ; Panel ; Pts = 25 ; Halton; Correlated $
```

```
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable               DOCVIS
Log likelihood function    -23415.68734
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.79197***      .04842    37.01  .0000     1.69706   1.88687
  FEMALE|     .40467***      .01497    27.04  .0000      .37533    .43400
  HHNINC|    -.04210         .04615     -.91  .3616     -.13255    .04834
    EDUC|    -.07738***      .00432   -17.92  .0000     -.08584   -.06892
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1340
Scale Factor for Marginal Effects  3.1340
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  FEMALE|    1.28057***      .04821    26.56  .0000     1.18608   1.37506   #
  HHNINC|    -.13195         .14463     -.91  .3616     -.41542    .15152
    EDUC|    -.24252***      .01364   -17.78  .0000     -.26925   -.21578
--------+--------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------


-----------------------------------------------------------------------------
Random Coefficients  Poisson  Model
Dependent variable               DOCVIS
Log likelihood function    -16498.80774
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
Constant|    1.38470***      .04129    33.53  .0000     1.30377   1.46563
        |Means for random parameters
  FEMALE|     .35720***      .01030    34.69  .0000      .33702    .37738
  HHNINC|     .48114***      .03342    14.40  .0000      .41564    .54664
    EDUC|    -.07369***      .00371   -19.89  .0000     -.08095   -.06643
        |Diagonal elements of Cholesky matrix
  FEMALE|     .26889***      .01080    24.89  .0000      .24771    .29006
  HHNINC|    3.65976***      .03546   103.21  .0000     3.59026   3.72926
    EDUC|     .11090***      .00054   206.14  .0000      .10984    .11195
        |Below diagonal elements of Cholesky matrix
lHHN_FEM|    2.00274***      .03653    54.83  .0000     1.93115   2.07434
lEDU_FEM|     .04662***      .00137    34.06  .0000      .04394    .04930
lEDU_HHN|     .14733***      .00117   126.26  .0000      .14504    .14962
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       .268886
       2|       4.17191
       3|       .190205
Implied correlation matrix of random parameters
Cor_Beta|             1             2             3
--------+------------------------------------------
       1|       1.00000       .480054       .245098
       2|       .480054       1.00000       .797161
       3|       .245098       .797161       1.00000
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   2.4536
Scale Factor for Marginal Effects  2.4536
--------+--------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCVIS|      Effect    Elasticity      z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
  FEMALE|     .87640***      .15101    17.35  .0000      .77738    .97542
  HHNINC|    1.18050***      .16806    14.01  .0000     1.01538   1.34563
    EDUC|    -.18080***     -.80623  -102.64  .0000     -.18425   -.17735
--------+--------------------------------------------------------------------


-----------------------------------------------------------------------------
Random Coefficients  Poisson  Model
Log likelihood function    -16414.99455
First order autocorrelation model
POISSON regression model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
Constant|    2.48699***      .03663    67.90  .0000     2.41520   2.55877
        |Means for random parameters
  FEMALE|     .39451***      .01138    34.67  .0000      .37221    .41681
  HHNINC|    3.01874***      .05227    57.75  .0000     2.91629   3.12119
    EDUC|    -.21165***      .00331   -63.98  .0000     -.21814   -.20517
        |Diagonal elements of Cholesky matrix
  FEMALE|     .12644***      .01253    10.09  .0000      .10187    .15100
  HHNINC|    2.81944***      .06118    46.09  .0000     2.69953   2.93934
    EDUC|     .02389***      .00128    18.70  .0000      .02139    .02639
        |Below diagonal elements of Cholesky matrix
lHHN_FEM|     .11513**       .05861     1.96  .0495      .00025    .23002
lEDU_FEM|     .06450***      .00198    32.50  .0000      .06061    .06839
lEDU_HHN|     .03971***      .00167    23.77  .0000      .03644    .04299
        |First order autocorrelation parameters
ar1FEMAL|    -.05159***      .01442    -3.58  .0003     -.07986   -.02333
ar1HHNIN|    -.19906***      .00869   -22.91  .0000     -.21608   -.18203
 ar1EDUC|     .91174***      .01147    79.49  .0000      .88926    .93423
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Implied standard deviations of random parameters
S.D_Beta|           1
--------+--------------
       1|       .126438
       2|       2.82179
       3|      .0794220
Implied correlation matrix of random parameters
Cor_Beta|           1             2             3
--------+------------------------------------------
       1|       1.00000      .0408021       .812105
       2|      .0408021       1.00000       .532726
       3|       .812105       .532726       1.00000
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   4.0251
Scale Factor for Marginal Effects  4.0251
--------+--------------------------------------------------------------------
        |     Partial                         Prob.      95% Confidence
  DOCVIS|      Effect    Elasticity      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  FEMALE|    1.58793***      .16679    19.48  .0000     1.42816   1.74771
  HHNINC|    12.1507***     1.05443    28.13  .0000     11.3041   12.9972
    EDUC|    -.85192***    -2.31566   -48.63  .0000     -.88625   -.81758
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable               DOCVIS
Log likelihood function    -13644.44645
Restricted log likelihood  -23415.68734
Chi squared [   1 d.f.]     19542.48178
Significance level               .00000
McFadden Pseudo R-squared      .4172946
Estimation based on N =   6209, K =   5
Inf.Cr.AIC  =  27298.9 AIC/N =    4.397
Model estimated: Jul 30, 2011, 20:28:35
NegBin form 2; Psi(i) = theta
Tests of Model Restrictions on Neg.Bin.
Model               Logl ChiSquared[df]
Poisson(b=0)   -24176.44   ******** [**]
Poisson        -23415.69     1521.5 [ 3]
Negative Bin.  -13644.45    19542.5 [ 1]
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    1.87069***      .11081    16.88  .0000     1.65351   2.08788
  FEMALE|     .40514***      .03568    11.35  .0000      .33521    .47508
  HHNINC|    -.04981         .10937     -.46  .6488     -.26417    .16454
    EDUC|    -.08450***      .00997    -8.47  .0000     -.10405   -.06495
        |Dispersion parameter for count data model
   Alpha|    1.92310***      .04172    46.10  .0000     1.84134   2.00486
--------+--------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point   3.1383
Scale Factor for Marginal Effects  3.1383
--------+-----------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
  FEMALE|    1.28294***      .12229    10.49  .0000     1.04325   1.52263   #
  HHNINC|    -.15633         .34307     -.46  .6486     -.82873    .51607
    EDUC|    -.26519***      .03211    -8.26  .0000     -.32812   -.20225
--------+-----------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------


--------------------------------------------------------------------------------
Random Coefficients  NegBnReg Model
Dependent variable               DOCVIS
Log likelihood function    -13049.40963
Negative binomial regression model
Simulation based on  25 Halton draws
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Nonrandom parameters
Constant|    1.33550***      .09784    13.65  .0000     1.14373   1.52727
        |Means for random parameters
  FEMALE|     .50627***      .03212    15.76  .0000      .44333    .56922
  HHNINC|     .37829***      .10215     3.70  .0002      .17809    .57849
    EDUC|    -.08845***      .00877   -10.09  .0000     -.10563   -.07127
        |Diagonal elements of Cholesky matrix
  FEMALE|     .06045*        .03166     1.91  .0562     -.00161    .12251
  HHNINC|     .92641***      .11019     8.41  .0000      .71044   1.14237
    EDUC|     .00337**       .00141     2.40  .0166      .00061    .00613
        |Below diagonal elements of Cholesky matrix
lHHN_FEM|   -1.39265***      .10589   -13.15  .0000    -1.60019  -1.18510
lEDU_FEM|    -.11817***      .00400   -29.55  .0000     -.12601   -.11033
lEDU_HHN|    -.02964***      .00372    -7.98  .0000     -.03692   -.02236
        |Dispersion parameter for NegBin distribution
ScalParm|    1.06614***      .02241    47.56  .0000     1.02220   1.11007
--------+-----------------------------------------------------------------------
Implied standard deviations of random parameters
S.D_Beta|            1
--------+--------------
       1|      .0604485
       2|       1.67263
       3|       .121880
Implied correlation matrix of random parameters
Cor_Beta|            1             2             3
--------+------------------------------------------
       1|       1.00000      -.832608      -.969579
       2|      -.832608       1.00000       .672574
       3|      -.969579       .672574       1.00000
```

```
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   2.0421
Scale Factor for Marginal Effects  2.0421
--------+--------------------------------------------------------------------
        |     Partial                             Prob.      95% Confidence
  DOCVIS|      Effect    Elasticity      z      |z|>Z*          Interval
--------+--------------------------------------------------------------------
  FEMALE|    1.03384***       .21404      7.85  .0000       .77575   1.29193
  HHNINC|     .77249***       .13214      3.79  .0002       .37266   1.17232
    EDUC|    -.18062***     -.96772  -312.54  .0000     -.18175   -.17949
--------+--------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
-----------------------------------------------------------------------------
```

E44.6.2 ZIP Models with Random Parameters

The random parameters model may be extended to the zero inflated Poisson and negative binomials models. The random parameters specification applies to the parameters in the regression model, not the regime splitting model. The model is

$$Y_{it} = 0 \text{ with probability } q_{it},$$

$$Y_{it} \sim \text{Poisson}(\lambda_{it}) \text{ or NegBin}(\lambda_{it},\theta) \text{ with probability } 1 - q_{it}$$

$$\text{Prob}[Y_{it} = 0] = q_{it} + [1 - q_{it}]R_{it}(0), \quad \text{Prob}[Y_{it} = j > 0] = [1 - q_{it}]R_{it}(j)$$

where $$R_{it}(y) = \text{Poisson probability} = e^{-\lambda it}\lambda_{it}^{yit} / y_{it}!, \quad \lambda_{it} = \exp(\boldsymbol{\beta}_i{}'\mathbf{x}_{it})$$

(the random parameters appear in λ_{it}) or,

$$R_{it}(j) = \text{negative binomial probability} = \Gamma(\theta+y_{it})/[y_{it}!\Gamma(\theta)]\, u_{it}^{\theta}\, [1 - u_{it}]^{yit}$$

$$\theta = 1/\alpha, \text{ where } \alpha \text{ is the overdispersion parameter}$$

$$u_{it} = \theta / [\theta + \lambda_{it}],$$

$$q_{it} \sim \text{Logistic}[v_{it}], \quad v_{it} = \boldsymbol{\gamma}'\mathbf{z}_{it}.$$

The command form is

POISSON	**; Lhs**	**= dependent variable**
or NEGBIN	**; Rhs**	**= list of all variables in $\mathbf{x}_i$, including one**
	; ZIP	**; Rh2 = list of variables for regime split, including one**
	; Panel	
	; RPM	**(for random parameters model)**
or	**; RPM**	**= list of variables in $\mathbf{z}_i$**
	; Fcn	**= specification of random parameters**
	; Pts	**= r (number of replications – this is optional)**
	; Cor	**(for correlated parameters – optional) $**

The ***; Fcn = list*** *specification is applied only to the Rhs variables.*

E44.7 Latent Class Models

The count model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is the Poisson

$$\text{Prob}[Y_{it} = y_{it}|\ \lambda_{it}] \ = \ \exp[-\lambda_{it}] \times\lambda_{it}^{yit} \ / \ y_{it}! \ \text{ where } \lambda_{it} \ = \ \exp[\boldsymbol{\beta}'\mathbf{x}_{it}]$$

or negative binomial probability,

$$\text{Prob}[Y_{it} = y_{it}|\ \lambda_{it}, \tau] \ = \ \Gamma(\tau + y_{it})/[y_{it}!\Gamma(\tau)]\ u_{it}^{\tau}\ [1 - u_{it}]^{\,yit} \ \text{ where } \ u_i \ = \tau / [\tau + \lambda_{it}].$$

(We have changed the symbol for the dispersion parameter to avoid a conflict with our generic notation for the latent class probabilities.) Henceforth, we use the term 'group' to indicate the T_i observations on respondent i in periods t=1,...,T_i. The following extends to the negative binomial model as well, but for the moment, we focus on the Poisson model.

Unobserved heterogeneity in the distribution of Y_{it} is assumed to impact the mean (and variance) λ_{it}. The continuous distribution of the heterogeneity is approximated by using a finite number of 'points of support.' The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$ where J is chosen by the analyst.)

The probability of observing y_{it} given that the individual is in class j is

$$P(i,t|j) \ = \ \text{Prob}[Y_{it} = y_{it}|\lambda_{it},j]$$

where the mean $\lambda_{it}|j$ is specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model $\lambda_{it}|j \ = \ \exp[\ \boldsymbol{\beta}'\mathbf{x}_{it} \ + \ \delta_j]$. We formulate this more generally as,

$$\lambda_{it}|j \ = \ \exp[\boldsymbol{\beta}_j'\mathbf{x}_{it}] \text{ and, for the negative binomial model, } \theta|j \ = \ \theta_j.$$

In this formulation, each class has its own parameter vector, $(\boldsymbol{\beta}_j,\theta_j)$though the variables that enter the mean are assumed to be the same. The negative binomial model has a separate dispersion parameter in each class as well. This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters. The prior probabilities for the latent classes are formulated as constants.

$$F_j \ = \ \frac{\exp(\theta_j)}{\sum_{m=1}^{J}\exp(\theta_m)}, \ \theta_J \ = \ 0, \ \Sigma_j\, F_j = 1.$$

The class probabilities may also be functions of a set of covariates, in which case

$$F_{ij} \ = \ \frac{\exp(\boldsymbol{\theta}_j'\mathbf{z}_i)}{\sum_{m=1}^{J}\exp(\boldsymbol{\theta}_m'\mathbf{z}_i)}.$$

An extension to the zero inflation models and a further generalization are presented in Section E44.7.

The estimation command for this model is

```
POISSON      ; Lhs  = ... ; Rhs = independent variables
or NEGBIN    ; LCM (for latent class model)
             ; Pts   = the desired number of classes, 2, 3, ..., 9
             ; Pds  = panel data specification $
```

(The model must be fit with panel data.) The default number of support points is five. But, this is fairly high. You may set J to 2, 3, 4, or 5. To specify that the class probabilities are functions of covariates, use

; LCM = the set of variables

(The default is **; LCM** = **one**. You may omit the '**= one**' if your class probabilities are constant.)

NOTE: For the case in which class probabilities have covariates, it is assumed that these are the same in every period. You should repeat these variables for each observation within the group. The program uses the first row, so, in fact, any data, including zeros, will suffice. However, do not mark observations 2 - T_i for these variables as missing. This will flag the observation as bad data to be bypassed.

Other options are the standard ones for Poisson and negative binomial models, including

; Par	to keep individual specific parameter estimates.
; Keep = name	to retain fitted values
; Res = name	to retain residuals
; Prob = name	to retain estimated probabilities for observed outcome

Some particular values computed for the latent class model are

; Group = the index of the most likely latent class
; Cprob = estimated probability for the most likely latent class

(Computation of these values is described in the technical details.) Other options include

; Maxit = n	to set maximum iterations
; Rst = list	to specify fixed value and equality restrictions
; CML: spec	to define linear constraints
; Test: spec	to define Wald tests

and so on. You can use the **; Rst = list** option to structure the latent class model so that different variables appear in different classes. Alternatively, you can use this to force the Heckman and Singer form of the model as follows, where we use a three class model as an example:

```
NAMELIST    ; x = ... one, list of variables $
CALC        ; k1 =  Col(x) - 1 $
POISSON     ; Lhs = ... ; Rhs = x ; LCM ; Pts = 3
            ; Rst = d1,k1_b, d2,k1_0, d3,k1_0, t1,t2,t3 $
```

Estimates retained by this model include

Matrices:	*b*	=	full parameter vector, $[\beta_1', \beta_2', \ldots F_1, \ldots, F_J]$
	varb	=	full covariance matrix
	beta_i	=	individual specific parameters, if **; Par** is requested.

Note that *b* and *varb* involve $J \times (K+1)$ estimates. Two additional matrices are created

b_class	=	a $J \times K$ matrix with each row equal to the corresponding β_j
class_pr	=	a $J \times 1$ vector containing the estimated class probabilities

Scalars:	*kreg*	=	number of variables in Rhs list
	nreg	=	total number of observations used for estimation
	logl	=	maximized value of the log likelihood function
	exitcode	=	exit status of the estimation procedure

The latent class estimator allows for truncation (not censoring) at zero. The model specification is

; TPM

with no other specifications. This is for the conditional distribution $y_i|y_i > 0$, as appears in hurdle models.

E44.7.1 Testing for Latent Heterogeneity

In order to test for latent class effects, you must compare a model with the effects to one without. This is not a parametric restriction on the latent class model. Note, thus, if θ_j is set equal to zero, this just produces $F_j = 1/J$. Alternatively, forcing all coefficient vectors to equal zero destroys the identifiability of the latent class probabilities – their standard errors will go to $+\infty$. (Try it.) Therefore, in order to test for class effects, the restricted and unrestricted models must be fit separately. One can use a likelihood ratio test, based on the following computations: For the latent class model the unrestricted log likelihood is,

$$\log L_U = \sum\nolimits_{i=1}^{N} \log \sum\nolimits_{j=1}^{J} F_j \prod\nolimits_{t=1}^{T_i} P(i,t|j).$$

For the Poisson or negative binomial model with no latent class sorting, the log likelihood is

$$\log L_R = \sum\nolimits_{i=1}^{N} \log \prod\nolimits_{t=1}^{T_i} P(i,t).$$

In both models, observations within the groups are assumed to be independent. Taking logs in the second expression produces the conventional log likelihood function for the count model,

$$\log L_R = \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T_i} P(i,t).$$

Therefore, it appears that a conventional likelihood ratio statistic can be computed. The degrees of freedom would be (J-1)(1+K). The first J-1 would be for the free latent class probabilities while the latter K(J-1) would be for the additional slope parameters in the last J-1 latent classes. The problem with this approach is that the model is not identified under the restrictions, so this is not a conventional LR test. That is, without the latent class sorting, the extra slope parameters cannot be estimated, and without variation across classes in the slope parameters, the class parameters cannot be estimated. The upshot is that if this is a valid LR statistic, then surely the degrees of freedom is fewer than (J-1)(1+K). But, whether it is appears not to be conclusively determined in the literature. (See Heckman and Singer for discussion.)

E44.7.2 Application

To illustrate the technique, we have applied the technique to the German health care data once again. A three class model for *docvis* is fit without restrictions, then restricted so as to produce the Heckman and Singer form of the model in which only the constant terms differ.

```
POISSON     ; Lhs  = docvis ; Rhs = one,female,hhninc,educ
            ; Panel ; Pts = 3 ; LCM ; Partial Effects $
POISSON     ; Lhs  = docvis ; Rhs = one,female,hhninc,educ
            ; Panel ; Pts = 3 ; LCM ; Partials Effects
            ; Rst  = b01,3_b,b02,3_b,b03,3_b,t1,t2,t3 $
```

```
Normal exit:  42 iterations. Status=0, F=    17537.48
-----------------------------------------------------------------------------
Latent Class / Panel Poisson  Model
Dependent variable                DOCVIS
Log likelihood function    -17537.47833
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    3.78019***      .14204    26.61  .0000     3.50180   4.05858
  FEMALE|     .11507**       .04547     2.53  .0114      .02595    .20418
  HHNINC|     .29716***      .11515     2.58  .0099      .07147    .52286
    EDUC|    -.13751***      .01304   -10.54  .0000     -.16308   -.11195
        |Model parameters for latent class 2
Constant|    1.13460***      .18518     6.13  .0000      .77166   1.49755
  FEMALE|     .47494***      .05972     7.95  .0000      .35790    .59198
  HHNINC|     .22802*        .12580     1.81  .0699     -.01854    .47458
    EDUC|    -.14129***      .01650    -8.56  .0000     -.17363   -.10896
        |Model parameters for latent class 3
Constant|    2.26186***      .09039    25.02  .0000     2.08469   2.43903
  FEMALE|     .24702***      .03307     7.47  .0000      .18221    .31182
  HHNINC|     .28702***      .07523     3.82  .0001      .13957    .43447
    EDUC|    -.10320***      .00852   -12.12  .0000     -.11989   -.08651
        |Estimated prior probabilities for class membership
Class1Pr|     .07954***      .01004     7.93  .0000      .05987    .09921
Class2Pr|     .47122***      .01887    24.97  .0000      .43423    .50821
Class3Pr|     .44924***      .01830    24.55  .0000      .41337    .48510
--------+--------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   2.0799
Scale Factor for Marginal Effects  2.0799
B for latent class model is a wghted avrg.
--------+-----------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
  DOCVIS|      Effect    Elasticity      z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
  FEMALE|     .71532***       .14540     6.05  .0000       .48361    .94702
  HHNINC|     .54082***       .09083     3.58  .0003       .24478    .83685
    EDUC|    -.25766***    -1.35536   -34.92  .0000      -.27212   -.24319
--------+-----------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

--------------------------------------------------------------------------------
Latent Class / Panel Poisson  Model
Dependent variable               DOCVIS
Log likelihood function    -17566.10677
Model fit with  3 latent classes.
--------+-----------------------------------------------------------------------
        |                 Standard             Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    3.57127***      .08933    39.98  .0000     3.39619   3.74635
  FEMALE|     .22546***      .02848     7.92  .0000      .16965    .28127
  HHNINC|     .29045***      .06140     4.73  .0000      .17010    .41080
    EDUC|    -.12384***      .00813   -15.23  .0000     -.13978   -.10791
        |Model parameters for latent class 2
Constant|    1.07754***      .09347    11.53  .0000      .89435   1.26073
  FEMALE|     .22546***      .02848     7.92  .0000      .16965    .28127
  HHNINC|     .29045***      .06140     4.73  .0000      .17010    .41080
    EDUC|    -.12384***      .00813   -15.23  .0000     -.13978   -.10791
        |Model parameters for latent class 3
Constant|    2.49700***      .08922    27.99  .0000     2.32213   2.67187
  FEMALE|     .22546***      .02848     7.92  .0000      .16965    .28127
  HHNINC|     .29045***      .06140     4.73  .0000      .17010    .41080
    EDUC|    -.12384***      .00813   -15.23  .0000     -.13978   -.10791
        |Estimated prior probabilities for class membership
Class1Pr|     .08083***      .00989     8.18  .0000      .06146    .10021
Class2Pr|     .47999***      .01859    25.81  .0000      .44355    .51644
Class3Pr|     .43918***      .01824    24.08  .0000      .40342    .47493
--------+-----------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   2.1050
Scale Factor for Marginal Effects  2.1050
B for latent class model is a wghted avrg.
```

```
--------+-----------------------------------------------------------------
        |     Partial                          Prob.       95% Confidence
  DOCVIS|      Effect    Elasticity      z    |z|>Z*          Interval
--------+-----------------------------------------------------------------
  FEMALE|     .47460***       .09532     6.92  .0000       .34014     .60906
  HHNINC|     .61140***       .10145     4.50  .0000       .34483     .87798
    EDUC|    -.26069***    -1.35494   -42.92  .0000      -.27260    -.24879
--------+-----------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E44.7.3 Latent Class Model with Zero Inflation

You can extend the latent class model to the zero inflation models as well. The extended model departs from the basic probabilities as usual, the Poisson

$$R(i,t|j) = \exp[-\lambda_{it\,|\,j}] \times \lambda_{it\,|\,j}^{\;yit} \;/\; y_{it}!$$

where $\lambda_{it|j} = \exp[\boldsymbol{\beta}_j'\mathbf{x}_{it}]$

or negative binomial probability,

$$R(i,t|j) = \Gamma(\tau_j + y_{it})/[y_{it}!\Gamma(\tau_j)]\; u_{it\,|\,j}^{\;\tau j}\,[1 - u_{it\,|\,j}]^{\;yit}$$

$$\tau_j = 1/\alpha_j, \text{ where } \alpha_j \text{ is the overdispersion parameter}$$

$$u_{it|j} = \tau_j / [\tau_j + \lambda_{it|j}].$$

The mean $\lambda_{it} \mid j$ and, if negative binomial, overdispersion, τ_j, are specific to the group. The zero inflation model then adds

$$Y_{it}|\, j = 0 \text{ with probability } q_{it|j},$$

$$Y_{it}|\, j \sim \text{Poisson}(\lambda_{it\,|\,j}) \text{ or NegBin}(\lambda_{it\,|\,j},\theta_j) \text{ with probability } 1 - q_{it\,|\,j}$$

where $q_{it\,|\,j} = \text{Logit probability } (\boldsymbol{\gamma}_j'\mathbf{z}_{it}) = \dfrac{\exp(\boldsymbol{\delta}_j'\mathbf{z}_{it})}{1+\exp(\boldsymbol{\delta}_j'\mathbf{z}_{it})}.$

Thus, $\text{Prob}[Y_{it} = 0 \mid j] = q_{it|j} + [1 - q_{it|j}]R_{it}(0 \mid j),$

$$\text{Prob}[Y_{it} = m > 0 \mid j] = [1 - q_{it|j}]R_{it}(m \mid j).$$

Combining terms, the preceding define

$$P(i,t \mid j) = \text{Prob}[Y_{it} = y_{it} \mid j]$$

and the rest of the analysis is the same as in the previous section.

The command for this model just adds the ZIP specification to the earlier latent class specification:

POISSON	**; Lhs**	**= ...**
or NEGBIN	**; Rhs**	**= independent variables**
	; ZIP	**; Rh2 = variables in regime probability**
	; LCM	**(for latent class model)**
	; Pts	**= the desired number of classes, 2, 3, 4, or 5**
	; Pds	**= panel data specification $**

All other options and parts of the command are the same as before.

E44.7.4 Technical Details on Estimating Latent Class Models

The sequence of T_i observations for individual i, given group j is $\mathbf{y}(i|j) = [y(i,1|j),y(i,2|j),...,y(i,T_i|j)]$. Observations for individual i in different periods are assumed to be independent. Thus, the joint probability of the sequence of observations $[\mathbf{y}(i|j)]$ is

$$P(i|j) \quad = \quad \prod_{t=1}^{T_i} P(i,t|j).$$

We denote the mass, or probability in interval (group) j as F_j, $j = 1,...,J$, such that $F_1 + F_2 + ... + F_J = 1$. Then, the posterior probability of an observed sequence of observations is

$$P(i) \quad = \quad \sum_{j=1}^{J} F_j P(i|j)$$

where F_j is the prior probability of membership in the jth class. We parameterize the group probabilities with

$$F_j \quad = \quad \frac{\exp(\theta_j)}{\sum_{m=1}^{J}\exp(\theta_m)}, \ \theta_J \ = \ 0, \ \Sigma_j F_j = 1.$$

where $\theta_J = 0$, since $\Sigma_j F_j = 1$. The class probabilities may also be functions of a set of covariates, in which case

$$F_{ij} \quad = \quad \frac{\exp(\boldsymbol{\theta}'_j \mathbf{z}_i)}{\sum_{m=1}^{J}\exp(\boldsymbol{\theta}'_m \mathbf{z}_i)}, \ \boldsymbol{\theta}_J \ = \ \mathbf{0}, \ \Sigma_j F_j = 1.$$

NOTE: In this formulation, the covariates are assumed to be the same in every period. Data on $\mathbf{z}_i$ must be present in every period, but the first observation in each group is used to compute the prior probabilities.

The log likelihood function for the observed sample is

$$\log L = \sum_{i=1}^{N} \log[P(i)]$$

$$= \sum_{i=1}^{N} \log \sum_{j=1}^{J} F_{ij} \prod_{t=1}^{T_i} P(i,t|j).$$

This function is maximized with respect to the vector of parameters

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_J),\ \boldsymbol{\theta}_1,...,\boldsymbol{\theta}_J.$$

subject to the restriction that $\boldsymbol{\theta}_J = \mathbf{0}$. (Other restrictions may be imposed as well.)

Among the useful results of this formulation is a posterior estimate of the probabilities of particular group membership; using Bayes theorem,

$$P(j \mid i) = P(i, j) / P(i)$$

$$= \frac{P(i \mid j)F_{ij}}{\sum_{j=1}^{J} P(i \mid j)F_{ij}}$$

Using this result, we compute j^* = the index of the group with the highest posterior probability. The predicted values, residuals, and predicted probabilities for the observed outcomes are then computed as those associated with group j^*. That is, for example,

$$\text{Fitted value}_{it} = \lambda_{it|j^*} = \exp(\boldsymbol{\beta}_{j^*}'\mathbf{x}_{it})$$

and so on.

Maximization of the log likelihood does not require any unusual techniques or approaches. (Some authors, e.g., Cockburn (1999) have used the EM algorithm for a Poisson model of this sort, but this is a means to an end, not a necessity. We have found that the conventional approach used here works without problems, and is much simpler.) The gradient of the log likelihood function is

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{N} \frac{1}{P_i} F_j P_{i|j} \sum_{t=1}^{T_i} \frac{\partial \log P_{it|j}}{\partial \boldsymbol{\beta}_j}$$

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}_j} = \sum_{i=1}^{N} \frac{1}{P_i} \sum_{m=1}^{J} P_{i|m} F_{im} [1(m=j) - F_j]\mathbf{z}_i$$

The gradients in the first term are $(y_{it} - \lambda_{it|j})\mathbf{x}_{it}$ for the Poisson model and $[\tau_j(y_{it} + q_{it|j}) - \tau_j]\,\mathbf{x}_{it}$ where θ_j is the overdispersion parameter and $q_{it|j} = \tau_j/(\tau_j + \lambda_{it|j})$ for the negative binomial model. The BHHH estimator is used for estimating the asymptotic covariance matrix of the maximum likelihood estimates.

E44.8 GMM Estimators for Count Models with Panel Data

This section develops *LIMDEP* commands to compute GMM estimators for panel data count data models See Blundell, Griffith, and Windmeijer (2002) and also Romeu (2004) for a review of Windmeijer's software that does this. These are not hard coded in *LIMDEP*, but can be directly implemented with trivial changes to the programs listed below. The reader is referred to the articles for theoretical development of the models. The remainder of this section presents the *LIMDEP* code for estimation.

Step 1. Set up the data. The Rhs variables in the model are assumed to be *xa*,*xb*,*xc*,... We assume for the illustration below that this is *xa*,*xb*,*xc*,*xd*,*xe*. The Lhs variable is *y*.

Step 2. Create lagged values of all variables. Use namelists *x* for the current Rhs values and *x*1 for the lagged values. Variables *y* and *y*1 are the current lagged values of the dependent variable.

Step 3. Set up panel data indicators such that variable *t* is the period number. The panel need not be balanced. Also set up the usual group count variable, called, say, *ti*. We also need the group identifier. This would normally be *_stratum* created by the panel estimator. We'll call it *group* where needed below.

The data setup commands for our constructed example are as follows:

```
NAMELIST ; x = xa,xb,xc,xd,xe $
CREATE   ; xa1 = xa[-1] ; xb1 = xb[-1] ; xc1 = xc[-1]
         ; xd1 = xd[-1] ; xe1 = xe[-1]$
CREATE   ; y1= y[-1]$
NAMELIST ; x1 = xa1,xb1,xc1,xd1,xe1 $
```

The sample must also be set to eliminate the lost observation due to the lagging.

```
SAMPLE   ; 2 - n $
```

Step 4. If the models are being estimated using instrumental variables, define namelists for these as well.

```
NAMELIST ; z = za,zb,zc,zd,ze,zf $ ... the list of instrumental variables $
```

Estimators are shown for four model groups: cross sections, panel data, presample means and linear feedback models, respectively. Throughout, the nomenclature used is based on Romeu's Table 1. The table has two columns of model definitions, which we label A and B. In the A column, there are six models, which we label 1A,...,6A. In the second column, we, use 1B, 2B, 3B1, 3B2, 4B1, 4B2, 5B and 6B.

E44.8.1 Cross Section Estimators

For a model with exogenous or predetermined regressors, we use the whole sample. With additive errors, treat it as a cross section, Poisson regression – Models 1A and 3A.

```
SAMPLE     ; All $
POISSON    ; Lhs = y ; Rhs = x $
```

This is for exogenous or predetermined regressors and multiplicative errors, Models 2A and 4A.

```
POISSON    ; Lhs = y ; Rhs = x $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5
           ; Fn1 = (y-exp(x'b1))/exp(x'b1) * xa
           ; Fn2 = (y-exp(x'b1))/exp(x'b1) * xb
           ; Fn3 = (y-exp(x'b1))/exp(x'b1) * xc
           ; Fn4 = (y-exp(x'b1))/exp(x'b1) * xd
           ; Fn5 = (y-exp(x'b1))/exp(x'b1) * xe $
```

With endogenous regressors, the number of equations changes to the number of instrumental variables. Assumed here to be *za,zb,zc,zd,ze,zf*. These are overidentified so we use a two step estimator. The two routines are for additive (Model 5A) and multiplicative errors (Model 6A), respectively.

```
POISSON    ; Lhs = y ; Rhs = x $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5
           ; Fn1 = (y-exp(x'b1)) * za ; Fn2 = (y-exp(x'b1)) * zb
           ; Fn3 = (y-exp(x'b1)) * zc ; Fn4 = (y-exp(x'b1)) * zd
           ; Fn5 = (y-exp(x'b1)) * ze ; Fn6 = (y-exp(x'b1)) * zf $
MATRIX     ; optimal w = <sigma> $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5 ; sigma = optimal w
           ; Fn1 = (y-exp(x'b1)) * za ; Fn2 = (y-exp(x'b1)) * zb
           ; Fn3 = (y-exp(x'b1)) * zc ; Fn4 = (y-exp(x'b1)) * zd
           ; Fn5 = (y-exp(x'b1)) * ze ; Fn6 = (y-exp(x'b1)) * zf $
```

This is for the multiplicative errors model.

```
POISSON    ; Lhs = y ; Rhs = x $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5
           ; Fn1 = (y-exp(x'b1))/exp(x'b1) * za
           ; Fn2 = (y-exp(x'b1))/exp(x'b1) * zb
           ; Fn3 = (y-exp(x'b1))/exp(x'b1) * zc
           ; Fn4 = (y-exp(x'b1))/exp(x'b1) * zd
           ; Fn5 = (y-exp(x'b1))/exp(x'b1) * ze
           ; Fn6 = (y-exp(x'b1))/exp(x'b1) * zf $
MATRIX     ; optimal w = <sigma> $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5 ; sigma = optimal w
           ; Fn1 = (y-exp(x'b1))/exp(x'b1) * xa
           ; Fn2 = (y-exp(x'b1))/exp(x'b1) * xb
           ; Fn3 = (y-exp(x'b1))/exp(x'b1) * xc
           ; Fn4 = (y-exp(x'b1))/exp(x'b1) * xd
           ; Fn5 = (y-exp(x'b1))/exp(x'b1) * xe $
```

E44.8.2 Panel Data Estimators

The panel data estimators for exogenous regressors are not valid with multiplicative errors. (This is cell 2B (no model) in Romeu’s table.) With additive errors, the estimator is the familiar fixed effects estimator, Model 1B.

```
POISSON    ; Lhs = y ; Rhs = x  ; Panel  ; FEM $
```

For additive errors with predetermined and endogenous regressors and also for multiplicative errors, create the instrumental variables from the current and lagged values of z. Collect these in namelist *z*. For the examples, *z* = *z*1,*z*2,*z*3,*z*4,*z*5,*z*6,*z*7. There are some differences across cases in the specific instrumental variables used. Lose an observation in each group because of the lagging.

```
REJECT     ; t = 1 $
```

These are the Chamberlain forms, 3B1 and 4B1 and corresponding dynamic forms

```
POISSON    ; Lhs = y ; Rhs = x $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5
           ; Fn1 = (y*exp(x1'b1)/exp(x'b1) - y1) * z1
           ; Fn2 = (y*exp(x1'b1)/exp(x'b1) - y1) * z2
           ; Fn3 = (y*exp(x1'b1)/exp(x'b1) - y1) * z3
           ; Fn4 = (y*exp(x1'b1)/exp(x'b1) - y1) * z4
           ; Fn5 = (y*exp(x1'b1)/exp(x'b1) - y1) * z5
           ; Fn6 = (y*exp(x1'b1)/exp(x'b1) - y1) * z6
           ; Fn7 = (y*exp(x1'b1)/exp(x'b1) - y1) * z7 $
MATRIX     ; optimal w = <sigma> $
GMME       ; Start = b ; Labels = b1,b2,b3,b4,b5 ; sigma = optimal w
           ; Fn1 = (y*exp(x1'b1)/exp(x'b1) - y1) * z1
           ; Fn2 = (y*exp(x1'b1)/exp(x'b1) - y1) * z2
           ; Fn3 = (y*exp(x1'b1)/exp(x'b1) - y1) * z3
           ; Fn4 = (y*exp(x1'b1)/exp(x'b1) - y1) * z4
           ; Fn5 = (y*exp(x1'b1)/exp(x'b1) - y1) * z5
           ; Fn6 = (y*exp(x1'b1)/exp(x'b1) - y1) * z6
           ; Fn7 = (y*exp(x1'b1)/exp(x'b1) - y1) * z7 $
```

The Wooldridge forms 3B2, 4B2 and 5B, differ in the choices of instruments and the dynamic form of the model.

```
POISSON     ; Lhs = y ; Rhs = x $
GMME        ; Start = b ; Labels = b1,b2,b3,b4,b5
            ; Fn1 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z1
            ; Fn2 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z2
            ; Fn3 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z3
            ; Fn4 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z4
            ; Fn5 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z5
            ; Fn6 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z6
            ; Fn7 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z7 $
MATRIX      ; optimal w = <sigma> $
GMME        ; Start = b ; labels = b1,b2,b3,b4,b5 ; sigma = optimal w
            ; Fn1 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z1
            ; Fn2 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z2
            ; Fn3 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z3
            ; Fn4 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z4
            ; Fn5 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z5
            ; Fn6 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z6
            ; Fn7 = (y/exp(x'b1) - y1/exp(x1'b1) ) * z7 $
```

E44.8.3 Presample Means Estimators

The presample is defined as observations 1 to t_0. The sample is $t_1 = t_0+1$ to t.

```
SAMPLE      ; All $
REJECT      ; t  >  t0  $
MATRIX      ; pmeans = Gxbr(y,group) $
SAMPLE      ; All $
CREATE      ; logpmi = Log(pmeans(group)) $
REJECT      ; t < t1 $
POISSON     ; Lhs = y ; Rhs = x,logpmi $
GMME        ; Start = b ; Labels = b1,b2,b3,b4,b5,f
            ; Fn1 = (y-exp(x'b1-f*logpmi)) * xa
            ; Fn2 = (y-exp(x'b1-f*logpmi)) * xb
            ; Fn3 = (y-exp(x'b1-f*logpmi)) * xc
            ; Fn4 = (y-exp(x'b1-f*logpmi)) * xd
            ; Fn5 = (y-exp(x'b1-f*logpmi)) * xe
            ; Fn6 = (y-exp(x'b1-f*logpmi)) * logpmi $
MATRIX      ; optimal w = <sigma> $
GMME        ; Start = b ; Labels = b1,b2,b3,b4,b5 ; sigma = optimal w
            ; Fn1 = (y-exp(x'b1-f*logpmi)) * xa
            ; Fn2 = (y-exp(x'b1-f*logpmi)) * xb
            ; Fn3 = (y-exp(x'b1-f*logpmi)) * xc
            ; Fn4 = (y-exp(x'b1-f*logpmi)) * xd
            ; Fn5 = (y-exp(x'b1-f*logpmi)) * xe
            ; Fn6 = (y-exp(x'b1-f*logpmi)) * logpmi $
```

E44.8.4 Panel Data Linear Feedback Model Estimators

The linear feedback forms use the second lag of y. The lagged value is created, then the sample is restricted to the useable observations.

```
SAMPLE    ; All $
CREATE    ; y2 = y[-2] $
REJECT    ; t <= 2 $
```

This is the Chamberlain form, 3B1 and 4B1, and the dynamic form

```
POISSON   ; Lhs = y ; Rhs = x,y1 $
GMME      ; Start = b,0 ; Labels = b1,b2,b3,b4,b5,c
          ; Fn1 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z1
          ; Fn2 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z2
          ; Fn3 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z3
          ; Fn4 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z4
          ; Fn5 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z5
          ; Fn6 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z6
          ; Fn7 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z7 $
MATRIX    ; optimal w = <sigma> $
GMME      ; Start = b ; Labels = b1,b2,b3,b4,b5,c ; sigma = optimal w
          ; Fn1 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z1
          ; Fn2 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z2
          ; Fn3 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z3
          ; Fn4 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z4
          ; Fn5 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z5
          ; Fn6 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z6
          ; Fn7 = ((y-c*y1)*exp(x1'b1)/exp(x'b1) - (y1-c*y2)) * z7 $
```

This is the Wooldridge form for 3B2, 4B2 and 5B.

```
POISSON   ; Lhs = y ; Rhs = x $
GMME      ; Start = b,0 ; Labels = b1,b2,b3,b4,b5,c
          ; Fn1 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z1
          ; Fn2 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z2
          ; Fn3 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z3
          ; Fn4 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z4
          ; Fn5 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z5
          ; Fn6 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z6
          ; Fn7 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z7 $
MATRIX    ; optimal w = <sigma> $
GMME      ; Start = b ; Labels = b1,b2,b3,b4,b5,c ; sigma = optimal w
          ; Fn1 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z1
          ; Fn2 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z2
          ; Fn3 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z3
          ; Fn4 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z4
          ; Fn5 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z5
          ; Fn6 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z6
          ; Fn7 = ((y-c*y1)/exp(x'b1) - (y1-c*y2)/exp(x1'b1) ) * z7 $
```

E45: The Tobit Model for Censored Data

E45.1 Introduction

The model and estimators described in this chapter is based on the following general structure:

Latent Underlying Regression: $y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \varepsilon_i \sim N[0,\sigma^2]$.

Observed Dependent Variable: if $y_i^* \leq L_i$, then $y_i = L_i$ (lower tail censoring)

if $y_i^* \geq U_i$, then $y_i = U_i$ (upper tail censoring)

if $L_i < y_i^* < U_i$, then $y_i = y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$.

Various modifications of the model are considered in the sections to follow. Within this framework, the most familiar form is the lower censoring only, at zero variant. Truncation, in which only data in the third group are observed, is a related case which is discussed in Chapter E47. In practice, most of the received applications involve censoring, rather than truncation. The thresholds, L_i and U_i, may be constants or variables. We accommodate censoring in the upper or lower (or both) tails of the distribution. The most familiar case of this model in the literature is the 'tobit' model, in which $U_i = +\infty$ and $L_i = 0$, i.e., the case in which the observed data contain a cluster of zeros. In the standard 'censored regression,' or tobit model, the censored range of y_i^* is the half of the line below zero. (For convenience, we will drop the observation subscript at this point.) If y^* is not positive, a zero is observed for y, otherwise the observation is y^*. Models of expenditure are typical. We also allow censoring of the upper tail ('on the right'). A model of the demand for tickets to sporting events might be an application, since the actual demand is only observed if it is not more than the capacity of the facility (stadium, etc.). A somewhat more elaborate specification is obtained when the range of y^* is censored in both tails. This is the 'two limit probit' model. An application might be a model of weekly hours worked, in which less than half time is reported as 20 and more than 40 is reported as 'full time,' i.e., 40 or more.

The preceding gives the basic model. We also allow for several variations, including a model with heteroscedasticity, models for panel data, two different models of sample selection, and models with nonnormally distributed disturbances. Numerous variants and features of this model are gathered in Chapter E47. The basic model is developed in this chapter.

NOTE: The mere presence of a clump of zeros in the data set does not, by itself, adequately motivate the tobit model. The specification of the model also implies that the nonlimit observations will have a continuous distribution with observations near the limit points. In general, if you try to fit a tobit model, e.g., to financial data in which there is a clump of zeros, and the nonzero observations are ordinary financial variables far from zero, the model is as likely as not to break down during estimation. In such a case, the model of sample selection is probably a more appropriate specification.

The current theoretical literature contains a large amount of material devoted to semiparametric and nonparametric estimation of censored data models. See, e.g., the work of Bo Honoré. (www.princeton.edu/~honore). Section E45.10 provides an estimator for Powell's symmetrically censored least squares estimator at the end of this chapter.

E45.2 Commands

The basic command for estimation of the censored regression, or tobit model is

TOBIT ; Lhs = y ; Rhs = ... $

The default value for the censoring limit is zero, at the left (i.e., the familiar case). Censoring limits can be varied in two fashions. To specify upper, rather than lower tail censoring, add

; Upper

to the model. With no other changes, this would specify a model in which the observed values of the dependent variable would be either zero or negative rather than zero or positive. The specific limit point to use can be changed by using

; Limit = limit value

where '**limit value**' is either a fixed value (number or scalar) or the name of a variable. For example, the model of the demand for sporting events at stadiums with fixed capacities which sell out a significant proportion of the time might be

TOBIT ; Lhs = tickets
; Rhs = one, price, ...
; Upper censoring
; Limit = capacity $

Models with censoring in both tails of the distribution are requested by changing the **; Limit** specification to

; Limits = lower limit, upper limit

where '**lower limit**' and '**upper limit**' are either numbers, scalars, or the names of variables (or one of each). For example, in a labor supply model, we might have

; Limits = 20,40

NOTE: A few of the variants of the tobit model discussed below do not allow variation in the specification of the censoring limits. In particular, in the *nested and bivariate tobit*, and the *sample selection models*, only the default case of censoring from below (at the left) at zero is supported. In these cases, a **; Limit = value** specification will be ignored. For these models, if your censoring is *upper*, instead, multiply the dependent variable by -1, then reverse the signs of the coefficients after estimation. If censoring is at a nonzero value, subtract this value from the Lhs variable before estimation and before the sign switch above.

Starting values for estimation are obtained by ordinary least squares regression of the dependent variable on the regressors. A full set of OLS results is given before any other output is displayed if you request it with **; OLS**. As has been widely documented, these OLS estimates are inconsistent in this setting (usually biased toward zero). The results are presented for comparison purposes only; the actual OLS coefficients are not used for any other purposes by this program. If you do not provide other starting values, the OLS estimates, [$\mathbf{b}$,s] of [$\boldsymbol{\beta}$,σ], are used to begin the iterations.

You may provide your own starting values with

; Start = values for β then σ

The least squares estimates are followed by the iterations.

You may impose fixed value and within equation equality constraints on the coefficients by using the

; Rst = specification

Other restrictions may be imposed with

; CML: specification of linear restrictions

These are discussed in Chapter R13. Note, once again, the parameters that enter this model are $[\beta_1,\beta_2,...,\beta_K,\sigma]$. If you use these options, you must provide exactly K+1 identifiers for the parameters. As in all models, the option will allow you to constrain the variance to equal one of the slopes, but this is likely to impede convergence, and is unlikely to produce a satisfactory model specification.

During the iterations, the parameters are transformed using Olsen's (1978) transformation,

$$[\boldsymbol{\gamma}, \theta] = [\boldsymbol{\beta}/\sigma, 1/\sigma].$$

If you are generating technical output from the iterations in your output file, the reported parameter vector will be scaled. It is unscaled when the iterations are complete. Maximum likelihood estimates are displayed at exit from the iterations. This will include a table of diagnostic statistics and some notation about the specific model along with the standard output. The MLE of σ will appear with the other parameter estimates. The display includes the log likelihood, the values or identity of the lower and upper bounds, and the estimates of $[\boldsymbol{\beta},\sigma]$. As usual, the ancillary parameter, σ, is included with the rest of the estimated parameter vector in the output table.

Other options for the tobit model are the standard ones for nonlinear models, including

; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**.
; List	to display predicted values
; Parameters	to include the estimate of σ in the retained parameter vector
; Maxit = n	to set maximum iterations
; Alg = name	to select algorithm
; Tlf, ; Tlb, ; Tlg	to set the convergence criteria (use **; Set** to keep these settings)
; Output = value	to control the technical output during iterations
; Keep = name	to retain fitted values
; Res = name	to retain residuals
; Partial Effects	

and so on. Sample clustering for the estimated asymptotic covariance matrix may be requested with

; Cluster = specification.

E45.3 Results for the Tobit Model

You may request the display of ordinary least squares results by adding

; OLS

to the command. These will be suppressed if you do not include this request. The OLS values will be used as the starting values for the iterations. Maximum likelihood estimates are presented, as in the example below. Note that unlike most of the discrete choice models, there is no restricted log likelihood presented. The maximum likelihood estimates for a model that contains only a constant term are no less complicated than one with covariates, and there is no closed form solution for the (β,σ) parameter pair for this model. For a general test of the joint significance of all the variables in the model, we suggest the standard trio of tests, which can be carried out as follows: First set up the Rhs variables in the model.

```
NAMELIST    ; xvars = the x variables in the model, without the constant term $
CALC        ; kx  = Col(xvars) $
TOBIT       ; Lhs = y ; Rhs = one $
CALC        ; l0 = logl $
```

This command will produce the Lagrange multiplier statistic.

```
TOBIT       ; Lhs = y ; Rhs = xvars,one ; Start = kx_0,b,s ; Maxit = 0 $
TOBIT       ; Lhs = y ; Rhs = xvars,one $
```

Compute the likelihood ratio statistic.

```
CALC        ; List ; lr = 2*(logl - l0) ; 1 - Chi(lr,kx) $
```

This computes a Wald statistic.

```
MATRIX      ; beta = b(1:kx) ; vb = varb(1:kx,1:kx)
            ; List ; Wald = beta'<vb>beta $
CALC        ; List ; 1 - Chi(wald,kx) $
```

The application below demonstrates use of the commands. Retained output from the model includes

Matrices: *b*, *varb*

Scalars:

s	= estimated σ
ybar, *sy*, *kreg*	= number of coefficients,
nreg	= number of observations
nonlimts	= number of nonlimit observations in estimating sample

Variables: *logl_obs*, *genres_1*, *genres_2*

Last Model: *b_variable names*, *sigma*

Last Function: E[y|$\mathbf{x}$] – see the development in the next section.

The diagnostic information for the model also includes Fin and Schmidt's LM test for the model specification against the alternative suggested by Cragg as well as a test for nonnormality. The tests are described in Sections E45.9.2 and E45.9.3.

E45.4 Partial Effects

The partial effects in the tobit model when censoring is at the left, at zero, are computed using

$$E[y|\mathbf{x}] = \Phi(\boldsymbol{\beta}'\mathbf{x}/\sigma)[\boldsymbol{\beta}'\mathbf{x} + \sigma\phi(\boldsymbol{\beta}'\mathbf{x}/\sigma)/\Phi(\boldsymbol{\beta}'\mathbf{x}/\sigma)].$$

After some algebra, we find

$$\partial E[y|\mathbf{x}]/\partial\mathbf{x} = \Phi(\boldsymbol{\beta}'\mathbf{x}/\sigma)\boldsymbol{\beta}.$$

The preceding is a broad result which carries over to more general models. That is,

$$\partial E[y|\mathbf{x}]/\partial\mathbf{x} = \text{Prob(nonlimit)}\boldsymbol{\beta}$$

for all specifications of the censoring limits, whether in one tail or both. To obtain a display of the marginal effects for the tobit model, add

; Partial Effects

to the **TOBIT** command. A full listing of the marginal effects computed at the sample means, including standard errors, the estimated conditional mean, and the scale factor, will be included in the model output. An example appears below. The partial effects for the tobit model can be obtained with

```
NAMELIST    ; x = the Rhs of the model $
PARTIALS    ; Effects: x ; Means ; Summary $
```

By using the average partial effects instead – omit the **; Means** – you can use the full range of options with the **PARTIALS** and **SIMULATE** commands.

E45.4.1 Notes About Partial Effects in the Tobit Model

The conditional mean function for the latent variable is the latent regression,

$$E[y_i^* \mid \mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta}.$$

In analysis of this regression for prediction purposes or for analysis of partial effects,

$$m_k^* = \partial E[y_i^* \mid \mathbf{x}_i]/\partial x_{ik} = \beta_k$$

is treated as is normally done in conventional linear regression analysis. Standard errors and confidence intervals for predictions of the conditional mean,

$$\hat{E}[y_i^* \mid \mathbf{x}_i] = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$$

and estimates of marginal effects,

$$\hat{m}_k{}^* = \hat{\beta}_k$$

are computable using conventional forms and the estimated asymptotic covariance matrix for the maximum likelihood estimators.

The more difficult computation involves the conditional mean function for the observed random variable, y_i. This is

$$E[y_i \mid \mathbf{x}_i] = [1\text{-}\Phi(\alpha_i)]L_i + \Phi(\alpha_i)\times [\mathbf{x}_i'\boldsymbol{\beta} + \sigma\, \lambda(\alpha_i)].$$

For this conditional mean, the marginal effects are surprisingly simple,

$$m_k = \partial E[y_i \mid \mathbf{x}_i] / \partial x_{ik} = \beta_k \times [1 - \Phi(\alpha_i)].$$

We note at this point a general result that we will use later. It is shown in Greene (1999) that the simple result

marginal effect = coefficient × probability of noncensored observation

is general for censoring in either or both tails of the distribution, *and extends beyond the normal distribution to any continuous distribution for* ε.

E45.4.2 Partial Effect for a Dummy Variable

Typical applications of the censored regression model involve discrete independent variables, often binary variables indicating presence or absence of a condition or whether or not some treatment was experienced. In our application below, we include a dummy variable for whether there are children in the home and whether or not the individual lives in a city. Whether differentiation of the conditional mean provides an accurate measure of the marginal impact of presence or absence of a treatment, or for some other dummy variable, depends on the variables in the model and other factors. Generally, the result will only be approximate, and the correct computation would be

$$\begin{aligned}\text{Impact} &= E[y_i \mid \mathbf{x}_i^1] - E[y_i \mid \mathbf{x}_i^0] \\ &= [1\text{-}\Phi(\alpha_i^1)]L_i + \Phi(\alpha_i^1) \times [\mathbf{x}_i^{1\prime}\boldsymbol{\beta} + \sigma\, \lambda(\alpha_i^1)] - \\ &\quad [1\text{-}\Phi(\alpha_i^0)]L_i + \Phi(\alpha_i^0) \times [\mathbf{x}_i^{0\prime}\boldsymbol{\beta} + \sigma\, \lambda(\alpha_i^0)],\end{aligned}$$

where the superscripts ‘1’ and ‘0’ indicate that in the computation, the dummy variable in the vector $\mathbf{x}_i$ takes values 1 and 0, respectively. Because it uses the means of the data, the internal calculation of partial effects does not accommodate this feature of the data. You should use **PARTIALS** to obtain the partial effects for the tobit model. The example below demonstrates.

E45.5 Predictions and Fit Measures

A listing of predictions is requested with

; List

The predictions and residuals are retained with

; Keep = name to retain predicted values
and **; Res = name** to retain residuals

There is a possible ambiguity in the computation of predictions in this model. Consider, first, the classical normal regression model with standard deviation σ_i. (We do this to avoid having to treat separately the tobit model with heteroscedasticity.) The conditional mean function is

$$E[y_i \mid \mathbf{x}_i] = \boldsymbol{\beta}'\mathbf{x}_i.$$

But, if y_i is restricted to the range $[L_i,U_i]$, the conditional mean becomes

$$E[y_i|\mathbf{x}_i , L_i < y_i < U_i] = \boldsymbol{\beta}'\mathbf{x}_i + \sigma_i \frac{\phi_L - \phi_U}{\Phi_U - \Phi_L}$$

where $\phi_j = \phi[(j - \boldsymbol{\beta}'\mathbf{x}_i) / \sigma_i], j = L_i,U_i$

and $\Phi_j = \Phi[(j - \boldsymbol{\beta}'\mathbf{x}_i) / \sigma_i], j = L_i,U_i$.

(This is the conditional mean function for the truncated regression model in Section E47.4.) With censoring in only one tail, either L_i will be $-\infty$ or U_i will be $+\infty$, in which case, ϕ_j will equal zero and Φ_j will be zero (for L_i) or one (for U_i). For the tobit model, then,

$$E[y_i{*}\mathbf{x}_i] = L_i\text{Prob}[y_i = L_i] + U_i\text{Prob}[y_i = \text{U}_i] + \text{Prob}[L_i < y_i < U_i]\text{E}[y_i \mid L_i < y_i < U_i].$$

This is $L_i\Phi_L + U_i(1 - \Phi_U) + (\Phi_U - \Phi_L)\boldsymbol{\beta}'\mathbf{x}_i + \sigma_i(\phi_L - \phi_U)$.

LIMDEP reports this as the prediction for the tobit model. Once again, in the case of censoring in only one tail, one of the densities is zero, and one of the tail probabilities is either zero or one.

The prediction displayed by **; List** and retained with **; Keep = name** is the conditional mean function listed above. We emphasize, *the prediction is not* $\boldsymbol{\beta}'\mathbf{x}$. The residual that is kept with **; Res** is the difference between actual and predicted values. If you require the linear index, you can obtain it with

```
NAMELIST   ; x = ... the Rhs for your tobit model $
TOBIT      ; Lhs = ... ; Rhs = x $
CREATE     ; xb = x ' b $
```

With **; List** the two additional variables displayed are the estimate of $\boldsymbol{\beta}'\mathbf{x}$ and the estimate of $[\Phi_U - \Phi_L]$. The latter is the estimated probability that the observation is a nonlimit observation. A different kind of residual, Chesher and Irish's (1987) 'generalized residual,' is discussed in Section E45.9.4.

You may use **SIMULATE** to analyze the predictions in the tobit model. The function analyzed by **SIMULATE** is the conditional mean function given earlier.

As in any nonlinear model, there is no obvious, well behaved counterpart to the R^2 in a linear regression (with a constant term) which is fit by ordinary least squares. Many surrogates have been suggested. A lengthy catalog appears in Veall and Zimmermann (1992). These are largely of three types:

1. Correlations between actual and predicted values for the nonlimit observations: Variations on two themes are suggested, one based on the squared correlation of the actual values of y_i and the predictions of y^*, $\hat{E}[y_i^*|\mathbf{x}_i] = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ and one based on the squared correlation of the actual values of y_i and the predictions of $y_i^*|(y_i^* > L_i)$, $\hat{E}[y_i^*|\mathbf{x}_i, y_i^* > L_i] = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \hat{\sigma}\lambda(\hat{\alpha}_i)$. The obvious defect here is that the limit observations are not included in the computation. But, this presents a bit of a dilemma. Simply including the limit observations in the computation would not solve the problem, because the 'fit' aspect in the limit range of the distribution is the model's ability to predict that an observation will be a limit observation, not its ability to predict the limit value, itself.

2. Mixtures of the correlations in type 1 above and the predicted probabilities for the limit observations: These represent an attempt to cover the defect in type 1. Ultimately, these measures end up mixing residuals in the nonlimit observations, which are of the scale of the observed continuous responses with residuals in the limit observations, which are

   $$e_i^0 = (1-d_i) - \Phi(\hat{\alpha}_i) = -\Phi(\hat{\alpha}_i).$$

   The authors of the survey appear skeptical of these measures, perhaps appropriately so.

3. Transformations of the log likelihood function which are bounded by zero and one: The primary virtue of these measures is that they are bounded and usually improve as the model improves. The most widely used is McFadden's

   pseudo-R2' = 1 - logL/logL0,

   where the latter is for a model with only a constant term. Since this mimics behavior of the log likelihood function, itself, the value added of the normalization seems modest. *The measures do not relate to a proportion of variation explained*, and they only range from zero to one because of the normalization. On the other hand, for purposes of comparing two models, one of which is nested within the other, the difference in the pseudo-R^2s will be a simple function of the likelihood ratio statistic that could be used to test the hypothesis of the restrictions. Of course, since this is the case, one might want to proceed directly to the likelihood ratio statistic and not bother with the 'fit measures.'

The authors of the survey suggest two criteria for fit measures: They should, at least roughly, mimic the OLS-R^2, and they should converge to the OLS-R^2 as the censoring probability goes to zero (since, in this case, the model converges to a linear regression model). The fit measures suggested earlier are based on the continuous data. However, prediction of the limit values is part of the purpose of the model. With that in mind, *LIMDEP* presents two alternatives which appear to satisfy the first criterion, and do meet the second:

$$R^2_{ANOVA} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \overline{\hat{y}}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} = \frac{\text{Var[predicted conditional mean]}}{\text{Var[dependent variable]}}$$

and

$$R^2_{DECOMPOSITION} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

$$= \frac{\text{Variation of predicted mean}}{\text{Variation of predicted mean} + \text{Residual variation}}$$

Both measures use the full sample of observations. In both cases, the conditional mean function, or prediction is $E[y_i \mid \mathbf{x}_i] = [1-\Phi(\alpha_i)]L_i + \Phi(\alpha_i)\times[\mathbf{x}_i'\boldsymbol{\beta} + \sigma\,\lambda(\alpha_i)]$. The first fit measure takes the variance of the estimated conditional mean divided by the variance of the observed variable. In the population, for any $\mathbf{x}_i$, the total variance equals the variance of the conditional mean plus the residual variance. The numerator estimates a sample average of the first of these while the denominator averages the sum of the two. The second measure takes the variance of the conditional mean function around the overall mean of the data in the numerator. The denominator contains the sum of the numerator and a residual variance, the true value minus the conditional mean function. In a linear regression, both measures equal R^2 by construction. (*Note that the second does not equal zero in a model with only a constant*.)

E45.6 Robust and Cluster Corrected Covariance Matrix

Generally, the objective of a robust covariance matrix estimator is to use an estimator of the asymptotic covariance matrix of the MLE that is robust to certain misspecifications of the model. The estimator typically used is the 'sandwich' estimator, $\mathbf{V} = \mathbf{H}^{-1} \times \mathbf{BHHH} \times \mathbf{H}^{-1}$, where **H** is the negative of the second derivatives matrix of the log likelihood function and **BHHH** is the Berndt et al. outer products of first derivatives estimator. Under certain circumstances (see Greene (2011)), the MLE is consistent in the presence of certain misspecifications of the model, though the standard estimators of the asymptotic covariance matrix is inappropriate. The sandwich estimator provides the needed correction. However, the maximum likelihood estimators of the coefficients in the censored regression models are inconsistent in the presence of

- heteroscedasticity
- omitted variables, even if they are orthogonal to the included ones,
- incorrect assumption of the normal distribution,
- incorrect functional form,
- measurement error,
- fixed effects in panel data (omitted variables),
- random effects in a panel data (autocorrelation).

That leaves very little for the robust estimator to be robust to. One possibility is unobserved heterogeneity in a cross section, but only if it is orthogonal to the included variables. It is difficult to construct a case for the estimator – for example, this model is quite far removed from the linear exponential families analyzed by Gourieroux, Monfort, and Trognon (1984). In the end, for better or worse, the specification of the censored normal regression model is fairly fragile, and robust estimation of the asymptotic covariance is essentially a moot point.

The preceding notwithstanding, there are two robust covariance matrices obtainable with the tobit estimator. The **; Cluster** specification provides both. The ordinary sandwich estimator can be obtained with

; Cluster = 1

while if you have clustered data, use

; Cluster = fixed number of observations or stratification variable

The estimator for stratified and clustered data that uses

; Stratum = specification

is also supported. Details appear in Section R10.2.

E45.7 Application of the Tobit Model

We will demonstrate a few of the tobit estimators and carry out some specification tests and secondary analyses of the models. The data used are the Mroz (1987) data on female labor supply. This data set contains 753 observations on women's labor market experience. (The data are provided in data file Mroz.dat.)

The data file is taken from the 1976 panel study of income dynamics, and is based on data for the previous year, 1975. Of the 753 observations, the first 428 are for women with positive hours worked in 1975, while the remaining 345 observations are for women who did not work for pay in 1975. The listing below lists a few observations to illustrate. There are 19 variables in the data set and one which is to be constructed from the data read in:

lfp = a dummy variable = 1 if woman worked in 1975, else 0
whrs = wife's hours of work in 1975
kl6 = number of children less than 6 years old in household
k618 = number of children between ages 6 and 18 in household
wa = wife's age
we = wife's educational attainment, in years
ww = wife's average hourly earnings, in 1975 dollars
rpwg = wife's wage reported at the time of the 1976 interview
hhrs = husband's hours worked in 1975
ha = husband's age
he = husband's educational attainment, in years
hw = husband's wage, in 1975 dollars
faminc = family income, in 1975 dollars
mtr = marginal tax rate facing the wife
wmed = wife's mother's educational attainment, in years
wfed = wife's father's educational attainment, in years
un = unemployment rate in county of residence, in percentage points
cit = dummy variable = 1 if live in large city (SMSA), else 0
ax = actual years of wife's previous labor market experience
prin = *faminc* - (*whrs***ww*) = wife's property income (computed)

IMPORT $

```
LFP WHRS KL6 K618 WA WE WW RPWG HHRS HA HE HW FAMINC MTR WMED WFED UN CIT AX
 1 1610 1 0 32 12  3.3540  2.65 2708 34 12  4.0288 16310 .7215 12  7  5.0 0 14
 1 1656 0 2 30 12  1.3889  2.65 2310 30  9  8.4416 21800 .6615  7  7 11.0 1  5
 1 1980 1 3 35 12  4.5455  4.04 3072 40 12  3.5807 21040 .6915 12  7  5.0 0 15
 1  456 0 3 34 12  1.0965  3.25 1920 53 10  3.5417  7300 .7815  7  7  5.0 0  6
 0    0 0 0 54 14  0.0000  0.00 1960 58 14  7.9082 33856 .7215 12 12  9.5 1 10
 0    0 1 2 30 12  0.0000  3.00 2940 31 17  6.9728 20500 .6915 12 12  7.5 1  4
 0    0 0 0 55 12  0.0000  0.00 2467 56 11  4.9181 28600 .5815  7  7  5.0 1  0
 0    0 0 1 51 10  0.0000  0.00 2256 56 12  8.3112 18750 .6915 10 10 11.0 0 10
 0    0 0 1 44 12  0.0000  0.00 1680 46 12  7.1429 20300 .7215  7  7  9.5 1  5
```

```
NAMELIST    ; x = kl6,k618,wa,we,,cit,one $
CREATE      ; logwage = 0 ; If(ww > 0) logwage = Log(ww) $
CREATE      ; prin = faminc - (whrs * ww) $
TOBIT       ; Lhs = whrs ; Rhs = x ; Partial Effects $
PARTIALS    ; Effects: x ; Summary $
PARTIALS    ; Effects: wa & wa = 25(5)65 ; Plot(ci) $
CALC        ; logl1 = logl $
```

These are the tobit estimates of an hours equation with the partial effects computed at the means.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                 WHRS
Log likelihood function     -3903.79391
Estimation based on N =    753, K =   7
Inf.Cr.AIC  =   7821.6 AIC/N =   10.387
Threshold values for the model:
Lower=     .0000     Upper=+infinity
LM test [df] for tobit=     32.508[  6]
Normality Test, LM    =     10.378[  2]
ANOVA  based fit measure =    .048112
DECOMP based fit measure =    .164940
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
     KL6|   -1075.62***    126.0104    -8.54  .0000    -1322.60   -828.65
    K618|   -127.723***    42.74150    -2.99  .0028    -211.495   -43.952
      WA|   -40.7847***     7.73797    -5.27  .0000    -55.9509  -25.6186
      WE|    98.8168***    23.18132     4.26  .0000     53.3823  144.2514
     CIT|   -93.6141       108.0917     -.87  .3865   -305.4700  118.2418
Constant|    1308.73***    482.7473     2.71  .0067      362.56   2254.89
        |Disturbance standard deviation
   Sigma|    1280.45***    48.15479    26.59  .0000     1186.07   1374.83
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point 674.3503
Scale Factor for Marginal Effects   .5924
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
    WHRS|      Effect         Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     KL6|   -637.200***    73.29245    -8.69  .0000    -780.850  -493.549
    K618|   -75.6635***    25.35208    -2.98  .0028   -125.3526  -25.9743
      WA|   -24.1609***     4.56870    -5.29  .0000    -33.1154  -15.2064
      WE|    58.5391***    13.65949     4.29  .0000     31.7670   85.3112
     CIT|   -55.4571       64.03526     -.87  .3865   -180.9639   70.0498
--------+--------------------------------------------------------------------
```

These are the average partial effects using the same model results. A second analysis examines the partial effect of *age* at various values of *age*.

PARTIALS ; Effects: x ; Summary

```
-----------------------------------------------------------------------
Partial Effects for Tobit (Censored) Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
-----------------------------------------------------------------------
      KL6       -633.69216   73.01543    8.68  -776.79977  -490.58456
      K618       -75.24695   25.17579    2.99  -124.59059   -25.90332
      WA         -24.02790    4.51593    5.32   -32.87896   -15.17684
      WE          58.21688   13.53082    4.30    31.69696    84.73681
   *  CIT        -55.50839   64.48895     .86  -181.90441    70.88764
-----------------------------------------------------------------------
Partial Effects  Analysis for Tobit (Censored) Regression Function
-----------------------------------------------------------------------
Effects on function with respect to WA
Results are computed by average over sample observations
Partial effects for continuous WA       computed by differentiation
Effect is computed as derivative     = df(.)/dx
-----------------------------------------------------------------------
df/dWA             Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
-----------------------------------------------------------------------
APE. Function    -24.02790    4.51593    5.32   -32.87896   -15.17684
WA       = 25.00 -31.29486    7.00604    4.47   -45.02670   -17.56303
WA       = 30.00 -29.45665    6.43442    4.58   -42.06811   -16.84518
WA       = 35.00 -27.44400    5.75781    4.77   -38.72932   -16.15869
WA       = 40.00 -25.28392    4.99237    5.06   -35.06897   -15.49887
WA       = 45.00 -23.01246    4.16318    5.53   -31.17229   -14.85263
WA       = 50.00 -20.67312    3.30307    6.26   -27.14713   -14.19911
WA       = 55.00 -18.31445    2.45094    7.47   -23.11829   -13.51060
WA       = 60.00 -15.98700    1.65148    9.68   -19.22390   -12.75010
WA       = 65.00 -13.74003     .96520   14.24   -15.63182   -11.84824
```

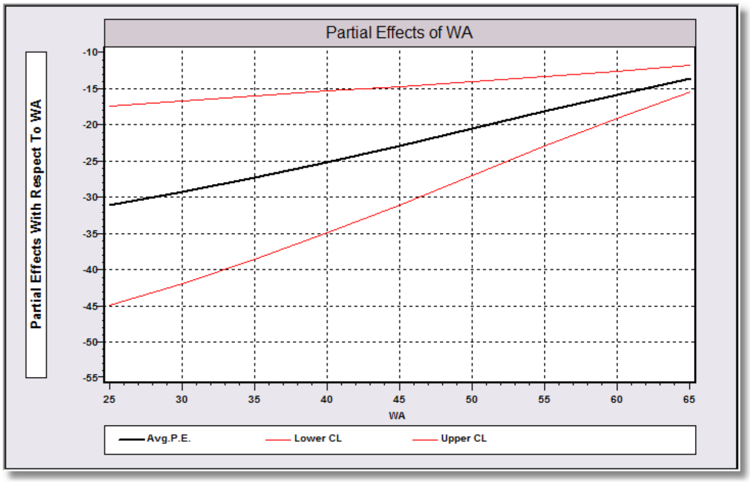


Figure E45.1 Partial Effects of Wife's Age

The following set of commands tests the hypothesis that the slope coefficients in the model are all zero, using the Lagrange multiplier, Wald and likelihood ratio tests. The first estimated model contains only a constant term.

```
CALC      ; kx  = Col(x) - 1 $
TOBIT     ; Lhs = whrs ; Rhs = x $
CALC      ; logl1 = logl $
MATRIX    ; beta = b(1:kx) ; vb = Varb(1:kx,1:kx)
          ; List ; Wald = beta'<vb>beta  $
CALC      ; List ; 1 - Chi(wald,kx) $
TOBIT     ; Lhs = whrs ; Rhs = one $
CALC      ; logl0 = logl $
TOBIT     ; Lhs = whrs ; Rhs = x ; Start = kx_0,b,s ; Maxit = 0 $
CALC      ; List ; LR = 2*(logl1 - logl0) ; 1 - Chi(lr,kx) $
```

Using the full model, we use **MATRIX** to compute the Wald statistic.

```
    WALD|             1
--------+--------------
       1|       94.7034
```

We then carry out the LM test by using the zero slope values as starting values and suppressing the iterations. The LM statistic is 97.61745.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                   WHRS
LM Stat. at start values         97.61745
LM statistic kept as scalar      LMSTAT
Log likelihood function      -3954.89176
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
     KL6|         0.0      120.8266      .00 1.0000 -.23682D+03  .23682D+03
    K618|         0.0      45.02409      .00 1.0000 -.88246D+02  .88246D+02
      WA|         0.0       8.17050      .00 1.0000 -.16014D+02  .16014D+02
      WE|         0.0      24.35427      .00 1.0000 -.47733D+02  .47733D+02
     CIT|         0.0      114.7817      .00 1.0000 -.22497D+03  .22497D+03
Constant|     312.841      514.4513      .61  .5431    -695.465  1321.147
        |Disturbance standard deviation
   Sigma|     1375.21***   53.16139    25.87  .0000     1271.02   1479.41
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The likelihood ratio statistic is computed using the log likelihood from the full model and the log likelihood from the model with only a constant term. The statistic equals 102.1957013.

```
 [CALC] LR      =    102.1957013
```

E45.8 Technical Details

Estimation of the censored regression model is quite routine. For fully parametric specifications – that is, in which the full distribution is specified, maximum likelihood is the accepted method. The likelihood is formulated as follows: For observations which are censored, terms in the log likelihood are the probability of observing the discrete value. For uncensored observations, the term is the usual one, the density for the continuous random variable. Thus, for the censored normal regression model with censoring only in the lower tail,

$$\log L = \sum_{Observations\ with\ y_i = L_i} \log\Phi\left(\frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) + \sum_{Observations\ with\ y_i = y_i*} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right].$$

The model is estimated using Olsen's (1978) transformation of the parameters, $\theta = 1/\sigma$ and $\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta}$. Let $d_i = 1$ for the noncensored observations and 0 otherwise. Then, the log likelihood simplifies to

$$\log L = \sum_{d_i=0} \log\Phi\left(\theta L_i - \mathbf{x}_i'\boldsymbol{\gamma}\right) + \sum_{d_i=1}(-1/2)\left(\log(2\pi) - \log\theta^2 + (\theta y_i - \mathbf{x}_i'\boldsymbol{\gamma})^2\right)$$

$$\frac{\partial \log - L}{\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}} = \begin{pmatrix} \sum_{d_i=0} -\dfrac{\phi(a_i)}{\boldsymbol{\Phi}(a_i)}\mathbf{x}_i + \sum_{d_i=1} e_i\mathbf{x}_i \\ \sum_{d_i=0}\dfrac{\phi(a_i)}{\boldsymbol{\Phi}(a_i)}L_i - \sum_{d_i=1}\left(e_i y_i - \dfrac{2}{\theta}\right)\end{pmatrix}$$

$$= \sum_{i=1}^{n}(1-d_i)\lambda_i^0\begin{pmatrix}\mathbf{x}_i\\ -L_i\end{pmatrix} + d_i\left[e_i\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix} + \begin{pmatrix}\mathbf{0}\\ 1/\theta\end{pmatrix}\right]$$

$$= \sum_{i=1}^{n}\mathbf{g}_i,$$

in which $a_i = \theta L_i$ - $\mathbf{x}_i'\boldsymbol{\gamma}$, $\lambda_i^0 = -\phi(a_i)/\Phi(a_i)$, and $e_i = \theta y_i$ - $\mathbf{x}_i'\boldsymbol{\gamma}$. The Hessian is

$$\frac{\partial^2 \log L}{\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}'} = -\sum_{i=1}^{n}(1-d_i)\delta_i^0\begin{pmatrix}\mathbf{x}_i\\ -L_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ -L_i\end{pmatrix}' + d_i\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}' + d_i\begin{pmatrix}\mathbf{0}\\ 1/\theta\end{pmatrix}\begin{pmatrix}\mathbf{0}\\ 1/\theta\end{pmatrix}'.$$

The asymptotic covariance matrix for the maximum likelihood estimator is usually estimated by inserting the MLEs of $\boldsymbol{\gamma}$ and θ into the Hessian, then inverting. But, the BHHH estimator (sum of outer products of gradients)

$$\textbf{BHHH} = \left[\sum_{i=1}^{n}\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'\right]^{-1}$$

is used occasionally instead.

The preceding obtains the estimated asymptotic covariance matrix for the MLEs of $\boldsymbol{\gamma}$ and θ. To recover the counterpart for the original parameters, $\boldsymbol{\beta}$ and σ, we use the delta method. Let $\hat{\mathbf{V}}$ be the estimated covariance, and let

$$\mathbf{G} = \begin{bmatrix} \partial\boldsymbol{\beta}/\partial\boldsymbol{\gamma}' & \partial\boldsymbol{\beta}/\partial\theta \\ \partial\sigma/\partial\boldsymbol{\gamma}' & \partial\sigma/\partial\theta \end{bmatrix} = \begin{bmatrix} \frac{1}{\theta}\mathbf{I} & \frac{-1}{\theta^2}\boldsymbol{\gamma} \\ \mathbf{0}' & \frac{-1}{\theta^2} \end{bmatrix}.$$

Then, the estimated asymptotic covariance matrix for the MLEs of $\boldsymbol{\beta}$ and σ is

$$\hat{\mathbf{S}}\left(\hat{\boldsymbol{\beta}},\hat{\sigma}\right) = \hat{\mathbf{G}}\hat{\mathbf{V}}\hat{\mathbf{G}}''.$$

The Hessian for this model is negative definite for all values of the parameters. As such, estimation by Newton's method is the standard approach. (See, e.g., Fair (1978), Pratt (1981), Olsen (1978), and Amemiya (1984) on various aspects of computation of the maximum likelihood estimator.) Convergence to a maximum of the log likelihood is usually routine, the more so when Olsen's transformation is used.

NOTE: If convergence is not achieved in a relatively small number of Newton iterations, this is usually indicative of a problem with the data. Widely disparate scaling of the variables or near collinearity is likely to cause this situation.

The maximum likelihood estimator is consistent, efficient, and asymptotically normally distributed, in the fashion of other familiar estimators, such as in the probit model. The fact that the density function for the observed random variable, y_i, is a mixture of discrete and continuous underlying distributions is a complication of some magnitude that was addressed in Amemiya's (1973) seminal paper on the subject. The end result is that maximum likelihood estimation can proceed as usual in spite of the complication.

The full specification with censoring in both tails adds some complication. The full model is

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \text{ where } \varepsilon_i|\mathbf{x} \sim N[0,\sigma^2].$$

$$y_i = \begin{cases} L_i \ \textit{if} \ \ y_i^* \le L_i \\ y_i^* \ \textit{if} \ \ L_i < y_i^* < U_i \\ U_i \ \textit{if} \ \ y_i^* \ge U_i \end{cases}$$

The algebraic results for this formulation are essentially the same. The log likelihood now has three terms:

$$\log L = \sum\nolimits_{y_i=L_i} \log \boldsymbol{\Phi}\left(\theta L_i - \mathbf{x}_i'\boldsymbol{\gamma}\right) + \sum\nolimits_{y_i=U_i} \log\left[1-\boldsymbol{\Phi}\left(\theta U_i - \mathbf{x}_i'\boldsymbol{\gamma}\right)\right] + \sum\nolimits_{y_i=y_i^*} (-1/2)\left(\log(2\pi) - \log\theta^2 + (\theta y_i - \mathbf{x}_i\boldsymbol{\gamma})^2\right)$$

Let $\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta}$, and $\theta = (1/\sigma)$,

L_i, U_i = lower, upper censoring limits (may be $-\infty$, $+\infty$, a number, or variable)

and $\varepsilon_i = \theta y_i - \boldsymbol{\gamma}'\mathbf{x}_i$.

Individual terms in the log likelihood function and derivatives are as follows. To avoid a possible source of confusion, we will now use P_i to denote a term in the log likelihood, rather than L_i which to this point has stood both for the lower limit value and the term now denoted P_i. For nonlimit observations:

$$\log P_i = -(1/2)\varepsilon_i^2 + \log\theta - ½\log 2\pi,$$

$$\partial\log P_i/\partial\boldsymbol{\gamma} = \varepsilon_i\mathbf{x}_i,$$

$$\partial\log P_i/\partial\theta = -\varepsilon_i y_i + 1/\theta,$$

$$\partial^2\log P_i/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}' = -\mathbf{x}_i\mathbf{x}_i',$$

$$\partial^2\log P_i/\partial\boldsymbol{\gamma}\partial\theta = \mathbf{x}_i y_i,$$

$$\partial^2\log P_i/\partial\theta^2 = -y_i^2 - 1/\theta^2.$$

For limit observations:

$$z_i = \boldsymbol{\gamma}'\mathbf{x}_i - \theta U_i \text{ if } y_i \geq U_i \text{ or } z_i = \theta L_i - \boldsymbol{\gamma}'\mathbf{x}_i \text{ if } y_i \leq L_i,$$

$$\log P_i = \log\Phi(z_i),$$

$$\partial\log P_i/\partial\boldsymbol{\gamma} = [\phi(z_i)/\Phi(z_i)]\mathbf{x}_i \text{ if } y_i \geq U_i\text{, reverse sign if } y_i \leq L_i,$$

$$\partial\log P_i/\partial\theta = -[\phi(z_i)/\Phi(z_i)]U_i \text{ or } +[...]L_i \text{ if } y_i \leq L_i.$$

Let $\delta_i = (\phi/\Phi)[z_i + \phi/\Phi].$

Then,

$$\partial^2\log P_i/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}' = -\delta_i\mathbf{x}_i\mathbf{x}_i',$$

$$\partial^2\log P_i/\partial\boldsymbol{\gamma}\partial\theta = \delta_i U_i \text{ or } \delta_i L_i,$$

$$\partial^2\log P_i/\partial\theta^2 = -\delta U_i^2 \text{ or } -\delta L_i^2.$$

Actual Hessians are used to estimate the asymptotic covariance matrices.

Remaining results are based on

$$\begin{aligned}\text{Prob(noncensored)} &= 1 - \text{Prob(censored in lower tail)} – \text{Prob(censored upper tail)}\\ &= 1 - \Phi(\alpha_i^L) - [1-\Phi(\alpha_i^U)]\\ &= \Phi(\alpha_i^U) - \Phi(\alpha_i^L)\end{aligned}$$

where $\alpha_i^L = (L_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma$ and $\alpha_i^U = (U_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma$. This produces the conditional mean function

$$E[y_i|\mathbf{x}_i] = \Phi(\alpha_i^L)L_i + [1-\Phi(\alpha_i^U)]U_i + [\Phi(\alpha_i^U) - \Phi(\alpha_i^L)]\left[\mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\phi(\alpha_i^L) - \phi(\alpha_i^U)}{\Phi(\alpha_i^U) - \Phi(\alpha_i^L)}\right].$$

The conditional mean function for the uncensored (truncated) variable appears in the brackets on the first line above. Other characteristics of the model are essentially as before. The marginal effects are once again a fraction of the underlying regression slopes

$$m_k = \beta_k[\Phi(\alpha_i^U) - \Phi(\alpha_i^L)].$$

The computation for dummy variables would proceed as before, by evaluating the conditional mean at the two points and computing the difference. A standard error would be computed by the delta method. The formidable algebra implied by the now quite complicated functional form suggests that the payoff to numerical differentiation using **WALD** will be considerable. (A general program for censoring using *LIMDEP* is given above.)

The McDonald and Moffitt decomposition described below is still obtainable in two parts based on the truncated variance, but the split is now more complicated;

$$\delta_i = \left[\frac{\phi(\alpha_i^L) - \phi(\alpha_i^U)}{\Phi(\alpha_i^U) - \Phi(\alpha_i^L)}\right]^2 - \frac{\alpha_i^L\phi(\alpha_i^L) - \alpha_i^U\phi(\alpha_i^U)}{\Phi(\alpha_i^U) - \Phi(\alpha_i^L)}.$$

E45.9 Specification Analysis

The tobit model in the form given above has provided a workhorse for a wealth of empirical research. In view of its widespread application, it is natural to expect there to be a variety of specification tests and analyses of the model. Several of the more common of these are described in this section. Some, such as the LM test for Cragg's model are built into the program, while others are implemented using sets of *LIMDEP* commands.

E45.9.1 McDonald and Moffitt's Decomposition of the Conditional Mean

A frequently cited result due to McDonald and Moffitt (1980) decomposes changes in the conditional mean into two parts. The conditional mean for the tobit model with simple zero lower tail censoring is

$$\mathrm{E}[y|\mathbf{x}] = 0 \times \mathrm{Prob}(y = 0) + \mathrm{Prob}(y > 0) \times E(y|\mathbf{x}, y > 0).$$

It follows that,

$$\partial E[y|\mathbf{x}]/\partial\mathbf{x} = \mathrm{Prob}(y > 0) \times \partial E[y|\mathbf{x}, y > 0]/\partial\mathbf{x} + E[y|\mathbf{x}, y > 0] \times \partial\mathrm{Prob}(y > 0)/\partial\mathbf{x}.$$

This breaks the slope into two parts:

- the change in *y* given nonlimit times the probability of being above the limit value,
- the change in the probability of being above the limit times the conditional mean.

(See their paper for discussion.) The formal counterparts to these expressions are:

$$\partial E[\mathrm{y}|\mathbf{x}]/\partial\mathbf{x} = [\Phi(1 - (\phi/\Phi)(\boldsymbol{\beta}'\mathbf{x}/\sigma + \phi/\Phi)) + \phi(\boldsymbol{\beta}'\mathbf{x}/\boldsymbol{\sigma} + \phi/\Phi)]\boldsymbol{\beta}.$$

The routines below do this computation twice. The first time, for illustration, it is done with **CALC** just to get the value of the expressions. Second, we use **WALD** to estimate a standard error for each part.

A small application based on a specific model that one might have fit could be as follows. This computes the decomposition.

```
NAMELIST    ; x = the set of variables $
TOBIT       ; Lhs = y ; Rhs = x ; Parameters $
CALC        ; kx = Col(x) $
MATRIX      ; xb = Mean(x) ; beta = b(1:kx) $
CALC        ; bxs = beta'xb/s ; mu = N01(bxs)/Phi(bxs)
            ; p = Phi(bxs)
            ; p1 = p*(1-bxs*mu-mu^2)
            ; p2 = N01(bxs)*(bxs + mu) $
WALD        ; Labels = kx_b,v ; Start = b  ; Var = varb
            ; Fn1 = b1'xb/v
            ; Fn2 = Phi(Fn1)
            ; Fn3 = N01(Fn1)/Fn2
            ; Fn4 = Fn2*(1-Fn3*(Fn1+Fn3))
            ; Fn5 = Fn2*Fn3*(Fn1+Fn3) $
```

As an alternative, the following is a general procedure that you can use with any model. First, set the model up with these three commands. The rest is standard. The routine computes the full set of marginal effects for the estimated model, decomposed by the formula given above.

```
NAMELIST    ; x = the list of Rhs variables $
CREATE      ; y = the dependent variable $
CALC        ; li = the lower limit value (usually zero) $
```

The remaining computations can use a standard procedure.

```
PROC  = mcdnm(y,x, li) $
TOBIT       ; Lhs = y ; Rhs = x ; Par ; Limit = li ; Partial Effects $
MATRIX      ; xb = Mean(x) $
CALC        ; k = Col(x) $
WALD        ; Labels = k_b,v
            ; Start = b ; Var = varb
            ; Fn1 = alpha = (li - b1'xb)/v
            ; Fn2 = p_censrd = 1-Phi(Fn1)
            ; Fn3 = lambda = N01(Fn1) / Fn1
            ; Fn4 = delta = Fn3*Fn3 - Fn1*Fn3
            ; Fn5 = firtsprt = Fn2*Fn4
            ; Fn6 = secndprt = Fn2*(1-Fn4)
            ; Fn7 = effect = Fn5 + Fn6 $
CALC        ; Part1 = Waldfns(5)
            ; Part2 = Waldfns(6) $
MATRIX      ; beta= b(1:k)
            ; me1= part1 * beta
            ; me2= part2 * beta
            ; me = me1 + me2
            ; List ; me12 = [me1,me2,me] $
ENDPROC $
EXEC        ; Proc  =  mcdnm(y,x,0)$
```

There is a modification that might be useful. The routine does not actually compute the standard error for the marginal effects. It computes the standard errors for the scale factors used to compute the marginal effects (Fn5, Fn6 and Fn7). The actual scaled coefficient vector, in parts, is given by the **MATRIX** command. If you want to compute the decomposition with standard errors for a particular coefficient, you can do so by simply adding these lines to the routine. We suppose that *b*3 is the coefficient of interest. You could add

; Fn8 = b3*Fn5 ; Fn9 = b3*Fn6 ; Fn10 = b3*Fn7

to the **WALD** command to accomplish this.

The following applies the procedure to the labor supply data. The model is set up with

```
NAMELIST   ; x = one,kl6,k618,wa,we,cit $ the list of Rhs variables
CREATE     ; y = whrs $  the dependent variable
CALC       ; li = 0 $ the lower limit value (usually zero)
EXECUTE    ; Proc = mcdnm(y,x,li) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                     Y
Estimation criterion        -3903.79391
Estimation based on N =    753, K =   7
Inf.Cr.AIC  =   7821.6 AIC/N =   10.387
Threshold values for the model:
Lower=     .0000     Upper=+infinity
LM test [df] for tobit=     32.508[  6]
Normality Test, LM    =     10.378[  2]
ANOVA  based fit measure =    .048112
DECOMP based fit measure =    .164940
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    1308.73***    482.7473     2.71  .0067      362.56   2254.89
     KL6|   -1075.62***    126.0104    -8.54  .0000    -1322.60   -828.65
    K618|   -127.723***    42.74150    -2.99  .0028    -211.495   -43.952
      WA|   -40.7847***     7.73797    -5.27  .0000    -55.9509  -25.6186
      WE|    98.8168***    23.18132     4.26  .0000     53.3823  144.2514
     CIT|   -93.6141       108.0917     -.87  .3865   -305.4700  118.2418
        |Disturbance standard deviation
   Sigma|    1280.45***    48.15479    26.59  .0000     1186.07   1374.83
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------------
```

```
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point 674.3503
Scale Factor for Marginal Effects   .5924
```

Y	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
KL6	-637.200***	73.29245	-8.69	.0000	-780.850	-493.549
K618	-75.6635***	25.35208	-2.98	.0028	-125.3526	-25.9743
WA	-24.1609***	4.56870	-5.29	.0000	-33.1154	-15.2064
WE	58.5391***	13.65949	4.29	.0000	31.7670	85.3112
CIT	-55.4571	64.03526	-.87	.3865	-180.9639	70.0498

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

```
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
VC matrix for the functions is singular.
Standard errors are reported, but the
Wald statistic cannot be computed.
Functions are computed at means of variables
```

WaldFcns	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
ALPHA	-.23372***	.04711	-4.96	.0000	-.32606	-.14138
P_CENSRD	.59240***	.01829	32.39	.0000	.55655	.62825
LAMBDA	-1.66090***	.35308	-4.70	.0000	-2.35294	-.96887
DELTA	2.37041**	1.16860	2.03	.0425	.07998	4.66083
FIRTSPRT	1.40423**	.64893	2.16	.0305	.13235	2.67611
SECNDPRT	-.81183	.66722	-1.22	.2237	-2.11955	.49590
EFFECT	.59240***	.01829	32.39	.0000	.55655	.62825

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

ME12	1	2	3
1	1837.75	-1062.46	775.290
2	-1510.42	873.222	-637.200
3	-179.353	103.690	-75.6635
4	-57.2711	33.1102	-24.1609
5	138.761	-80.2223	58.5391
6	-131.456	75.9986	-55.4571

E45.9.2 Testing Cragg's Specification of the Tobit Model

A variant of the tobit model which has been used in some studies is that of Cragg (1971), in which the tobit model above applies, but the probability of a nonlimit outcome is determined apart from the level of the nonlimit outcome. Fin and Schmidt (1984) suggest, for example, that the probability of a fire in a building and the amount of the damage when a fire occurs might both depend on the age of the building, but in opposite directions. The tobit model precludes this. The model reduces to the following:

$$\text{Prob}(y^* > 0) = \Phi(\boldsymbol{\gamma}'\mathbf{z}),$$

$$\text{Prob}(y^* \leq 0) = 1 - \Phi(\boldsymbol{\gamma}'\mathbf{z}),$$

if $y^* > 0$, a truncated regression in $\boldsymbol{\beta}'\mathbf{x}$ applies.

This is a combination of the probit model and the truncated regression model. As it stands, the model can be estimated in two parts simply by using a probit model for the indicator of whether y^* is positive or not and a truncated regression model for the nonlimit observations. The tobit model results if it is assumed that $\mathbf{z} = \mathbf{x}$ and $\boldsymbol{\gamma} = \boldsymbol{\beta}$. Given the first, the second is a testable restriction. The log likelihood of the unrestricted model is simply the sum of those of the probit and truncated regressions. This can be compared to the log likelihood for the tobit model, which will be smaller than this sum. An unresolved side issue is that if the first equation does give the probability of a positive observation, then the relationship of the disturbance in the latent regression underlying the probit model to that in the truncated regression is unclear. It is unlikely that they could be independent. In the tobit model, the probit disturbance is $1/\sigma$ times that in the truncated regression. In Cragg's model, the relationship is ambiguous.

The tobit log likelihood function may be written

$$\begin{aligned}\log L &= \Sigma_i\,(1-I_i)\log\text{Prob}(y_i = 0) + I_i\log\text{Prob}(y_i > 0)\\ &\quad + \Sigma_i \log[f(y_i \mid y_i > 0) / \text{Prob}(y_i > 0)],\end{aligned}$$

where $I_i = 1$ if $y_i > 0$, and 0 otherwise.

The first part is the log likelihood function for a probit model. The second line is the log likelihood function for the truncated regression. The reformulation simply adds, then subtracts the second term in the first line. In the tobit model,

$$\text{Prob}(y > 0) = \Phi(\boldsymbol{\beta}'\mathbf{x}/\sigma) = \Phi(\boldsymbol{\gamma}'\mathbf{x})$$

and $$f(y \mid y > 0) = (1/\sigma)\phi(\boldsymbol{\beta}'\mathbf{x}/\sigma).$$

To obtain Cragg's formulation, it is necessary only to release the restriction that $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$. This is testable with a likelihood ratio test just by estimating the three implied models and computing

$$\lambda = 2(\log L_{probit} + \log L_{truncation} - \log L_{tobit}).$$

The following commands carry out the test for the small example in the preceding section. We omit the model results as they are only of secondary interest. Note, in the original data set, *lfp* is the variable corresponding to I_i. The test statistic is 60.77. The critical value with five degrees of freedom is 12.59, so the hypothesis of the tobit model is rejected. The 'p value' is zero.

```
TOBIT       ; Quietly ; Lhs = whrs ; Rhs = x $
CALC        ; ltobit = logl $
PROBIT      ; Quietly ;  Lhs= lfp ; Rhs = x $
CALC        ; lprobit= logl $
TRUNCATE    ; Quietly ; Lhs = whrs ; Rhs = x $
CALC        ; ltrunc = logl
            ; List ; chisq  =  2*(lprobit + ltrunc - ltobit)
            ; 1 - Chi(chisq,kreg) $
```

```
[CALC] CHISQ   =     60.7665209
[CALC] *Result*=       .0000000
Calculator: Computed   3 scalar results
```

There is a disadvantage to the preceding method. The approach requires estimation of all three models. Another approach which in the end may be simpler to use is the LM test devised by Fin and Schmidt (1984). This requires estimation only of the tobit (i.e., restricted) model. To set this up, we reparameterize the model as follows, incorporating Olsen's transformation at the outset:

Limit probability = Prob($y^* \leq 0$) = $1 - \Phi(\boldsymbol{\delta}'\mathbf{x})$

Density for the nonlimit observation = $\eta\phi(\eta y - \boldsymbol{\gamma}'\mathbf{x}) / \Phi(\boldsymbol{\gamma}'\mathbf{x})$

This is Cragg's model as it stands. Now, let $\boldsymbol{\theta}$ be a free parameter vector and let

$$\boldsymbol{\delta} = \boldsymbol{\gamma} + \boldsymbol{\theta}.$$

Insert this in the definition above, to obtain the same model:

Limit probability = Prob($y^* \leq 0$) = $1 - \Phi[(\boldsymbol{\gamma} + \boldsymbol{\theta})'\mathbf{x}]$

Truncated normal density for the nonlimit observation = $\eta\phi(\eta y - \boldsymbol{\gamma}'\mathbf{x}) / \Phi(\boldsymbol{\gamma}'\mathbf{x})$

As before, let $I_i = 1$ indicate a nonlimit observation, and let $I_i = 0$ indicate a limit observation. Then the log likelihood for this model is that for a tobit model;

$$\log L = \Sigma_i (1-I_i)\log\{1 - \Phi[(\boldsymbol{\gamma} + \boldsymbol{\theta})'\mathbf{x}_i]\} + I_i\log\{\eta\phi(\eta y_i - \boldsymbol{\gamma}'\mathbf{x}_i) / \Phi(\boldsymbol{\gamma}'\mathbf{x}_i)\}$$

Cragg's model results if $\boldsymbol{\theta}$ is a free parameter vector while the tobit model is obtained by the restriction $\boldsymbol{\theta} = \mathbf{0}$. Given this simple formulation, Fin and Schmidt devised a Lagrange multiplier statistic to test the restriction. Let the function $\lambda(z) = \phi(z)/\Phi(z) = \lambda_1$. Let $\lambda(-z) = \lambda_0$. In the following, z will equal $\boldsymbol{\gamma}'\mathbf{x}$, and for convenience, let Φ denote the CDF, $\Phi(z)$. Then, Fin and Schmidt's LM statistic is computed as follows: Let

$$a_i = \lambda_{0i}\lambda_{1i}$$

$$b_i = \Phi_i \times (1 - z_i\lambda_{1i} - \lambda_1^{\,2})$$

$$c_i = \Phi_i \times (z_i + \lambda_{1i})/\eta$$

$$d_i = \Phi_i \times (2 + z_i^{\,2} + z_i\lambda_{1i})$$

$$e_i = \eta y_i - \boldsymbol{\gamma}'\mathbf{x}_i$$

Let $\mathbf{A}$ and $\mathbf{B}$, denote $n \times n$ diagonal matrices formed from these quantities and let $\mathbf{c}$ and $\mathbf{d}$ denote $n \times 1$ vectors formed from c_i and d_i. Now, denote the full $n \times K$ data matrix as $\mathbf{X}$. Then,

$$\mathbf{g} = \Sigma_{nonlimit\ observations}\ (\lambda_{1i} - e_i)\mathbf{x}_i$$

$$\mathbf{H} = (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} - [\mathbf{X}'\mathbf{B}\mathbf{X} - (\mathbf{X}'\mathbf{c})(\mathbf{c}'\mathbf{X})/\mathbf{d}'\mathbf{d}]^{-1}$$

and

$$\text{LM} = \mathbf{g}'\mathbf{H}\mathbf{g}.$$

This LM statistic is computed automatically by *LIMDEP* when it fits a tobit model, and is reported with the standard output. Here is the diagnostic table from the model estimated in the example above:

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                      Y
Estimation criterion          -3903.79391
Estimation based on N =    753, K =    7
Inf.Cr.AIC  =    7821.6 AIC/N =   10.387
Threshold values for the model:
Lower=      .0000     Upper=+infinity
LM test [df] for tobit=      32.508[  6]
Normality Test, LM    =      10.378[  2]
ANOVA  based fit measure =    .048112
DECOMP based fit measure =    .164940
--------+--------------------------------------------------------------------
```

The LM statistic is shown above the fit measures – for this model and data set, the value is 32.508, with six degrees of freedom. The critical value, as noted earlier is 12.59, so the hypothesis of the model would be rejected.

E45.9.3 Testing for Nonnormality

Tests of the normality assumption in the censored regression model have been developed by many authors. Generally, they are based on the same approach as is used in the linear regression model; the skewness (third moment) and kurtosis (fourth moment) of the 'residuals' are compared to what would be expected if the underlying distribution were normal (zero and three, respectively). The obstacle in this setting is that it is difficult to obtain a 'clean' set of residuals. For the censored regression model, the obvious choice, $y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ has an equally obvious defect. The residuals for the limit observations are clearly not going to conform to a normal distribution even if the model is correct, by construction.

Pagan and Vella's Conditional Moment Test

Pagan and Vella (1989) devised a conditional moment test for normality in the censored regression model. The test is based on the moment restrictions:

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

$$E_y[E[\varepsilon_i^3 \mid y_i]] = 0$$

$$E_y[E[\varepsilon_i^4 \mid y_i] - 3\sigma^4] = 0.$$

Note that these are not the moments only of the truncated distribution, since they are averaged over the distribution of the observed y. For observations for which $y_i > L_i$, we observe y_i^*, so the moments are the familiar ones. For the limit observations, we require the moments of the truncated distribution. Pagan and Vella provide the following useful result (adapted for current purposes): Denoting $E[\varepsilon_i^j \mid y_i = L_i]$ as μ_j, their result is

$$\mu_j = (j\text{-}1)\sigma^2\mu_{j-2} + \alpha_i^{j-1}\sigma^j\lambda_i(\alpha_i)$$

where, as usual, $\alpha_i = (L_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma$, $\lambda_i(\alpha_i) = -\phi(\alpha_i)/\Phi(\alpha_i) = \lambda_i$, and, the initial values for the recursion are $\mu_{-1} = 0$ and $\mu_0 = 1$. Collecting terms, we have

	$y = L_i$	$y > L_i$
$E[\varepsilon_i \mid y_i]$	$\sigma[\lambda_i]$	0
$E[\varepsilon_i^2 \mid y_i]$	$\sigma^2[1 + \lambda_i\alpha_i]$	σ^2
$E[\varepsilon_i^3 \mid y_i]$	$\sigma^3[\lambda_i(2 + \alpha_i^2)]$	0
$E[\varepsilon_i^4 \mid y_i]$	$\sigma^4[3 + \lambda_i(3\alpha_i + \alpha_i^3)]$	$3\sigma^4$

To carry out the conditional moment test, we first compute

$$m_{i1} = E[\varepsilon_i^3 \mid y_i] \text{ and } m_{i2} = E[\varepsilon_i^4 \mid y_i] - 3\sigma^4$$

using the appropriate form from above, the sample data, and the maximum likelihood estimates of the parameters. Let the $n\times 2$ matrix, **M**, be constructed from these two columns of functions. Let the $n\times(K+1)$ matrix, **G**, contain the derivatives of the log likelihood function for the K elements in $\boldsymbol{\beta}'$ followed by that for σ. The transpose of the ith row of **G** is

$$\mathbf{g}_i = \frac{1}{\sigma}\left\{ d_i \begin{bmatrix} \left(\dfrac{\varepsilon_i}{\sigma}\right)\mathbf{x}_i \\ \left(\dfrac{\varepsilon_i}{\sigma}\right)^2 - 1 \end{bmatrix} + (1-d_i)\begin{bmatrix} \lambda_i \mathbf{x}_i \\ \lambda_i \alpha_i \end{bmatrix} \right\}$$

Then, **G** is computed with the sample data and the maximum likelihood estimates of the parameters. Finally, let **i** denote an $n\times 1$ column vector of ones. Then, the conditional moment statistic for this test is

$$\chi^2[2] = \mathbf{i}'\mathbf{M}[\mathbf{M}'\mathbf{M} - \mathbf{M}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{M}]^{-1}\mathbf{M}'\mathbf{i}.$$

Under the hypothesis of normality, the statistic has a limiting chi squared distribution with two degrees of freedom.

The program below gives a general set of computations. For a particular application, it is necessary to provide the definitions of *x*, *y*, and *li*.

```
NAMELIST   ; x = the Rhs variables in the regression $
CALC       ; li = the lower censoring value (usually zero) $
CREATE     ; y = the Lhs variable in the model $
TOBIT      ; Lhs = y ; Rhs = x ; Limit = li $
CREATE     ; d = y > li ; d0 = 1 - d ; xb = x'b ; e = y - xb
           ; alpha = (li - xb)/s  ; lambda = -N01(alpha)/(1-Phi(alpha))
           ; m1 = d*e^3 + d0*s^3*lambda*(2+alpha^2)
           ; m2 = (d*e^4 + d0*s^4*(3+lambda*(3*alpha+alpha^3))) - 3*s^4
           ; qx = (d*e/s + d0*lambda)/s
           ; qs = (d*((e/s)^2-1) + d0*lambda*alpha)/s $
NAMELIST   ; m = m1,m2  $
MATRIX     ; dd = Bhhh(x,one,qx,qs)
           ; mdb= m'[qx]x ; mds = m'qs ; md = [mdb, mds]
           ; v = m'm - md * <dd> * md'
           ; List ; cmtest = 1'm * <v> * m'1 $
CALC       ; List ; pvalue = 1 - Chi(cmtest,2) $
```

```
Matrix CMTEST   has  1 rows and  1 columns.
            1
       +--------------
      1|   30.16118
```

Applying the foregoing to our labor supply application produces a value of 30.16118. The critical chi squared value for two degrees of freedom is 5.99, so the hypothesis of normality is rejected based on this test.

Chesher and Irish's Generalized Residuals Test

The generalized residual based test devised by Chesher and Irish (1987) is based on a similar logic. The generalized residuals are computed as the derivatives of the log likelihood with respect to the constant term in the model. The Chesher and Irish test is based on the hypothesis that these will behave as if drawn from a normal distribution, the same as in the Pagan and Vella test. The computations are as follows, where d_{1i} denotes a nonlimit observation and $d_{0i} = 1 - d_{1i}$.

$$\varepsilon_I = y_i - \mathbf{x}_i'\boldsymbol{\beta}$$

$$a_i = [(y_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma]$$

$$\alpha_i = [(l_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma] \text{ where } l_i \text{ is the lower censoring limit, usually zero}$$

or

$$\alpha_i = [(\mathbf{x}_i'\boldsymbol{\beta} - u_i)/\sigma] \text{ where } u_i \text{ is the upper censoring limit, usually zero}$$

$$\lambda_i = \phi(\alpha_i)/\Phi(\alpha_i) \text{ where } \phi(.) = \text{normal PDF}, \Phi(.) = \text{normal CDF (lower)}$$

or

$$\lambda_i = -\phi(\alpha_i)/\Phi(\alpha_i) \text{ where } \phi(.) = \text{normal PDF}, \Phi(.) = \text{normal CDF (upper)}$$

$$e_{1i} = -d_{0i}\,\lambda_i + d_{1i}\,a_i$$

$$e_{2i} = -d_{0i}\,\alpha_i\lambda_i + d_{1i}\,(a_i^2 - 1)$$

$$e_{3i} = -d_{0i}\,(2+\alpha_i^2)\lambda_i + d_{1i}\,a_i^3$$

$$e_{4i} = -d_{0i}\,(3\,\alpha_i + \alpha_i^3)\lambda_i + d_{1i}\,(a_i^4 - 3)$$

$$\mathbf{c}_i = [\mathrm{e}_{1i}\mathbf{x}_i', \mathrm{e}_{2i}, \mathrm{e}_{3i}, \mathrm{e}_{4i}]'$$

Define the $n\times(K+3)$ matrix $\mathbf{C}$ so that its $\mathbf{i}$th row is $\mathbf{c}_i'$ and as before, let $\mathbf{i}$ be an $n\times1$ column of ones. Then, the test statistic is

$$\chi^2[2] = \mathbf{i}'\mathbf{C}\,(\mathbf{C}'\mathbf{C})^{-1}\mathbf{D}'\mathbf{i}$$

This would be straightforward to program in similar fashion to the Pagan and Vella test. It is reported automatically with the tobit model output, as shown below

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                    Y
Estimation criterion         -3903.79391
Estimation based on N =    753, K =    7
Inf.Cr.AIC  =    7821.6 AIC/N =   10.387
Threshold values for the model:
Lower=     .0000     Upper=+infinity
LM test [df] for tobit=      32.508[  6]
Normality Test, LM     =     10.378[  2]
ANOVA  based fit measure =    .048112
DECOMP based fit measure =    .164940
--------+--------------------------------------------------------------------
```

E45.9.4 Generalized Residuals

The conditional moment test suggested by Pagan and Vella (1989) is one of a class of tests devised by Chesher and Irish (1987) based on what they label generalized residuals. Their specification tests are based on the standard approach to tests of specification in the regression model. Tests for omitted variables are based on covariances of residuals from the model with the omitted variables in question; tests of heteroscedasticity are based on covariances of squares of residuals with hypothesized exogenous variables; tests of normality are based on the means and variances of third and fourth moments of residuals, and so on. The problem with the extension of these methods to censored data models is that the residuals in the censored data are ill defined. Chesher and Irish defined the generalized residuals for these purposes. For censored data models based on an 'index function,' $\mathbf{x}_i'\boldsymbol{\beta}$ (which includes most of the cases we have examined) the generalized residuals for the ith observation are

$$e(1) = \frac{\partial \log f_i}{\partial \beta_1} \text{ and } e(2) = \frac{\partial \log f_i}{\partial \sigma}$$

where $\log f_i$ is the term for the ith observation in the log likelihood for the model and β_1 is the constant term in the regression. (Chesher and Irish do their analysis in terms of σ^2, but to maintain consistency with our earlier results, we will modify their formulations.) Their testing procedures extend to more general censored regression models, including the categorical data model we examined earlier, so we will consider the extension to that model as well. For an observation which is not censored,

$$e(1) = \frac{1}{\sigma}\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) \text{ and } e(2) = \frac{1}{\sigma}\left[\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)^2 - 1\right].$$

For an observation which falls in the open region $(Lower_i, Upper_i]$,

$$f_i = \Phi(\alpha_i^{Upper}) - \Phi(\alpha_i^{Lower}),$$

with $\alpha_i^{Lower} = (Lower_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma$ and $\alpha_i^{Upper} = (Upper_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma$,

so $$e(1) = -\frac{1}{\sigma}\left(\frac{\phi(\alpha_i^{Upper}) - \phi(\alpha_i^{Lower})}{\boldsymbol{\Phi}(\alpha_i^{Upper}) - \boldsymbol{\Phi}(\alpha_i^{Lower})}\right) = \frac{1}{\sigma}\lambda(\alpha_i^{Lower}, \alpha_i^{Upper}).$$

Notice we have extended the definition of the $\lambda(\bullet)$ function for this purpose. For our basic censored regression model, $U_i = \infty$, $\Phi(\alpha_i^U) = 1$, $\phi(\alpha_i^U) = 0$, and the results we used earlier emerge. This definition of a 'residual' makes sense. For the uncensored observation, $e(1) = (1/\sigma^2)\{E[y_i^*|y_i, \mathbf{x}_i] - \mathbf{x}_i'\boldsymbol{\beta}\}$, since when an observation is uncensored, $y_i = y_i^*$. For the censored regression with simple lower censoring at L_i, $Lower = -\infty$, $Upper = L_i$, – we'll maintain this for the moment, and extend it to the more general case presently – and

$$e(1) = -\frac{1}{\sigma}\left(\frac{\phi(\alpha_i^L) - 0}{\boldsymbol{\Phi}(\alpha_i^L) - 0}\right) = \frac{1}{\sigma}\lambda(\alpha_i^L).$$

As we noted earlier,

$$E[y_i^*|y_i, \mathbf{x}_i] = E[y_i^*|y_i^* \le L_i, \mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\left(\frac{-\phi\left(\alpha_i^L\right)}{\Phi\left(\alpha_i^L\right)}\right),$$

so, once again, $e(1) = (1/\sigma^2)\{E[y_i^*|y_i, \mathbf{x}_i] - \mathbf{x}_i'\boldsymbol{\beta}\}$.

For uncensored observations, the other generalized residual is

$$\begin{aligned} e(2) &= \partial \log f_i/\partial\sigma = (1/\sigma^3)[(y_i^* - \mathbf{x}_i'\boldsymbol{\beta})^2 - \sigma^2] \\ &= (1/\sigma^3)\{[E[y_i^*|y_i, \mathbf{x}_i] - \mathbf{x}_i'\boldsymbol{\beta}]^2 - \sigma^2\}, \end{aligned}$$

since, once again, if the observation is uncensored, $y_i = y_i^*$. For a censored observation,

$$e(2) = \partial \log f_i/\partial\sigma = (1/\sigma)\alpha_i\lambda_i.$$

The previous expression produced the deviation of a square from its expectation, so the interpretation as a residual makes intuitive sense. In fact, this one does also. For the censored observation,

$$\begin{aligned} [E[y_i^*|y_i, \mathbf{x}_i] - \mathbf{x}_i'\boldsymbol{\beta}]^2 &= \mathrm{E}[\varepsilon_i^2|\varepsilon_i \le L_i - \mathbf{x}_i'\boldsymbol{\beta}] \\ &= \mathrm{Var}[\varepsilon_i|\varepsilon_i \le L_i - \mathbf{x}_i'\boldsymbol{\beta}] + \{\mathrm{E}[\varepsilon_i\,|\varepsilon_i \le L_i - \mathbf{x}_i'\boldsymbol{\beta}]\}^2 \\ &= \sigma^2[1 - \delta(\alpha_i)] + [\sigma\lambda(\alpha_i)]^2. \\ &= \sigma^2 + \sigma^2\alpha_i\lambda(\alpha_i). \end{aligned}$$

Therefore, $[E[y_i^*|y_i, \mathbf{x}_i] - \mathbf{x}_i'\boldsymbol{\beta}]^2 - \sigma^2 = \sigma^2\alpha_i\lambda(\alpha_i)$

which gives $e(2)$ the same interpretation for the censored observations.

Chesher and Irish provide the general expression for $e(2)$ in the categorical (grouped) data model.

$$e(2) = -\frac{1}{\sigma}\left(\frac{\alpha_i^{Upper}\phi\left(\alpha_i^{Upper}\right) - \alpha_i^{Lower}\phi\left(\alpha_i^{Lower}\right)}{\Phi\left(\alpha_i^{Upper}\right) - \Phi\left(\alpha_i^{Lower}\right)}\right)$$

as well as new expressions $e(3)$ and $e(4)$ which enter the computations of tests for normality. The latter two, with their testing procedure produce exactly the Pagan and Vella approach laid out earlier, so we'll not repeat them here. Chesher and Irish suggest several specification tests using their generalized residuals. The general approach, based on maximum likelihood estimates of the parameters and the BHHH estimator of the asymptotic covariance matrix for the estimators, uses the statistic

$$\chi^2[J] = \mathbf{i}'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{i}$$

where $\mathbf{R}$ is an $n\times(K+1+J)$ matrix of constructed observations. Each row of $\mathbf{R}$ consists of, first K+1 elements which are the derivatives of log f_i with respect to $\boldsymbol{\beta}$ and σ, followed by J elements which are the products of variables which are expected to be orthogonal to the generalized residuals. These first K+1 columns of $\mathbf{R}$ are the terms in the gradient of the log likelihood function, so for the first K+1 elements, at the maximum likelihood estimates, $\mathbf{i}'\mathbf{R} = \mathbf{0}$.

They suggest the following tests:

1. for J omitted variables, $z_1,..., J$ trailing elements in $\mathbf{R}_i$ are $e(1)_i\mathbf{z}_i'$,
2. for heteroscedasticity of a known functional form $\sigma_i^2 = \sigma^2 h(\mathbf{z}_i,\boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ is a $J\times1$ parameter vector such that $h(\mathbf{z}_i,\mathbf{0}) = 1$, J trailing elements are $e_i(2)\ \partial h(\mathbf{z}_i,\boldsymbol{\kappa})/\partial\boldsymbol{\kappa}|\boldsymbol{\kappa}=\mathbf{0}$,
3. for heteroscedasticity of unknown form, J trailing elements are the J unique terms in $e_i(2)$ $\mathbf{x}_i\otimes\mathbf{x}_i$, not including the constant term.

They also suggest a test for random parameter variation which is identical to test 3 save for the addition of $e(3)$ and $e(4)$ – see the earlier discussion of Pagan and Vella's test.

To operationalize this procedure, define as $\mathbf{g}_i'$ the first K+1 elements of the row in $\mathbf{R}$. As we observed, these are the elements of the derivatives of the log likelihood function. Denote the trailing J elements as $\mathbf{m}_i'$. If we arrange these in two sets of columns, then the matrix $\mathbf{R}$ becomes $[\mathbf{G},\mathbf{M}]$. With this in place, the test statistic becomes

$$\chi^2[J] \ = \ \begin{bmatrix}\mathbf{i'G} & \mathbf{i'M}\end{bmatrix}\begin{bmatrix}\mathbf{G'G} & \mathbf{G'M}\\ \mathbf{M'G} & \mathbf{M'M}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{G'i}\\ \mathbf{M'i}\end{bmatrix}.$$

Recall, $\mathbf{G'i}$ is the derivative vector (gradient) of the log likelihood function, which equals zero at the maximum likelihood estimates. Using the partitioned inverse formula, then, we reduce the statistic to

$$\chi^2[\text{J}] \ = \ \mathbf{i'M}[\mathbf{M'M} - \mathbf{M'G}(\mathbf{G'G})^{-1}\mathbf{G'M}]^{-1}\mathbf{M'i}$$

(which is identical to the Pagan and Vella statistic, as noted). Carrying out the test, therefore, requires computation of the moments in **M** and the derivatives in **G**, and a bit of matrix algebra. Chesher and Irish's results are quite convenient, and suggest a general strategy for a variety of other specification tests in the censored regression model.

Generalized residuals are computed automatically by the **TOBIT** (and **TRUNCATION** and **GROUPED**) estimators. After estimation, the variables *genres_1* and *genres_2* will contain the two generalized residuals noted above.

E45.9.5 Censoring with Unknown Censoring Limits

We have allowed for either lower or upper censoring based on known censoring points. It might be natural to ask if the censoring point could, itself, be unknown. One might know whether or not an observation had been censored, but the exact value of the censoring threshold might not be known. Can the censoring limit be treated as an unknown parameter, and estimated with the rest of the model? Unfortunately, no, as it is simple to show. Consider the log likelihood for the model with lower censoring that we examined earlier, and suppose that $\tilde{L}$ is an unknown parameter to be estimated. Then,

$$\partial\log L/\partial\tilde{L} \ = \ \phi/\Phi\times\theta \ > 0.$$

This expression is always positive. The only way to equate it to zero is to set θ to zero (infinite variance, σ^2, or to have $\phi(.)$ go to zero, which will happen if any coefficient in the model diverges. The upshot is that for identification in the censoring model, the censoring threshold must be known. It is not estimable.

E45.10 Powell's Symmetrically Censored LS Estimator

As noted earlier, a large body of research has focused on semiparametric alternatives to the normal based censored regression estimator. Powell's (1986) symmetrically censored least squares estimator is a simple one that can be implemented as a short procedure. The procedure and an application are as follows:

```
?==============================================================
? Powell's symmetrically censored least squares estimator
?==============================================================
? This is the only part of the estimator that is specific to the problem.
? Define the list of regressors and the dependent variable.
NAMELIST    ; x = one,wa,we,hhrs,ha,he,kl6,k618,ww,cit,ax $
CREATE      ; y = whrs $
? Use the tobit MLE as starting values for beta.
TOBIT       ; Quietly ; Lhs = y ; Rhs = x $
MATRIX      ; bj = b ; btobit = b ; vtobit = varb $
CALC        ; deltab = 1 $  Start delta large enough to begin.
PROCEDURE $   This procedure computes the scls estimator iteratively
CREATE      ; bx = x'bj ; bx2 = 2*bx ; ts = bx > 0 ; ys = Min(y,bx2) $
MATRIX      ; hj = <x'[ts]x> ; bj1 = hj  * x'[ts]ys ; db = bj1-bj $
? We check convergence using a scale free measure rather than db.
CALC        ; list(exec) ; deltab = Qfr(db,hj) $
MATRIX      ; bj = bj1 $
ENDPROC $
EXECUTE     ; While deltab > .00001 $
? Estimation is finished. Get covariance matrix and display results.
CREATE      ; vs = (y > 0)*(y < bx2) ; u2 = ts*(ys-bx)^2 $
MATRIX      ; c = x'[vs]x ; d = x'[u2]x  ; v = <c>*d*<c> $
DISPLAY     ; Labels = x ; Parameters = bj  ; Covariance = v
            ; Title = Symmetrically Censored Least Squares $
DISPLAY     ; Labels = x ; Parameters = btobit ; Covariance = vtobit
            ; Title = Maximum Likelihood Tobit Estimates $
?=============================================================
```

```
[CALC:Iteration=0001] DELTAB  =  11681.0721316
[CALC:Iteration=0001] DELTAB  =   1909.8563458
[CALC:Iteration=0001] DELTAB  =      8.4412047
[CALC:Iteration=0001] DELTAB  =     28.2620147
[CALC:Iteration=0001] DELTAB  =     26.4144440
[CALC:Iteration=0001] DELTAB  =     11.8704001
[CALC:Iteration=0001] DELTAB  =      4.7102712
 (Values omitted)
[CALC:Iteration=0001] DELTAB  =       .0000500
[CALC:Iteration=0001] DELTAB  =       .0000207
[CALC:Iteration=0001] DELTAB  =       .0000087
```

```
--------------------------------------------------------------------------------
Symmetrically Censored Least Squares
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
Constant|    2307.33***    501.4397     4.60  .0000     1324.53    3290.13
      WA|   -42.2714***    15.48364    -2.73  .0063    -72.6188   -11.9240
      WE|    25.3684       28.43786      .89  .3724    -30.3688    81.1056
    HHRS|    -.03411         .09091     -.38  .7075     -.21230     .14408
      HA|   -2.85768       12.46911     -.23  .8187   -27.29668   21.58133
      HE|   -40.9069**     20.12293    -2.03  .0421    -80.3472    -1.4667
     KL6|   -681.440***    166.7827    -4.09  .0000   -1008.328   -354.552
    K618|   -79.0379**     39.77414    -1.99  .0469   -156.9938    -1.0820
      WW|    76.7658***    26.65889     2.88  .0040     24.5153   129.0162
     CIT|    57.9004       99.87113      .58  .5621   -137.8434   253.6443
      AX|    46.8506***     7.20154     6.51  .0000     32.7359    60.9654
--------+-----------------------------------------------------------------------
Maximum Likelihood Tobit Estimates
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
Constant|    2037.32***    458.1890     4.45  .0000     1139.28    2935.35
      WA|   -45.4782***    12.47314    -3.65  .0003    -69.9252   -21.0313
      WE|    9.84812       25.11007      .39  .6949   -39.36671   59.06295
    HHRS|    -.07035         .07535     -.93  .3505     -.21804     .07734
      HA|   -7.69527       11.91816     -.65  .5185   -31.05443   15.66389
      HE|   -17.8252       18.51562     -.96  .3357    -54.1152    18.4647
     KL6|   -767.883***    109.1351    -7.04  .0000    -981.783   -553.982
    K618|   -21.6712       36.93462     -.59  .5574    -94.0618    50.7193
      WW|    157.123***    14.41339    10.90  .0000     128.873    185.373
     CIT|   -67.3054       94.64545     -.71  .4770   -252.8070   118.1963
      AX|    63.1848***     6.15842    10.26  .0000     51.1146    75.2551
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E45.11 Double Hurdle Model for Censored Regression

The double hurdle model is a generalization of Cragg's specification in Section E45.9.2. The rather lengthy development in the literature is summarized in Yen and Jones (1997), which is used as the background for the implementation here. The model consists of a binary choice (probit) participation equation,

$$d^* = \boldsymbol{\alpha}'\mathbf{z} + v,\ v \sim \mathrm{N}[0,1],$$

$$d \;\; = 1(d^* > 0),$$

a latent intensity equation,

$$y^* \;=\; \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim \mathrm{N}[0,\sigma],$$

and an observation mechanism that extends the tobit model,

$$y \;=\; y^* \text{ if } d^* > 0 \text{ and } y^* > 0 \text{ and } y = 0 \text{ otherwise.}$$

(The tobit model implicitly assumes that $y^* > 0$ implies $d = 1$ by construction.) Yen and Jones motivate the specification based on separate purchase decision and frequency of purchase decisions. Our implementation builds on the basic model and adds three extensions (all of which were suggested by Yen and Jones). We note in Section E45.11.5 correction of a substantive error in the development in the Yen and Jones presentation.

The model is an extension of the tobit model. The basic command for estimation of the double hurdle model is

TOBIT ; Lhs = dependent variable
; Rhs = independent variables
; Hurdle $

Remaining options are the same as for the tobit model with the exception that censoring is always at zero from below. If your data are censored at some other point, C or C_i, then simply create the new variable y_i - C_i to conform to the assumption. If your data are censored from above, then create the new variable C_i - y_i, then reverse the signs of the estimates of $\boldsymbol{\beta}$ after estimation. The basic specification also assumes that the variables in the participation equation are the same as in the intensity equation. This assumption is relaxed immediately in the next section. Section E45.11.2 extends the model to allow correlation between the intensity and participation equations. Section E45.11.3 provides a transformation somewhat similar to the Box-Cox model that allows for different functional forms, e.g., linear vs. logarithmic and forms in between.

E45.11.1 Basic Model with Heteroscedasticity

The central specification of the double hurdle model provides for different variables in the two equations. The specification to accommodate the general model is

TOBIT ; Lhs = dependent variable
; Rhs = independent variables in intensity equation
; Hurdle = independent variables in participation equation $

If the variables in the two equations are the same, then use **; Hurdle** without a list in the command, as shown in the introduction to this section. A second extension is to allow heteroscedasticity in the variance in the tobit equation. The model is then

$$d^* = \boldsymbol{\alpha}'\mathbf{z} + v,\ v \sim N[0,1]$$

$$d = 1(d^* > 0)$$

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim N[0,\sigma_i],\ \sigma_i = \sigma \times \exp(\boldsymbol{\delta}'\mathbf{h}_i)$$

$$y = y^* \text{ if } d^* > 0 \text{ and } y^* > 0 \text{ and } y = 0 \text{ otherwise.}$$

The original double hurdle model returns if $\boldsymbol{\delta} = \mathbf{0}$. The extended model command is

TOBIT ; Lhs = dependent variable
; Rhs = independent variables in intensity equation
; Hurdle = independent variables in participation equation
; Hfn = variables in variance ; Heteroscedasticity $

Note that consistent with the model equation above, the Hfn list should not contain *one*. As always, **; Heteroscedasticity** may be abbreviated to **; Het**.

E45.11.2 Endogenous Participation

The participation becomes endogenous if v and ε are correlated. This produces a relatively large extension of the model – given the formulation, the extension is easily tested. The model specification becomes

$$d^* = \boldsymbol{\alpha}'\mathbf{z} + v,\ v \sim N[0,1]$$

$$d = 1(d^* > 0)$$

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim N[0,\sigma_i],\ \sigma_i = \sigma \times \exp(\boldsymbol{\delta}'\mathbf{h}_i)$$

$$y = y^* \text{ if } d^* > 0 \text{ and } y^* > 0 \text{ and } y = 0 \text{ otherwise,}$$

$$\begin{pmatrix} v \\ \varepsilon \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_i \\ \rho\sigma_i & \sigma_i^2 \end{pmatrix}\right].$$

(Yen and Jones specified that the off diagonal element in the covariance matrix was constant. The (invalid) assumption is substantive, and would affect the estimation results.)

The endogeneity assumption is added with the command

```
TOBIT    ; Lhs = dependent variable
         ; Rhs = independent variables in intensity equation
         ; Hurdle = independent variables in participation equation
         ; Hfn = variables in variance ; Het
         ; Correlation $
```

Note that the list of variables in the hurdle equation and the heteroscedasticity specification are both optional. The simplest model with endogenous participation results with

```
TOBIT    ; Lhs = dependent variable
         ; Rhs = independent variables in intensity equation
         ; Hurdle
         ; Correlation $
```

Endogeneity and different participation and intensity equations or endogeneity and heteroscedasticity are specified with

```
TOBIT    ; Lhs = dependent variable
         ; Rhs = independent variables in intensity equation
         ; Hurdle = independent variables in participation equation
         ; Correlation $
```

and

```
TOBIT    ; Lhs = dependent variable
         ; Rhs = independent variables in intensity equation
         ; Hurdle
         ; Hfn = variables in variance ; Het
         ; Correlation $
```

E45.11.3 Inverse Hyperbolic Sine Transformation

The inverse hyperbolic sine transformation is suggested for this model as a device to extend the functional form beyond (or between) the usual levels or logarithms choices. The transformation is

$$T(y,\gamma) = \frac{\log\left[\gamma y + \left(1+\gamma^2 y^2\right)^{1/2}\right]}{\gamma}$$

where γ is the crucial new parameter that extends the model. The transformation approaches linearity ($T(y,\gamma) = y$) as γ approaches zero and approaches the log function as γ increases. The transformation is incorporated in the model with

$$d^* = \boldsymbol{\alpha}'\mathbf{z} + v,\ v \sim N[0,1]$$

$$d = 1(d^* > 0)$$

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim N[0,\sigma_i],\ \sigma_i = \sigma \times \exp(\boldsymbol{\delta}'\mathbf{h}_i)$$

$$T(y,\gamma) = y^* \text{ if } d^* > 0 \text{ and } y^* > 0 \text{ and } T(y,\gamma) = 0 \text{ otherwise,}$$

$$\begin{pmatrix} v \\ \varepsilon \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_i \\ \rho\sigma_i & \sigma_i^2 \end{pmatrix}\right].$$

To use the transformation, the command is modified to

```
TOBIT       ; Lhs = dependent variable
            ; Rhs = independent variables in intensity equation
            ; Hurdle = independent variables in participation equation
            ; Hfn = variables in variance ; Het
            ; Correlation
            ; Model = IHS $
```

All of the model permutations noted earlier may be used with this specification.

E45.11.4 Application

To illustrate the hurdle model, we have manipulated the income variable in the health care data set. The dependent variable is income = $\max\left(0, hhninc - \overline{hhninc}\right)$. The model is fit with the pooled data set – there is no panel data version of this specification.

```
SAMPLE      ; All $
CREATE      ; income = hhninc-xbr(hhninc) $
CREATE      ; income = max(0,income) $
NAMELIST    ; x = one,age,educ,hsat,married,hhkids $
NAMELIST    ; z = one,age,educ $
NAMELIST    ; h = female,married,age $
TOBIT       ; Lhs = income ; Rhs = x ; Hurdle = z ; Correlated ; Partial Effects
            ; Het ; Hfn = h ; Model = IHS ; Maxit = 15 $
```

Average partial effects are computed for three outcomes, Prob(y > 0), E[y] and E[y|y>0]. The partial effects are also computed for the three sets of variables in the model, **x** in the regression, **h** in the disturbance variance (heteroscedasticity) and **z** in the hurdle equation.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable               INCOME
Log likelihood function     -7592.81080
Estimation based on N =  27326, K =  15
Inf.Cr.AIC  =  15215.6 AIC/N =     .557
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    -.16284***      .04393    -3.71  .0002     -.24894   -.07674
     AGE|    -.00558***      .00053   -10.63  .0000     -.00661   -.00455
    EDUC|     .02512***      .00093    27.10  .0000      .02330    .02694
    HSAT|     .00243***      .00072     3.37  .0008      .00102    .00384
 MARRIED|     .16178***      .01146    14.11  .0000      .13932    .18425
  HHKIDS|    -.06916***      .00383   -18.05  .0000     -.07667   -.06165
        |Disturbance standard deviation
   Sigma|     .14803***      .01101    13.45  .0000      .12646    .16960
        |Heteroscedasticity terms in disturbance variance
  FEMALE|     .07770***      .01121     6.93  .0000      .05574    .09967
 MARRIED|    -.60960***      .02258   -26.99  .0000     -.65386   -.56534
     AGE|     .01717***      .00139    12.34  .0000      .01444    .01990
        |Hurdle Equation for IHS/Hurdle Model
Constant|   -2.13278***      .19219   -11.10  .0000    -2.50946  -1.75610
     AGE|     .07989***      .00629    12.70  .0000      .06756    .09222
    EDUC|     .00142         .00970      .15  .8834     -.01759    .02044
        |Correlation Between Hurdle and Latent Regression
Rho(u,e)|    -.30528**       .14504    -2.10  .0353     -.58956   -.02101
        |Parameter for inverse hyperbolic sine
   Gamma|    2.67679***      .09189    29.13  .0000     2.49669   2.85689
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
Average Partial Effects in Inverse Hyperbolic Sine Double Hurdle Model
Partial effects are added for variables common to components.
-----------------------------------------------------------------------
Average Partial Effects of Probability of Positive Outcome
-----------------------------------------------------------------------
Variable    Regression     Hurdle     Variance  Total Effect   Std.Error
-----------------------------------------------------------------------
     AGE    -.008669      .009750      .000141      .001223      .001394
    EDUC     .038992      .000174      .000000      .039166      .001198***
    HSAT     .003772      .000000      .000000      .003772      .001120***
 MARRIED     .251115      .000000     -.005008      .246107      .017772***
  HHKIDS    -.107350      .000000      .000000     -.107350      .005948***
  FEMALE     .000000      .000000      .000638      .000638      .000092***
-----------------------------------------------------------------------
```

```
--------------------------------------------------------------------
Average Partial Effects in Inverse Hyperbolic Sine Double Hurdle Model
Partial effects are added for variables common to components.
--------------------------------------------------------------------
Average Partial Effects of Unconditional Expected Value
--------------------------------------------------------------------
Variable    Regression    Hurdle     Variance  Total Effect   Std.Error
--------------------------------------------------------------------
     AGE    -.002727     .001932    -.000692    -.001488     .000331***
    EDUC     .012268     .000034     .000000     .012302     .000365***
    HSAT     .001187     .000000     .000000     .001187     .000352***
 MARRIED     .079008     .000000     .024568     .103576     .005774***
  HHKIDS    -.033775     .000000     .000000    -.033775     .001871***
  FEMALE     .000000     .000000    -.003132    -.003132     .000452***
--------------------------------------------------------------------

--------------------------------------------------------------------
Average Partial Effects in Inverse Hyperbolic Sine Double Hurdle Model
Partial effects are added for variables common to components.
--------------------------------------------------------------------
Average Partial Effects of Conditional (on positive) Expected Value
--------------------------------------------------------------------
Variable    Regression    Hurdle     Variance  Total Effect   Std.Error
--------------------------------------------------------------------
     AGE     .037692    -.042974    -.001696    -.006978     .006151
    EDUC    -.169541    -.000766     .000000    -.170307     .005226***
    HSAT    -.016401     .000000     .000000    -.016401     .004871***
 MARRIED   -1.091875     .000000     .060194   -1.031681     .077136***
  HHKIDS     .466769     .000000     .000000     .466769     .025863***
  FEMALE     .000000     .000000    -.007673    -.007673     .001106***
--------------------------------------------------------------------
```

E45.11.5 Technical Details

The following technical details will, for convenience, replicate the results in Yen and Jones. As noted earlier, there is a point at which a substantive correction is needed in their results.

The relevant components of the log likelihood for the model are

$$\text{Prob}[T(y_i,\gamma) = 0] = \text{Prob}(y_i^* < 0 \text{ or } d_i^* < 0|\mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i)$$

$$= 1 - \text{Prob}(y_i^* > 0 \text{ and } d_i^* > 0|\, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i)$$

$$(1) \qquad = 1 - \Phi_2\left[\boldsymbol{\alpha}'\mathbf{z}_i, \frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}, \rho\right]$$

where $\Phi_2(\ldots)$ is a bivariate normal CDF. This term applies for censored, or 'limit' observations. (In Yen and Jones (1997, eqn (5)), the ρ in the preceding appears as σ_{12}/σ_i, which varies by observation. However, it must be the case in the joint distribution that the covariance equals the product of the correlation and the two standard deviations, so 'σ_{12}' must equal $\rho\times\sigma_i\times 1$. This produces the constant value ρ in the bivariate normal probability above. The necessity for this is clear; what must appear as the third argument in the probability is a correlation, but the erroneous term σ_{12}/σ_i which appears in its place in Yen and Jones's eqn (5) cannot be bounded in (-1,1).)

For nonlimit observations, the contribution to the likelihood function is the joint density of the observed y transformed by $T(y_i,\gamma)$ and the observation mechanism $d_i = 1$. This term is

$$f[y_i|T(y_i,\gamma),\ \{d_i=1|T(y_i,\gamma) > 0\}] = f[y_i|T(y_i,\gamma) \mid \{d_i=1|T(y_i,\gamma) > 0\}] \times \text{Prob}\{d_i = 1|T(y_i,\gamma) > 0\}.$$

The first term is the density for the observed nonzero y_i,

(2) $$[T(y_i,\gamma) \mid \{d_i = 1|\mathrm{T}(y_i,\gamma) > 0\}] = \left(1+\gamma^2 y_i^2\right)^{-1/2}\frac{1}{\sigma_i}\phi\left[\frac{T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right]$$

where the leading term is the Jacobian of the transformation from $T(y_i,\gamma)$ back to y_i. The second term is

(3) $$\text{Prob}\{d_i = 1|T(y_i,\gamma) > 0\} = \Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i+\rho[(T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i)/\sigma_i]}{\sqrt{1-\rho^2}}\right].$$

The log likelihood consists of logs of (1) for the limit observations and (2)×(3) for the nonlimit observations.

There are three expectations and three margins in this model. The probability of a nonlimit observation is

(4) $$P = \text{Prob}(y_i^* > 0 \text{ and } d_i^* > 0|\ \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i) = \text{Prob}(y_i > 0|\ \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] = \Phi_2\left[\boldsymbol{\alpha}'\mathbf{z}_i,\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i},\rho\right].$$

The partial effects are fairly simple:

$$\frac{\partial\text{Prob}(y_i>0\,|\,\mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i)}{\partial\begin{pmatrix}\mathbf{x}_i\\ \mathbf{z}_i\\ \mathbf{h}_i\end{pmatrix}} = \frac{\partial\Phi_2\left[\boldsymbol{\alpha}'\mathbf{z}_i,\dfrac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i},\rho\right]}{\partial\begin{pmatrix}\mathbf{x}_i\\ \mathbf{z}_i\\ \mathbf{h}_i\end{pmatrix}} = \begin{pmatrix} g_{i2}\dfrac{\boldsymbol{\beta}}{\sigma_i}\\ g_{i1}\boldsymbol{\alpha}\\ -g_{i2}\left(\dfrac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\boldsymbol{\delta}\end{pmatrix} = \begin{pmatrix}P_{\mathbf{x}}\\ P_{\mathbf{z}}\\ P_{\mathbf{h}}\end{pmatrix}$$

where g_{i1} and g_{i2} are the partial derivatives of the bivariate normal CDF with respect to the first two arguments. For the bivariate normal PDF, $\Phi_2(\mathrm{w}_1,\mathrm{w}_2,\rho)$, these are obtained using

$$\frac{\partial\Phi_2(w_1,w_2,\rho)}{\partial w_1} = \phi(w_1)\Phi\left[\frac{w_2-\rho w_1}{\sqrt{1-\rho^2}}\right] \text{ and } \frac{\partial\Phi_2(w_1,w_2,\rho)}{\partial w_2} = \phi(w_2)\Phi\left[\frac{w_1-\rho w_2}{\sqrt{1-\rho^2}}\right].$$

(See Greene (2011, eqn (17-52).) Where parts have variables in common, the components are added. Note that for a variable that appears in all three vectors, the partial effect is the sum of the three terms.

The conditional expectation, $E[y_i|y_i > 0, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i]$ is

$$E+ = E[y_i|y_i > 0, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] \; = \; \int_0^\infty y_i \frac{f(y_i \mid \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i)}{\text{Prob}(y_i > 0 \mid \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i)} dy_i .$$

The necessary components appear earlier;

$$E+ \; = \; \mathrm{E}[y_i|y_i > 0, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] \; =$$

$$\frac{1}{\Phi_2\left[\boldsymbol{\alpha}'\mathbf{z}_i, \dfrac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}, \rho\right]} \int_0^\infty y_i\left(1+\gamma^2 y_i^2\right)^{-1/2} \frac{1}{\sigma_i}\phi\left[\frac{T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right]\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho[(T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i)/\sigma_i]}{\sqrt{1-\rho^2}}\right]dy_i .$$

The unconditional mean is $E = E[y_i|\, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] \; = P \times E+$. This is

$$E = \; E[y \mid \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] = \int_0^\infty y_i\left(1+\gamma^2 y_i^2\right)^{-1/2} \frac{1}{\sigma_i}\phi\left[\frac{T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right]\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho[(T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i)/\sigma_i]}{\sqrt{1-\rho^2}}\right]dy_i .$$

Computation of the expectations and the derivatives requires the integration shown above. The complexity of the computation is reduced somewhat by the following:

1. The limits of integration are not functions of any of the interesting variables.
2. Though the functions appear quite complex, they can be greatly simplified for present purposes.

Define quantities

$$J_i = \; y_i\left(1+\gamma^2 y_i^2\right)^{-1/2}$$

$$e_i \; = \; \frac{T(y_i,\gamma)-\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i} .$$

We will also use

$$\partial e_i/\partial \mathbf{x}_i = -\boldsymbol{\beta}/\sigma_i,$$

$$\partial \sigma_i/\partial \mathbf{h}_i = \sigma_i\boldsymbol{\delta},$$

$$\partial e_i/\partial \sigma_i = -e_i/\sigma_i,$$

$$\partial e_i/\partial \mathbf{h}_i = \partial e_i/\partial \sigma_i \times \partial \sigma_i/\partial \mathbf{h}_i \; = \; (-e_i/\sigma_i)(\sigma_i\boldsymbol{\delta}) = -e_i\boldsymbol{\delta}.$$

Then, the unconditional mean can be written

$$\text{(5)} \qquad E \; = \; E[y \mid \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] = \int_0^\infty J_i \frac{1}{\sigma_i}\phi\left[e_i\right]\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right]d_{\;i} .$$

The relevant derivatives are now found as follows. We differentiate first with respect to e_i and σ_i. Thus,

$$
\begin{aligned}
\frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial e_i} &= \int_0^\infty J_i \frac{1}{\sigma_i}\left(-e_i \phi[e_i]\right)\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right] dy_i \\
&\quad + \frac{\rho}{\sqrt{1-\rho^2}}\int_0^\infty J_i \frac{1}{\sigma_i}\phi[e_i]\phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right] dy_i \\
&= E_1 + E_2 \times \frac{\rho}{\sqrt{1-\rho^2}} \\
&= E_e.
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial \sigma_i} &= \frac{-1}{\sigma_i}\int_0^\infty J_i \frac{1}{\sigma_i}\phi[e_i]\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right] dy_i + \frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial e_i}\frac{\partial e_i}{\sigma_i} \\
&= \frac{-1}{\sigma_i}E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i] + E_e \frac{\partial e_i}{\sigma_i} \\
&= \frac{-1}{\sigma_i}E + \left(E_1 + E_2 \times \frac{\rho}{\sqrt{1-\rho^2}}\right)\left(\frac{-e_i}{\sigma_i}\right) \\
&= E_\sigma
\end{aligned} \tag{7}
$$

Collecting the parts, we obtain

$$
\begin{aligned}
\frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial \mathbf{h}_i} &= E_\sigma \times (\sigma_i \boldsymbol{\delta}) = E_{\mathbf{h}} \\
\frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial \mathbf{x}_i} &= E_e \frac{-1}{\sigma_i}\boldsymbol{\beta} = E_{\mathbf{x}} \\
\frac{\partial E[y \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i]}{\partial \mathbf{z}_i} &= \frac{\boldsymbol{\alpha}}{\sqrt{1-\rho^2}}\int_0^\infty J_i \frac{1}{\sigma_i}\phi[e_i]\phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right] dy_i = E_2 \times \frac{\boldsymbol{\alpha}}{\sqrt{1-\rho^2}} = E_{\mathbf{z}}
\end{aligned}
$$

Derivatives of $E[y_i|y_i > 0, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i]$ are simple given the parts already computed. Since

$$E[y_i|y_i > 0, \mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i] = [1/\Phi_2(\ldots)]\times E[y_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{h}_i],$$

so the relevant derivatives are simply $E_{\mathbf{x}}^{+} = (1/P)\{E_{\mathbf{x}} - E \times P_{\mathbf{x}}\}$ and likewise for $\mathbf{z}$ and $\mathbf{h}$.

Computation all of these parts requires the three integrals,

$$E = \int_0^\infty J_i \frac{1}{\sigma_i}\phi\left[e_i\right]\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right]dy_i,$$

$$E_1 = \int_0^\infty J_i \frac{1}{\sigma_i}\left(-e_i\phi\left[e_i\right]\right)\Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right]dy_i,$$

$$E_2 = \int_0^\infty J_i \frac{1}{\sigma_i}\phi\left[e_i\right]\phi\left[\frac{\boldsymbol{\alpha}'\mathbf{z}_i + \rho e_i}{\sqrt{1-\rho^2}}\right]dy_i.$$

None of these are in forms amenable to Hermite or Gaussian quadrature. We use a 15 point Newton-Cotes trapezoid method. There is a final complication in that the method is used for proper integrals (with finite limits). As the upper limit here is infinite, a stopping rule is required – simple experimentation will take intolerably long. In the integrals, the main weighting function is $\phi(e_i)$. The Jacobian, J_i is essentially linear in y_i and the second function in the integrand, $\Phi[.]$ or $\phi[.]$ quickly asymptotes to 1 or 0 as e_i increases. The task then is to determine the practical limit for y_i which is then used to compute e_i. We solve for the y_i at which e_i reaches +8.0, where the standard normal PDF is essentially zero. This, in turn is solved using a first order approximation to $T(y_i,\gamma)$.

In order to use the delta method to compute standard errors, we require the Jacobians of the partial effects with respect to the full parameter vector. These, in turn, can all be computed from

$$J_E = \partial E/\partial\boldsymbol{\theta}' \text{ where } \boldsymbol{\theta}' = [\boldsymbol{\beta}',\sigma,\boldsymbol{\delta}',\alpha,\rho,\boldsymbol{\gamma}']$$

$$J_1 = \partial E_1/\partial\boldsymbol{\theta}'$$

$$J_2 = \partial E_2/\partial\boldsymbol{\theta}'$$

These must also be computed from the integrals. Numerical derivatives of the integrals are used for these three vectors.

E46: Panel Data Models for Censored Data and Truncated Distributions

E46.1 Introduction – Model Frameworks

This chapter will describe the extensions to panel data settings of the tobit model developed in Chapter E45. The estimators shown here are the same for the truncated regression (TRUNCATION) and grouped data model (GROUPED) that are discussed in Chapter E47. The full set of estimators developed for the binary choice models in Chapters E30 and E31 and for count data models in Chapter E44 are available here as well. The results are collected here for both the censoring and truncation models as the models are largely the same, with a slight variation in the assumption about the data observation process.

The three models described in this chapter are the tobit model of the previous chapter and two variations on it, the truncated regression and the grouped data, or interval censored regression: All three structures are based on the latent regression structure,

$$y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ \varepsilon_{it} \sim N[0,\sigma^2].$$

The three differ in the observation mechanism for the observed dependent variable. The tobit model is

$$\text{if } y_{it}^* \leq L_{it}, \text{ then } y_{it} = L_{it} \text{ (lower tail censoring)}$$

$$\text{if } y_{it}^* \geq U_{it}, \text{ then } y_{it} = U_{it} \text{ (upper tail censoring)}$$

$$\text{if } L_{it} < y_{it}^* < U_{it}, \text{ then } y_{it} = y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}.$$

A special case of the censored data regression model arises when the range of the dependent variable is completely censored. This is the case when data are reported only by interval category. For example, income data might be reported only by range. We assume that the finite (internal) terminal points are known variables or constants. The dependent variable is coded $y = 1, 2, ..., J$ (not 0,..., as in the case of the ordered probability models). For example, consider a survey of incomes, which reports ranges:

$$y = \begin{array}{lrcl} 1 \text{ if} & & y^* < & \$\,15{,}000, \\ 2 \text{ if} & \$\,15{,}000 \leq & y^* < & \$\,30{,}000, \\ 3 \text{ if} & \$\,30{,}000 \leq & y^* < & \$\,50{,}000, \\ 4 \text{ if} & \$\,50{,}000 \leq & y^* < & \$\,75{,}000, \\ 5 \text{ if} & & y^* \geq & \$\,75{,}000. \end{array}$$

The observation mechanism, once again based on the latent regression model, is

$$y_{it} = j \text{ if } A_{i,j-1} \leq y_{it}^* < A_{ij},\ j = 1,...,J,\ A_0 = -\infty,\ A_J = +\infty.$$

The truncated regression model applies to the nonlimit observations in the tobit formulation. The observation mechanism is simply

$$\text{if } L_{it} < y_{it}^* < U_{it}, \text{ then } y_{it} = y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},$$

y_{it} is unobserved otherwise.

E46.2 Panel Data Frameworks

Chapter E45 analyzed in detail several variants of a single equation, latent regression model of censoring in the linear (or nonlinear) regression context. The estimator is assumed to be based on a cross section. Since it is typically applied in micro-level data, the extension of the censored and truncated regression models to panel data is a natural direction. There are several formulations for extensions to panel data setting. These include, where *f*(.) denotes the density for the observed random variable (i.e., the model),

- **Fixed effects:** $f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i d_{it})$, $\text{Cov}(d_{it},\mathbf{x}_{it})$ not necessarily zero
- **Random effects:** $f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i)$, $\text{Cov}(u_i,\mathbf{x}_{it}) = 0$
- **Random parameters:** $f(y_{it}) = f(\boldsymbol{\beta}_i'\mathbf{x}_{it})$

  $\boldsymbol{\beta}|i \sim h(\boldsymbol{\beta}|i)$ with mean vector $\boldsymbol{\beta}+\boldsymbol{\Delta}\mathbf{z}_i$ and covariance matrix $\boldsymbol{\Sigma}$
- **Latent class:** $f(y_{it}|\text{class } j) = f(\boldsymbol{\beta}_j'\mathbf{x}_{it})$, $\text{Prob}[\text{class} = j] = F_j(\boldsymbol{\theta})$

We will detail these for the tobit model. The commands and results are the same for the truncated and grouped data regressions, so they will be noted during the development.

To illustrate the various estimators, we will use an artificial data set containing 1,000 groups of 10 observations, or 10,000 observations in total. So that the applications can be replicated, we use the following data setup. We first set the seed for the random number generator so that data can be replicated and set the dimensions of the data set.

```
ROWS      ; 10000 $
CALC      ; Ran(12345) $
SAMPLE    ; 1-10000 $
```

If you are replicating these computations, note that the **ROWS** command may not be needed. When you start *LIMDEP*, the bar at the top of the project window will indicate the current setting of the number of rows in the data area. You will only need the **ROWS** command if the value shown is less than 10,000.

To begin, we create the group specific effects. (The values of **u** are used in generating the data at this point, while **v** is used later.)

```
MATRIX    ; u = Rndm(1000)
          ; v = Rndm(1000) $
```

The underlying data satisfy the assumptions of a fixed effects model. The group effect is correlated with one of the independent variables.

```
CREATE      ; i = Trn(10,0) $
CREATE      ; x1 = Rnn(0,1) ; x2 = Rnn(0,1)
            ; z1 = Rnn(0,1) ; z2 =  Rnd(2) - 1 $
MATRIX      ; x1b = Gxbr(x1,i) ; u = u + .5 * x1b $
CREATE      ; eit = Rnn(0,2) ; ui = u(i)  ; vi = .25 * v(i) $
CREATE      ; ys = x1 + x2 + eit + ui ; y = Max(0,ys) $
NAMELIST    ; x = x1,x2,z1,z2,one $
```

Note that the data are actually generated by a fixed effects model. We will also be including $z1$ and $z2$ in the equation, though the true coefficients on them are zero. The **MATRIX** command that creates $x1b$ also makes these group means part of the effects, thus inducing the correlation. The disturbance variance in the model is $\sigma_\varepsilon^2 = 2$.

This is the base case with no treatment for group effects. We compare the results for the tobit and truncated regressions. After this, we will focus on the tobit model.

```
TOBIT       ; Lhs = y ; Rhs = x ; Partial Effects $
TRUNCATE    ; Lhs = y ; Rhs = x ; Partial Effects $
```

The commands below estimate a fixed effects model, a random effects model and a random parameters model for the tobit framework. Results from these are shown below with development of the estimators. The truncation model is presented below as well. Note that the full set of results for the truncation model apply to the nonlimit data.

```
TOBIT       ; Lhs = y ; Rhs = x; Random Effects ; Pds = 10 ; Partial Effects $
TOBIT       ; Lhs = y ; Rhs = x
            ; RPM ; Fcn = one(n) ; Pds = 10 ; Pts = 20 ; Halton ; Partial Effects $
TOBIT       ; Lhs = y ; Rhs = x; FEM ; Pds = 10 ; Partial Effects $
```

The commands below simulates the conditions of a random parameters model – the coefficient on $x1$ has a normal distribution with mean 1 and standard deviation 0.25.

```
CREATE      ; y = (1+vi)*x1 + x2 + eit + ui
            ; y = (y > 0) * y $
TOBIT       ; Lhs = y ; Rhs = x
            ; RPM ; Fcn = one(n), x1(n) ; Correlated
            ; Pds = 10 ; Pts = 20 ; Halton $
```

This is the base case tobit model with no individual effects. Note that the three true values for the regression parameters are 0.0, 1.0, 1.0, 0.0, 0.0, respectively. The OLS starting values are not shown. The tobit and truncated regression MLEs are both consistent and are estimating the same parameters. The similarity in the two sets of results is to be expected. The tobit estimator is based on more information, so one would expect it to be more efficient (have smaller variances). This is, in fact, clearly evident in the results.

The base case results are followed by two estimators of the random effects model. The first one uses the Butler and Moffitt quadrature method. The second treats the random effects case as a random parameters model in which only the constant term is random. The model is estimated by maximum simulated likelihood. The results of the two methods are nearly identical. This is to be expected. It is striking, however, that the RP approach achieves the results with only 25 Halton draws, which is far less than what one would typically use in practice. The final set of results is the unconditional fixed effects estimator. The FEM includes estimates of the 1,000 dummy variable coefficients (not shown). In principle, the estimator is affected by the incidental parameters problem. However, with T = 10, this appears not to be the case here.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                      Y
Log likelihood function    -13953.44433
Estimation based on N =  10000, K =   6
Inf.Cr.AIC  =  27918.9 AIC/N =    2.792
Threshold values for the model:
Lower =      .0000    Upper = +infinity
LM test [df] for tobit=      5.568[  5]
Normality Test, LM    =       .385[  2]
ANOVA  based fit measure =      .126658
DECOMP based fit measure =      .287434
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
      X1|    1.03177***      .02649    38.95  .0000      .97985   1.08369
      X2|     .99321***      .02667    37.24  .0000      .94094   1.04548
      Z1|    -.02438         .02524     -.97  .3341     -.07385    .02509
      Z2|    -.13019***      .05038    -2.58  .0098     -.22894   -.03144
Constant|     .04118         .03824     1.08  .2815     -.03377    .11614
        |Disturbance standard deviation
   Sigma|    2.21774***      .02421    91.59  .0000     2.17028   2.26520
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Limited Dependent Variable Model - TRUNCATE
Dependent variable                      Y
Log likelihood function     -8069.80996
Estimation based on N =  10000, K =   6
Inf.Cr.AIC  =  16151.6 AIC/N =    1.615
Threshold values for the model:
Lower =      .0000    Upper = +infinity
Observations after truncation      4971
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
      X1|    1.11958***      .05898    18.98  .0000     1.00398   1.23518
      X2|     .95695***      .05811    16.47  .0000      .84306   1.07084
      Z1|     .00038         .04785      .01  .9937     -.09341    .09416
      Z2|    -.18004*        .09624    -1.87  .0614     -.36867    .00860
Constant|    -.00819         .13922     -.06  .9531     -.28106    .26469
        |Disturbance standard deviation
   Sigma|    2.23911***      .05200    43.06  .0000     2.13720   2.34102
```

```
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
(Tobit)
----------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point    .8739
Scale Factor for Marginal Effects   .4961
--------+-------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
       Y|      Effect         Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------
      X1|     .51184***      .01302    39.31  .0000       .48632    .53735
      X2|     .49271***      .01305    37.76  .0000       .46713    .51829
      Z1|    -.01210         .01252     -.97  .3341     -.03664    .01245
      Z2|    -.06458***      .02499    -2.58  .0098     -.11357   -.01560
--------+-------------------------------------------------------------
(Truncated Regression)
Conditional Mean at Sample Point   2.0032
Scale Factor for Marginal Effects   .4207
--------+-------------------------------------------------------------
      X1|     .47104***      .02067    22.79  .0000       .43053    .51155
      X2|     .40262***      .02141    18.80  .0000       .36065    .44459
      Z1|     .00016         .02013      .01  .9937     -.03930    .03962
      Z2|    -.07575*        .04044    -1.87  .0610     -.15500    .00350
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

E46.3 Fixed Effects Models

The fixed effects model has the desirable characteristic that it is not necessary to assume that the individual component is orthogonal to the included variables. As such, it is a more robust specification than the random effects estimator. The complication is that in a data set with n groups or individuals, each observed T_i times, the fixed effects specification creates n new parameters to be estimated. In practical terms, n could be enormous (thousands), so approaches are devised to 'sweep' the coefficient of the fixed effects out of the estimating equations. Other issues concern the 'incidental parameters problem' and the attendant inconsistency of the estimator of the main parameters. *LIMDEP* contains a full, unrestricted fixed effects estimator. The issues of small T_i and the incidental parameters problem must be resolved outside the program. There is some evidence that even for fairly small T_i the issue of small sample bias of the fixed effects estimator is overstated. (See Heckman (1981). See also Greene (2004b) for evidence specifically about the tobit and truncated regression models. This study is discussed in Section E46.3.3.) The practical issue of potentially large numbers of parameters has been overcome – *LIMDEP* is able to fit up to 100,000 individual dummy variable parameters even in a model with no minimal sufficient statistics, such as the tobit or truncated regression models.

The command for estimation is

TOBIT ; Lhs = dependent variable
or TRUNCATE ; Rhs = independent variables
or GROUPED ; Pds = panel specification
; FEM (for fixed effects model) $

The default limit value is zero, with left censoring or truncation. The limit value and censoring in the lower tail may be changed with

; Limit = the nonzero value
and/or **; Upper censoring or ; Upper truncation**

This estimator only supports censoring in one tail. You may request residuals, fitted values, partial effects, and all other optional features with this model. Restrictions that you would impose with **; Rst,** however, must be built into the model at the outset. The algorithm does not accommodate restrictions.

NOTE: Your Rhs list should not include a constant term, as the fixed effects model fits a complete set of constants for the set of groups. If you do include *one* in your Rhs list, it is removed prior to beginning estimation.

The fixed effects models are estimated by maximum likelihood. The fixed effects model assumes a group specific effect:

$$f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$$

where α_i is the parameter to be estimated. You may also fit a two way fixed effects model

$$f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \gamma_t)$$

where γ_t is an additional, time (period) specific effect. The time specific effect is requested by adding

; Time

to the command if the panel is balanced, and

; Time = variable name

if the panel is unbalanced. For the unbalanced panel, we assume that overall, the sample observation period is

$$t = 1,2,..., T_{max}$$

and that the 'Time' variable gives for the specific group, the particular values of *t* that apply to the observations. Thus, suppose your overall sample is five periods. The first group is three observations, periods 1, 2, 4, while the second group is 4 observations, 2, 3, 4, 5. Then, your panel specification would be

; Pds = Ti, for example, where *Ti* = 3, 3, 3, 4, 4, 4, 4
and **; Time = Pd**, for example, where *Pd* = 1, 2, 4, 2, 3, 4, 5.

NOTE: See the discussion in Chapter R23 for technical details on how this model is estimated. It places an important restriction on the two way fixed effects model.

The only fitting algorithm available is Newton's method, and some of the options for control of the optimization routine are not available. Those that are available are shown in the list below. This estimator cannot accommodate restrictions, so

; Rst = list
and **; CML: specification**

are both ignored.

Starting values for the iterations are obtained by fitting the basic model without fixed effects by ordinary least squares. If you request the display of these results with **; OLS**, you will see a constant term in these results even though you have not included one in your commands. This is used to get the starting value for the fixed effects. Iterations begin with the restricted model that forces all the fixed effects to equal the constant term in the restricted model. You may provide your own starting values for the slope parameters with

; Start = ... the list of values for β,σ,α.

Do not include a set of constants in your starting values. The last value, if it is included (it is optional), provides a common starting constant.

Results that are kept for this model are

Matrices: *b* = estimate of β
varb = asymptotic covariance matrix for estimate of β
alphafe = estimated fixed effects

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

The upper limit on the number of groups is 100,000.

Standard Model Specifications for the Fixed Effects Tobit and Truncated Regression Models

This is the full list of general specifications available for this model estimator.

Controlling Output from Model Commands

; Par keeps ancillary parameter σ in main results vector *b*.
; Margin displays marginal effects.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.

E46.3.1 Technical Notes

The fixed effects model is fit by 'brute force.' *LIMDEP* actually estimates the full K+N up to 100,150 coefficients by Newton's method. It is possible to fit the huge number of coefficients because we take advantage of the properties of the sparse second derivatives matrix. One of the implications, however, is that there is no covariance matrix computed for the fixed effects. It is possible to test for the fixed effects model with a likelihood ratio test (though the incidental parameters issue casts some doubt on the validity of this test), but since the covariance matrix is not computed, it is not possible to do any kind of inference for individual fixed effects. Marginal effects in the fixed effects model are computed at the means of the data and with the sample average of the fixed effects estimates as the constant term.

NOTE: The individual specific constant term cannot be computed for any group in which the dependent variable always takes the limit value (usually zero). The model results will show the count of such groups. For example, for the preceding data, the following output is produced:

```
-------------------------------------------------------------------------------
FIXED EFFECTS Tobit  Model
Maximum Likelihood Estimates
Dependent variable                    Y
Weighting variable                 None
Number of observations            10000
Iterations completed                  5
Log likelihood function       -12499.95
Sample is 10 pds and    1000 individuals.
Bypassed   12 groups with inestimable a(i).
TOBIT (censored) regression model
(Lower) truncation limit is     .00
-------------------------------------------------------------------------------
```

This shows that 12 of the 1,000 groups contained 10 observations in which *y* equals zero in all of them. The truncated regression estimator must also check for this condition in the data. In principle, your data will not contain limit observations for the truncation model. But, if in fact, it does, these observations are bypassed. If all of the observations in one or more groups are bypassed, then the same warning will appear for the truncation model.

The two way fixed effects estimator is computed by actually creating the time specific dummy variables and adding them to the model. This means that the 150 parameter limit on model size applies to the number of variables in the model plus the number of periods (minus one).

E46.3.2 Application

The following presents one and two way FEMs for the tobit model.

TOBIT ; Lhs = y ; Rhs = x ; FEM ; Pds = 10 ; Partial Effects $
TOBIT ; Lhs = y ; Rhs = x ; FEM ; Pds = 10 ; Time ; Partial Effects $

```
-----------------------------------------------------------------------------
FIXED EFFECTS Tobit  Model
Dependent variable                    Y
Log likelihood function    -12622.25665
Estimation based on N =  10000, K = 994
Inf.Cr.AIC  =  27232.5 AIC/N =    2.723
Sample is 10 pds and   1000 individuals
Skipped   11 groups with inestimable ai
TOBIT (censored) regression model
(Lower) truncation limit is     .00
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
      X1|     .98385***     .02420    40.65  .0000      .93641   1.03128
      X2|     .99171***     .02462    40.27  .0000      .94345   1.03998
      Z1|    -.03002        .02311    -1.30  .1941    -.07532    .01529
      Z2|    -.08437*       .04624    -1.82  .0680    -.17499    .00625
        | Variance parameter given is sigma
Std.Dev.|    1.86697***     .02005    93.09  .0000     1.82767   1.90628
--------+--------------------------------------------------------------------
```

Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Estimated E[y|means,mean alphai]= .748
Estimated scale factor for dE/dx= .501

Y	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence Interval	
X1	.49314***	.00575	37.52	.0000	.46738	.51890
X2	.49709***	-.00067	37.44	.0000	.47107	.52311
Z1	-.01505	.2748D-04	-1.30	.1941	-.03775	.00766
Z2	-.04229*	-.02839	-1.86	.0631	-.08689	.00231

FIXED EFFECTS Tobit Model
Dependent variable Y
Log likelihood function -12618.43315
Estimation based on N = 10000, K =1003
Sample is 10 pds and 1000 individuals
Skipped 11 groups with inestimable ai
No. of period specific effects= 9
TOBIT (censored) regression model
(Lower) truncation limit is .00

Y	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Index function for probability					
X1	.98294***	.02612	37.64	.0000	.93175	1.03412
X2	.99256***	.02799	35.46	.0000	.93771	1.04741
Z1	-.03015	.02310	-1.31	.1917	-.07542	.01512
Z2	-.08425*	.04627	-1.82	.0686	-.17493	.00644
Period1	-.07945	.10545	-.75	.4512	-.28613	.12723
Period2	-.03225	.10539	-.31	.7596	-.23882	.17432
Period3	-.13292	.13870	-.96	.3379	-.40477	.13893
Period4	-.01231	.09998	-.12	.9020	-.20827	.18365
Period5	-.15049	.14569	-1.03	.3016	-.43603	.13505
Period6	-.05039	.10948	-.46	.6453	-.26497	.16419
Period7	.01634	.09507	.17	.8635	-.17000	.20268
Period8	.02710	.09367	.29	.7723	-.15649	.21069
Period9	-.11898	.13299	-.89	.3710	-.37965	.14168
	Variance parameter given is sigma					
Std.Dev.	1.86605***	.02004	93.10	.0000	1.82676	1.90533

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Estimated E[y|means,mean alphai]= .747
Estimated scale factor for dE/dx= .501

Y	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence Interval	
X1	.49266***	.00575	35.38	.0000	.46537	.51995
X2	.49748***	-.00067	33.74	.0000	.46858	.52638
Z1	-.01511	.2762D-04	-1.31	.1917	-.03780	.00758
Z2	-.04223*	-.02836	-1.85	.0637	-.08686	.00241

This is the truncated regression model estimated with the same data. Note that the sample is different for this model, because the estimator skips the limit observations. This eliminates the same 12 groups for which *y* is always zero, However, since the estimator is skipping all limit observations, the sample is in fact, greatly reduced. Indeed, the command set

```
CREATE      ; d = y > 0 $
MATRIX      ; dbar = Gxbr(d,i) $
SAMPLE      ; 1-1000 $
CREATE      ; dd = dbar $
REJECT      ; dd < 1 $
SAMPLE      ; All $
REJECT      ; d = 0 $
SETPANEL    ; Group = i ; Pds = ti $
TRUNCATE    ; Lhs = y  ; Rhs = x ; Partial Effects ; FEM ; Pds = 10 ; Time $
```

```
-----------------------------------------------------------------------------
FIXED EFFECTS TrncRg Model
Dependent variable                    Y
Log likelihood function     -6376.58031
Estimation based on N =   4971, K =1003
Inf.Cr.AIC  =  14759.2 AIC/N =    2.969
Unbalanced panel has    989 individuals
Skipped    0 groups with inestimable ai
No. of period specific effects= 9
TRUNCATED regression model
(Lower) truncation limit is     .00
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
      X1|     .87000***      .05335    16.31  .0000       .76544    .97457
      X2|     .82205***      .02104    39.07  .0000       .78081    .86329
      Z1|     .00729         .03578      .20  .8385     -.06283    .07741
      Z2|    -.10906         .07344    -1.49  .1375     -.25300    .03487
 Period1|    -.37064         .84491     -.44  .6609    -2.02663   1.28535
 Period2|    -.48523        2.16078     -.22  .8223    -4.72028   3.74981
 Period3|    -.42006        1.94393     -.22  .8289    -4.23010   3.38998
 Period4|    -.19895        1.22310     -.16  .8708    -2.59618   2.19828
 Period5|    -.44021        2.01074     -.22  .8267    -4.38119   3.50078
 Period6|    -.41834        1.93849     -.22  .8291    -4.21771   3.38104
 Period7|    -.36124        1.75008     -.21  .8365    -3.79134   3.06885
 Period8|    -.44737        2.03715     -.22  .8262    -4.44011   3.54537
 Period9|    -.08877         .91656     -.10  .9228    -1.88519   1.70765
        |Variance parameter given is sigma
Std.Dev.|    1.68755***      .02963    56.96  .0000      1.62948   1.74561
--------+--------------------------------------------------------------------
```

The results are striking in another respect. Whereas the tobit estimator of the parameters seems to estimate them quite well, even with 10,000 observations, it appears that the tobit estimator has slightly underestimated σ^2. But, the truncated regression estimator, again, in spite of the large sample, has underestimated all of the parameters. These are precisely what would be predicted by the results on the incidental parameters problem in the next section.

E46.3.3 The Incidental Parameters Problem

Section R23.2.2 mentions the incidental parameters (IP) problem as a feature of the estimation of fixed effects models by maximum likelihood. As widely understood in econometrics, the IP problem is associated with a persistent upward bias in the parameter estimator in the FE model. Tables E46.1 and E46.2 below are extracted from the only received study of the IP problem in the tobit and truncated regression models, Greene (2004b). The remaining literature is focused on binary choice models. The tables describe a Monte Carlo study of censoring and truncation in the model

$$y_{it}^* = \alpha_i + \beta x_{it} + \delta_i d_{it} + \varepsilon_{it}$$

in which x_{it} is a continuous variable and d_{it} is a dummy variable. The underlying data are generated according to the fixed effects model – the effects are correlated with both x_{it} and d_{it}. The table entries show the estimates of the percentage biases of various estimators based on 1,000 replications of the model, with N also equal to 1,000. Surprisingly, the experiment suggests that the conventional wisdom is wrong for both the tobit and the truncated regression models. For the tobit case, the bias appears to manifest itself not in the estimator of β, but in the estimator of σ. The implications for the estimated marginal effects and for the estimated standard errors are shown in the lower rows of the table. Table E46.1 suggests that the truncated regression model works in the opposite direction – all model components appear to be biased downward. The overall conclusion here is also somewhat contradictory. Based on these results, one is tempted to conclude that once T reaches 5, the IP problem is relatively small for these particular models.

Estimate	T = 2	T = 3	T = 5	T = 8	T = 12	T = 15	T = 20
β	0.67	0.53	0.50	0.29	0.098	0.082	0.047
δ	0.33	0.90	0.57	0.54	0.32	0.16	0.14
σ	-36.14	-23.54	-13.78	-8.40	-5.54	-4.43	-3.30
ME_x	15.83	8.85	3.65	1.30	0.44	0.22	0.081
ME_d	19.67	11.85	5.08	2.16	0.89	0.46	0.27
S.E. (β)	-32.92	-19.00	-11.30	-8.36	-6.21	-4.98	0.63
S.E. (δ)	-32.87	-22.75	-12.66	-7.39	-5.56	-6.19	0.25

Table E46.1 Tobit Model, Behavior of the MLE/FE, Percentage Bias in Estimation

Estimate	T = 2	T = 3	T = 5	T = 8	T = 12	T = 15	T = 20
β	-17.13	-11.97	-7.64	-4.92	-3.41	-2.79	-2.11
δ	-22.81	-17.08	-11.21	-7.51	-5.16	-4.14	-3.27
σ	-35.36	-23.42	-14.28	-9.12	-6.21	-4.94	-3.75
ME_x	-7.52	-4.85	-2.87	-1.72	-1.14	-0.94	-0.67
ME_d	-11.64	-8.65	-5.49	-3.64	-2.41	-1.90	-1.53
S.E. (β)	-33.00	-21.36	-12.30	-8.41	-3.83	-6.17	-2.62
S.E. (δ)	-31.52	-16.81	-9.45	-3.82	-7.74	-1.43	-0.61

Table E46.2 Truncated Regression Model, Behavior of the MLE/FE, Percentage Bias

E46.4 Random Effects Models

The random effects model with censored data or truncation is based on the same latent regression used earlier, but with a different treatment of the common effect. Specifically,

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i.$$

$$\varepsilon_{it}, u_i \sim \text{ bivariate normal with means } (0,0), \text{ variances } (\sigma^2, \omega^2) \text{ and correlation } 0.$$

Data are observed by the mechanisms

$$y_{it} = \text{Max}(L_{it}, y_{it}^*) \text{ for the tobit model and}$$

$$y_{it} = y_{it}^* \text{ if } y_{it} \geq L_{it} \text{ and unobserved otherwise for the truncation model.}$$

The essential assumptions are that the random effect is the same in every period and the unique effect, ε_{it} is uncorrelated across periods. All effects are uncorrelated across individuals. Since the unique effects are independent across periods, all of our previous results apply to the conditional distribution of $y_{it}|u_i$.

As before, for the tobit model, let $d_{it} = 1$ if $y_{it} > L_{it}$ (uncensored) and 0 otherwise. Then, the density of the observed random variable, y_{it} is

$$f(y_{it}|u_i,d_{it} = 0) = \text{Prob}[y_{it}^* \leq L_{it} \mid u_i] = \Phi\left(\frac{L_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i}{\sigma}\right) \text{ (censored)}$$

$$f(y_{it}|u_i,d_{it} = 1) = \frac{1}{\sigma}\phi\left(\frac{y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i}{\sigma}\right) \text{ (uncensored).}$$

(For convenience, we leave the dependence on $\mathbf{x}_{it}$ implicit.) For purposes of formulating the log likelihood, we will combine these by writing

$$f(y_{it}|u_i) = [f(y_{it}|u_i,d_{it} = 0)]^{1-d_{it}} \times [f(y_{it}|u_i,d_{it} = 1)]^{d_{it}}.$$

Since, conditioned on u_i, the observations are independent, the joint density of the T_i observations for group i is the product of the individual densities;

$$f(y_{i1},y_{i2},\ldots,y_{iTi}|u_i) = \prod_{t=1}^{T_i} f(y_{it}|u_i).$$

To form the log likelihood function, we need the unconditional distribution, the log of which then enters the function to be maximized. The unconditional density is obtained by integrating u_i out of the conditional density

$$f(y_{i1},y_{i2},\ldots,y_{iTi}|u_i) = \int_{-\infty}^{\infty} f(y_{i1}, y_{i2}, \ldots, y_{iT_i}|u_i)g(u_i)du_i$$

Recall $g(u) = (1/\omega)\phi(u_i/\omega)$. Combining all terms, then summing the logs to obtain the log likelihood function, we have log

$$\log L_{tobit} = \sum_{i=1}^{n} \log\left\{\int_{-\infty}^{\infty} \frac{1}{\omega\sqrt{2\pi}}\exp\left(-\frac{u_i^2}{2\omega^2}\right)\prod_{t=1}^{T_i}\left[\Phi\left(\frac{L_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}-u_i}{\sigma}\right)\right]^{1-d_{it}}\left[\frac{1}{\sigma}\phi\left(\frac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}-u_i}{\sigma}\right)\right]^{d_{it}} du_i\right\}.$$

This function is to be maximized with respect to ($\boldsymbol{\beta}$,σ,ω). The same sequence of steps produces the counterpart for the truncated regression model,

$$\log L_{truncation} = \sum_{i=1}^{n} \log\left\{\int_{-\infty}^{\infty} \frac{1}{\omega\sqrt{2\pi}}\exp\left(-\frac{u_i^2}{2\omega^2}\right)\prod_{t=1}^{T_i}\left[\Phi\left(\frac{(\mathbf{x}_{it}'\boldsymbol{\beta}+u_i)-L_{it}}{\sigma}\right)\right]^{-1}\left[\frac{1}{\sigma}\phi\left(\frac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}-u_i}{\sigma}\right)\right] du_i\right\}.$$

There are two ways to maximize the log likelihoods, both of which will generally prove successful.

- The function and its derivatives can be evaluated by Hermite quadrature.

- Since the function and derivatives are equal to expectations, $E_u[h(...,u_i)]$, they can be approximated by simulation. At each point at which the function or derivative must be computed, the integral is replaced by the average of R function evaluations at random draws from the currently estimated distribution of u_i. The simulation method is considered in the next section.

  The quadrature based estimator can be requested with

```
TOBIT          ; Lhs  = dependent variable
or TRUNCATE    ; Rhs  = independent variables
               ; Pds  = panel specification of group sizes
               ; Random Effects $
```

(The quadrature based random effects estimator is not available for the GROUPED data model. The random parameters specification is provided for the **GROUPED** command, so a random effects model can be estimated by maximum simulated likelihood instead.)

The limit value and censoring in the lower tail may be changed with

```
               ; Limit = the nonzero value or variable name
and/or         ; Upper censoring
```

Censoring or truncation in both tails of the distribution are specified with

```
               ; Limits = lower specification, upper specification
```

where each specification may be a constant or the name of a variable. The other options, such as fitted values, marginal effects, and so on are the same as for the tobit model without the random effects treatment. Censoring may be in either or both tails and censoring limits may be constant or may vary by observation. As usual, zero, lower censoring is the default.

NOTE: There is no limit on the number of groups in this model. As always, the panel may be unbalanced. Also, in principle, there is no internal limit on the number of observations in a group. However, do note in this model, to compute the log likelihood, it is necessary actually to compute the joint probability for the T_i observations in a group. That is the product of probabilities, and, to the point, not the sum of the logs. Therefore, if your panel has a very large number of observations in a group – consider monthly observations on some variable, by firm, for a number of years – it will be necessary to compute the product of a large number of probabilities. It is possible that this value can become extremely small, and when so, accuracy is lost in the computations. On occasion, if the estimator claims it is unable to locate a maximum of the objective function, it is possible that this is the reason.

To illustrate the estimator, the following reports the random effects estimates of the model fit earlier.

TOBIT ; Lhs = y ; Rhs = x ; Random Effects ; Pds = 10 ; Partial Effects $
TOBIT ; Lhs = y ; Rhs = x
; RPM ; Fcn = one(n) ; Pds = 10 ; Pts = 20 ; Halton ; Partial Effects $

```
-----------------------------------------------------------------------------
Reestimated RANDOM EFFECTS Tobit Model
Dependent variable                   Y
Log likelihood function    -13659.74001
Restricted log likelihood  -13953.44433
Chi squared [   1 d.f.]       587.40864
Significance level               .00000
McFadden Pseudo R-squared      .0210489
Estimation based on N =  10000, K =   7
Inf.Cr.AIC  =  27333.5 AIC/N =    2.733
Model estimated: Aug 01, 2011, 22:21:56
Sample is 10 pds and    1000 individuals.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|    1.00083***      .02498    40.07  .0000      .95188   1.04979
      X2|     .98910***      .02671    37.03  .0000      .93676   1.04145
      Z1|    -.02806         .02424    -1.16  .2470    -.07557     .01945
      Z2|    -.10308**       .04788    -2.15  .0313    -.19693    -.00923
Constant|     .02745         .04647      .59  .5547    -.06363     .11853
Sigma(v)|    2.00007***      .02294    87.20  .0000    1.95511   2.04503
Sigma(u)|     .94842***      .03757    25.24  .0000      .87479   1.02206
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point    .7869
Scale Factor for Marginal Effects   .4956
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
       Y|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|     .49601***      .01348    36.80  .0000      .46959    .52242
      X2|     .49019***      .01590    30.83  .0000      .45903    .52135
      Z1|    -.01391         .01237    -1.12  .2610    -.03816    .01034
      Z2|    -.05109***      .01208    -4.23  .0000    -.07477   -.02741
```

Partial Effects for Tobit (Censored) Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable

(Delta method)	Partial Effect	Standard Error	|t|	95% Confidence Interval	
X1	.49542	.01413	35.06	.46772	.52312
X2	.48961	.01508	32.47	.46005	.51917
Z1	-.01389	.01200	1.16	-.03741	.00963
* Z2	-.05103	.02368	2.15	-.09745	-.00462

Random Coefficients Tobit Model
Dependent variable Y
Log likelihood function -13670.53292
Estimation based on N = 10000, K = 7
Inf.Cr.AIC = 27355.1 AIC/N = 2.736
Model estimated: Aug 01, 2011, 22:22:11
Sample is 10 pds and 1000 individuals
TOBIT (censored) regression model
(Lower) censoring limit is .00
Simulation based on 20 Halton draws

Y	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
X1	1.00158***	.02412	41.52	.0000	.95430	1.04886
X2	.99160***	.02577	38.48	.0000	.94109	1.04211
Z1	-.02678	.02343	-1.14	.2531	-.07270	.01915
Z2	-.09841**	.04630	-2.13	.0335	-.18915	-.00766
	Means for random parameters					
Constant	.04182	.03546	1.18	.2383	-.02768	.11133
	Scale parameters for dists. of random parameters					
Constant	.98345***	.02441	40.29	.0000	.93561	1.03128
	Variance parameter given is sigma					
Std.Dev.	2.00468***	.02183	91.81	.0000	1.96189	2.04747

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point .7971
Scale Factor for Marginal Effects .4989

Y	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence Interval	
X1	.49972***	.00336	39.83	.0000	.47513	.52431
X2	.49474***	-.00202	37.34	.0000	.46877	.52070
Z1	-.01336	.1880D-04	-1.14	.2529	-.03626	.00954
Z2	-.04910**	-.03090	-2.08	.0371	-.09526	-.00293

E46.5 Random Parameters Models

We have extended the random parameters model to the censored regression (tobit) and truncated regression models. (Full details on the random parameters model appear in Chapter R24.) The structure of the random parameters model is based on the conditional density

$$f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i) \;=\; f(\boldsymbol{\beta}_i'\mathbf{x}_{it}),\; i = 1,...,N,\; t = 1,...,T_i.$$

where $f(.)$ is the hybrid continuous/discrete density for the tobit model. The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) mean

$$E[\boldsymbol{\beta}_i \mid \mathbf{z}_i] \;=\; \boldsymbol{\beta} \;+\; \boldsymbol{\Delta}\mathbf{z}_i,$$

(the second term is optional – the mean may be constant),

$$\text{Var}\,[\boldsymbol{\beta}_i \mid \mathbf{z}_i] \;=\; \boldsymbol{\Sigma}.$$

The model is operationalized by writing

$$\boldsymbol{\beta}_i \;=\; \boldsymbol{\beta} \;+\; \boldsymbol{\Delta}\mathbf{z}_i \;+\; \boldsymbol{\Gamma}\mathbf{v}_i.$$

As noted earlier, the heterogeneity term is optional. In addition, it may be assumed that some of the parameters are nonrandom. It is convenient to analyze the model in this fully general form here. One could easily accommodate nonrandom parameters just by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$.

NOTE: If there is no heterogeneity in the mean, and only the constant term is considered random – the model may specify that some parameters are nonrandom – then this model is equivalent to the random effects model of the preceding section.

E46.5.1 Command for the Random Parameters Model

The basic model command for this form of the model is

```
      TOBIT          ; Lhs  = dependent variable
      or TRUNCATE    ; Rhs  = independent variables
      or GROUPED     ; Pds  = fixed periods or count variable
                     ; RPM
or                   ; RPM = list of variables in z
                     ; Fcn  = random parameters specification $
```

The limit value and censoring in the lower tail may be changed for the tobit and truncated regression models with

```
                     ; Limit = the nonzero value
and/or               ; Upper censoring
```

This estimator only supports censoring in one tail.

NOTE: For this model, your Rhs list should include a constant term.

NOTE: The **; Panel** specification is optional. You may fit these models with cross section data. There is nothing inherent in the model that limits it to a panel data application. However, identification can be a bit weak for cross section estimation, and it will often break down. Using this model in a cross section is likely to be successful only when the data and the model are strongly consistent with each other.

Standard Model Specifications for the Random Parameters Truncated Regression Model

This is the full list of general specifications applicable to this model estimator.

Controlling Output from Model Commands.

; Par keeps individual specific parameter estimates.
; Margin displays marginal effects. Marginal effects are computed by setting the heterogeneity terms to their expected values of zero.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Robust sandwich estimator or robust VC for TSCS and some discrete choice.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name sets algorithm. The default (and best) algorithm for estimation is BFGS. But, all other algorithms are available.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E46.5.2 Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one, x1, x2, x3, x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be constant. For those that you wish to specify as random, use

; Fcn = variable name (distribution), variable name (distribution), ...

Numerous distributions may be specified. Three that are commonly used are

n = standard normal distribution, variance = 1,
t = triangular (tent shaped) distribution in [-1,+1], variance = 1/6,
u = standard uniform distribution [-1,1], variance = 1/3.

All random variables have mean zero. Note that each of these is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. The latter two distributions are provided as one may wish to reduce the amount of variation in the tails of the distribution of the parameters across individuals and to limit the range of variation. (See Train, 2010). Other options for the distributions of random parameters are described in Chapter R24. To specify that the constant term and the coefficient on *x*1 are normally distributed with fixed mean and variance, use

; Fcn = one(n), x1(n)

This specifies that the first and second coefficients are not random while the remainder are. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

NOTE: The model with only a random constant term (**; Fcn = one(n)**) is precisely equivalent to the random effects model of the previous section.

E46.5.3 Application

The following example illustrates. Note that this simulates the assumptions of the model. The first instruction fits the RP model with only a random constant term, which is the random effects model. The second fits the model with both constant and one slope random.

```
SAMPLE    ; All $
CREATE    ; yrps = (1+vi)*x1 + x2 + eit + ui
          ; yrp = max(0,yrps) $
TOBIT     ; Lhs = yrp ; Rhs = x
          ; RPM ; Fcn = one(n), x1(n) ; Correlated
          ; Pds = 10 ; Pts = 20 ; Halton $
```

The results include estimates of the means and standard deviations of the distributions of the random parameters and the estimates of the nonrandom parameters. The log likelihood shown is conditioned on the random draws, so one might be cautious about using it to test hypotheses, for example, that the parameters are random at all by comparing it to the log likelihood from the basic model with all nonrandom coefficients. The random constant term model shown earlier is mathematically equivalent to the random effects model. The results for the quadrature based estimator are shown earlier with the simulation based estimates. They are strikingly close, in spite of the small number of draws used for the simulations. The marginal effects shown are for the simulation estimator. The results below are estimates of the model with two random parameters, the constant and the slope on $x1$. The true values of the two variance parameters for the random parameters are roughly 1.15 for the constant and .25 for the coefficient on $x1$. The true means are zero and one, while the true coefficient on $x2$ is also one.

Correlated Random Parameters

The default RP command defines an estimator for a model in which the covariance matrix of the random parameters is diagonal. To extend it to a model in which the parameters are freely correlated, add

; Correlation (or just ; Cor)

to the command. The example below estimates a random parameters model with correlated random constant term and random slope.

```
-----------------------------------------------------------------------------
Random Coefficients  Tobit    Model
Dependent variable                 YRP
Log likelihood function    -13750.50869
Estimation based on N =  10000, K =   9
Inf.Cr.AIC  =  27519.0 AIC/N =    2.752
Sample is 10 pds and   1000 individuals
TOBIT (censored) regression model
(Lower) censoring  limit is     .00
Simulation based on  20 Halton draws
-----------------------------------------------------------------------------
```

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YRP|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Nonrandom parameters
      X2|     .98788***      .02551    38.72  .0000      .93788   1.03788
      Z1|    -.01670         .02320     -.72  .4717     -.06217    .02878
      Z2|    -.10816**       .04612    -2.34  .0190     -.19856   -.01776
        |Means for random parameters
Constant|     .06495*        .03480     1.87  .0620     -.00326    .13316
      X1|     .97941***      .02410    40.64  .0000      .93218   1.02664
        |Diagonal elements of Cholesky matrix
Constant|    1.02785***      .02620    39.23  .0000      .97650   1.07920
      X1|     .25809***      .02389    10.80  .0000      .21126    .30492
        |Below diagonal elements of Cholesky matrix
 lX1_ONE|    -.03295         .02608    -1.26  .2064     -.08407    .01816
        |Variance parameter given is sigma
Std.Dev.|    1.99475***      .02141    93.19  .0000     1.95280   2.03671
--------+-------------------------------------------------------------
Implied covariance matrix of random parameters
Var_Beta|             1             2
--------+----------------------------
       1|       1.05647     -.0338705
       2|     -.0338705      .0676965
Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       1.02785
       2|       .260185
Implied correlation matrix of random parameters
Cor_Beta|             1             2
--------+----------------------------
       1|       1.00000      -.126651
       2|      -.126651       1.00000
```

E46.5.4 Model Specifications

There are several additional model specifications and estimation controls that you can use with the random parameters model..

Heterogeneity in the Means

The preceding examples have specified that the mean of the random variable is fixed over individuals. If there is measured heterogeneity in the means, in the form of

$$E[\beta_{ki}] = \beta_k + \Sigma_m \delta_{km} z_{mi}$$

where z_m is a variable that is measured for each individual, then the command may be modified to

; RPM = list of variables in z

In the data set, these variables must be repeated for each observation in the group.

Controlling the Simulation

There are two parameters of the simulations that you can change. *R* is the number of points in the simulation. Authors differ in the appropriate value. Train (1999) recommends several hundred. Bhat suggests 1,000 is an appropriate value. The program default is 100. You can choose the value with

; Pts = number of draws, R

The value of 20 that we set in our experiments above was chosen purely to produce an example that you could replicate without spending an inordinate amount of waiting for the results.

The standard approach to simulation estimation is to use random draws from the specified distribution. As suggested immediately above, good performance in this connection requires very large numbers of draws. The drawback to this approach is that with large samples and large models, this entails a huge amount of computation and can be very time consuming. Some authors have documented dramatic speed gains with no degradation in simulation performance through the use of a small number of Halton draws instead of a large number of random draws. Authors (e.g., Bhat (1999)) have found that a Halton sequence of draws with only one tenth the number of draws as a random sequence is equally effective. To use this approach, add

; Halton

to your model command. The results below show the same model as estimated immediately above using five Halton draws instead of 20 simulated random draws. The estimates are essentially the same. The estimator based on the Halton sequences required roughly 20 seconds and 14 iterations to converge; the one based on the pseudorandom numbers required about 70 seconds and 15 iterations to reach the same estimates. (With ever faster computers, this consideration may ultimately be minor. However, we have of late heard from users who are employing data sets involving hundreds of thousands of observations. In a data set this large, use of the Halton sequences approach may produce a benefit worth pursuing.) Halton sequences are discussed in Section R24.7.

In order to replicate an estimation, you must use the same random draws. One implication of this is that if you give the identical model command twice in sequence, you will not get the identical set of results because the random draws in the sequences will be different. To obtain the same results, you must reset the seed of the random number generator with a command such as

CALC ; Ran(seed value) $

(Note that we have used **; Ran(12345)** before each of our examples above, precisely for this reason. The specific value you use for the seed is not of consequence; any odd number will do. You can also achieve replicability by using Halton sequences, which are not random, but are deterministic sequences.

E46.5.5 Model Estimates

Results saved by this estimator are:

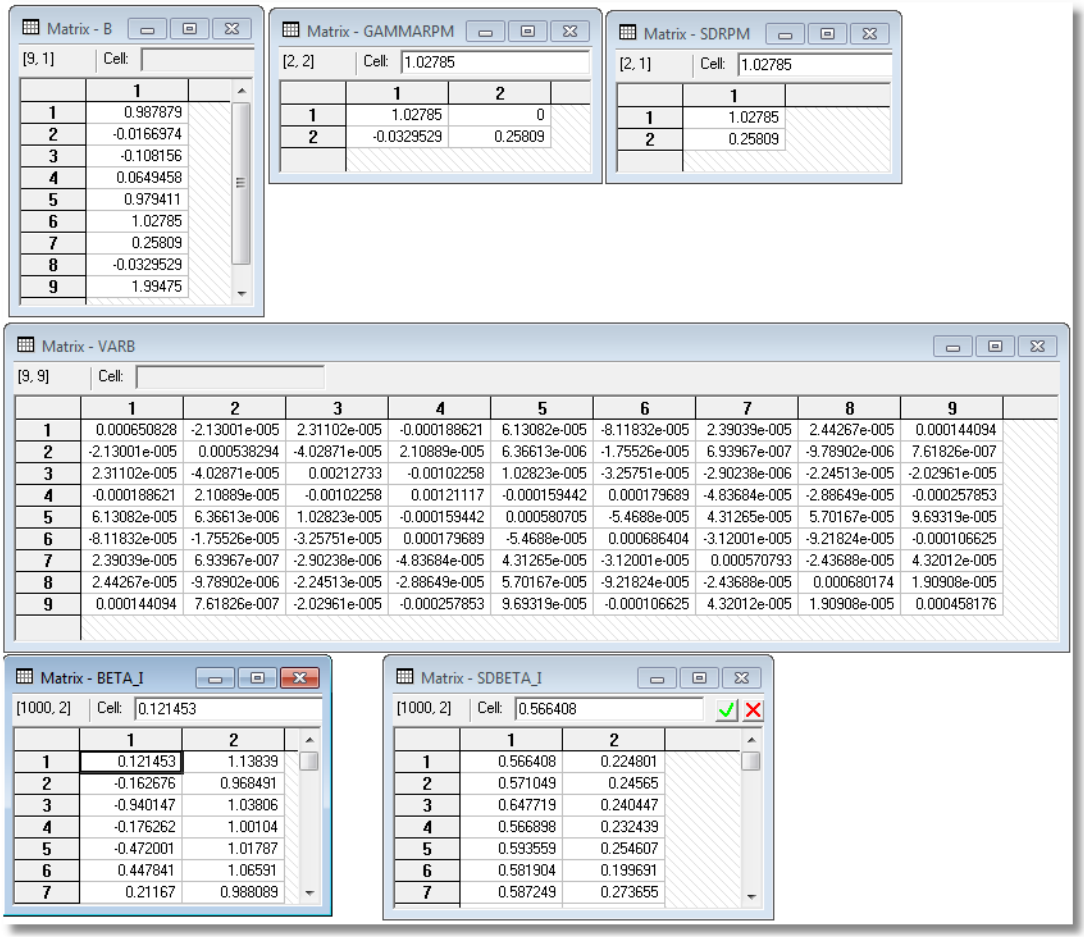
Matrices:	*b*	=	estimate of θ
	varb	=	asymptotic covariance matrix for estimate of θ
	beta_i	=	individual specific parameters, if **; Par** is requested

Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function

Last Model: *b_variables*

Last Function: None

Section R24.5 describes a method of estimating the conditional mean of the distribution from which β_i is drawn When you include **; Parameters** in your command, the matrices of conditional means and conditional standard deviations are kept with the output of the model. The matrices below are generated by the model command in the previous section. These are in addition to *b* and *varb* shown below. The matrix *sdrpm* saves the implied estimates of the standard deviations of the random parameters. These are reported with the output as a matrix (column vector) of implied standard deviations. The matrix *gammarpm* is the lower triangular matrix assembled from the estimated parameter vector. In the vector of estimated parameters (see vector *b*), the diagonal elements of Γ appear first, followed by the below diagonal element(s). Finally, *beta_i* and *sdbeta_i* are computed with a row for each individual.



Matrix - B [9, 1]

	1
1	0.987879
2	-0.0166974
3	-0.108156
4	0.0649458
5	0.979411
6	1.02785
7	0.25809
8	-0.0329529
9	1.99475

Matrix - GAMMARPM [2, 2] Cell: 1.02785

	1	2
1	1.02785	0
2	-0.0329529	0.25809

Matrix - SDRPM [2, 1] Cell: 1.02785

	1
1	1.02785
2	0.25809

Matrix - VARB [9, 9]

	1	2	3	4	5	6	7	8	9
1	0.000650828	-2.13001e-005	2.31102e-005	-0.000188621	6.13082e-005	-8.11832e-005	2.39039e-005	2.44267e-005	0.000144094
2	-2.13001e-005	0.000538294	-4.02871e-005	2.10889e-005	6.36613e-006	-1.75526e-005	6.93967e-007	-9.78902e-006	7.61826e-007
3	2.31102e-005	-4.02871e-005	0.00212733	-0.00102258	1.02823e-005	-3.25751e-005	-2.90238e-006	-2.24513e-005	-2.02961e-005
4	-0.000188621	2.10889e-005	-0.00102258	0.00121117	-0.000159442	0.000179689	-4.83684e-005	-2.88649e-005	-0.000257853
5	6.13082e-005	6.36613e-006	1.02823e-005	-0.000159442	0.000580705	-5.4688e-005	4.31265e-005	5.70167e-005	9.69319e-005
6	-8.11832e-005	-1.75526e-005	-3.25751e-005	0.000179689	-5.4688e-005	0.000686404	-3.12001e-005	-9.21824e-005	-0.000106625
7	2.39039e-005	6.93967e-007	-2.90238e-006	-4.83684e-005	4.31265e-005	-3.12001e-005	0.000570793	-2.43688e-005	4.32012e-005
8	2.44267e-005	-9.78902e-006	-2.24513e-005	-2.88649e-005	5.70167e-005	-9.21824e-005	-2.43688e-005	0.000680174	1.90908e-005
9	0.000144094	7.61826e-007	-2.02961e-005	-0.000257853	9.69319e-005	-0.000106625	4.32012e-005	1.90908e-005	0.000458176

Matrix - BETA_I [1000, 2] Cell: 0.121453

	1	2
1	0.121453	1.13839
2	-0.162676	0.968491
3	-0.940147	1.03806
4	-0.176262	1.00104
5	-0.472001	1.01787
6	0.447841	1.06591
7	0.21167	0.988089

Matrix - SDBETA_I [1000, 2] Cell: 0.566408

	1	2
1	0.566408	0.224801
2	0.571049	0.24565
3	0.647719	0.240447
4	0.566898	0.232439
5	0.593559	0.254607
6	0.581904	0.199691
7	0.587249	0.273655

Figure E28.2 Matrix Results for RP Censored Regression

E46.6 Latent Class Models

The tobit, truncated and grouped data regression models for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ are denoted

$$f(y_{it} \mid \mathbf{x}_{it}) \;=\; f(y_{it},\boldsymbol{\beta}'\mathbf{x}_{it}) \;=\; f(i,t).$$

Henceforth, we will use the term 'group' to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. Unobserved heterogeneity in the distribution of y_{it} is assumed to impact the density in the form of a random effect. The continuous distribution of the heterogeneity is approximated by using a finite number of 'points of support.' The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J 'classes' with a model which allows for heterogeneity as follows: The density of the observed y_{it} given that regime j applies is

$$f(i,t|j) \;=\; f(y_{it}|\, \mathbf{x}_{it}, j)$$

where the density is now specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$f(i,t|j) \;=\; f(y_{it},\, \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \delta_j),\ \text{Prob[class} = j] \;=\; F_{j.}$$

We formulate this approximation more generally as,

$$f(i,t|j) \;=\; f[y_{it} \mid \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \boldsymbol{\delta}_j'\mathbf{x}_{it},\, \sigma_j),$$

$$F_j \;=\; \exp(\theta_j) \,/\, \Sigma_j \exp(\theta_j),\ \text{with } \theta_J \;=0.$$

In this formulation, each group has its own parameter vector, $(\boldsymbol{\beta}_j',\sigma_j) = (\boldsymbol{\beta} + \boldsymbol{\delta}_j,\sigma_j)$ though the variables that enter the mean are assumed to be the same. (This can be changed by imposing restrictions on the full parameter vector, as described below.) This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters. (A further generalization is discussed below.) The class probabilities can also be extended in the form of a multinomial logit model;

$$\theta_{ij} \;=\; \boldsymbol{\theta}_j'\mathbf{z}_i$$

by adding the variable names for these variables in the command as shown below.

E46.6.1 Commands for Latent Class Models

The estimation command for this model is

TOBIT ; Lhs = ...
or TRUNCATE ; Rhs = independent variables
or GROUPED ; LCM = list of variables in z if desired (= list is optional)
; Pds = panel data specification
; Pts = number of latent classes $

The default number of support points is five. You may set *J* to 2, 3, ..., 9 with

; Pts = the value you wish

The limit value and censoring in the lower tail may be changed with

; Limit = the nonzero value
and/or **; Upper censoring or truncation**

This estimator only supports censoring in one tail. Other options are the standard ones for the tobit and truncation models. Some particular values computed for the latent class model are

; Parameters to keep the individual specific parameter estimates.
; Group = name to retain the index of the most likely latent class
; Cprob = name to retain the estimated probability for the most likely latent class

You can obtain a listing of these two results by using

; List

Other model specifications appear in the list below.

This estimator does not support restrictions with **; CML** or **; Test:** However, you can use the **; Rst = list** option to constrain the model, for example to structure the latent class model so that different variables appear in different classes or that classes have common parameters as in the Heckman and Singer form of the model. To use these options, note, first, that the structure of the parameter vector is as follows:

$$\beta_1,\sigma_1,\ \beta_2,\sigma_2,\ ...,\ \beta_J,\sigma_J,\ \theta_1,\ \theta_2,...,\theta_J.$$

That is, the model parameters for the classes appear first, followed by the structural parameters for the class probabilities. Note that θ_J will be set equal to zero by the program, but if you use **; Rst**, you must treat θ_J as a free parameter. The example below demonstrates. In the following, we set up a three class model in which the two slope parameters and the disturbance variance parameters are forced to the same in all three classes, but the constants differ. This would correspond to Heckman and Singer's formulation of a random effects model.

The commands are as follows:

```
TOBIT   ; Lhs = y
        ; Rhs = one,x1,x2
        ; LCM
        ; Pts  = 3
        ; Pds = 10
        ; Rst = a1,b1,b2,sigmav, a2,b1,b2,sigmav, a3,b1,b2,sigmav,
                theta1,theta2,theta3 $
```

Standard Model Specifications for the Latent Class Truncated Regression Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Margin displays marginal effects.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Robust requests a sandwich estimator or robust VC for TSCS and some discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E46.6.2 Model Estimates

Estimates retained by this model include

Matrices: *b* = full parameter vector, $[\boldsymbol{\beta}_1',\sigma_1, \boldsymbol{\beta}_2',\sigma_2, \ldots F_1,\ldots,F_J]$
varb = full covariance matrix
beta_i = individual specific parameters, if **; Par** is requested

Note that *b* and *varb* involve $J\times(K+2)$ estimates. Two additional matrices are created,

b_class = a $J\times K$ matrix with each row equal to the corresponding $\boldsymbol{\beta}_j$
class_pr = a $J\times 1$ vector containing the estimated class probabilities

Scalars: *kreg* = number of variables in Rhs list
nreg = total number of observations used for estimation
logl = maximized value of the log likelihood function
exitcode = exit status of the estimation procedure.

Last Function: None

E46.6.3 Applications

The first two sets of results fit three class latent class models to the simulated data used in the earlier examples. The third fits the Heckman and Singer random effects model noted earlier.

```
TOBIT      ; Lhs = y ; Rhs = one,x1,x2
           ; LCM
           ; Pts = 3
           ; Pds = 10 $
TRUNCATE   ; Lhs = y ; Rhs = one,x1,x2
           ; LCM
           ; Pts = 3
           ; Pds = 10 $
TOBIT      ; Lhs = y ; Rhs = x
           ; LCM
           ; Pts = 3
           ; Pds = 10
           ; Rst = b1,b2,b3,b4,a1,vv, b1,b2,b3,b4,a2,vv, b1,b2,b3,b4,a3,vv,
                   theta1,theta2,theta3 $
```

```
-----------------------------------------------------------------------------
Latent Class / Panel Tobit    Model
Dependent variable                    Y
Log likelihood function    -13657.65434
Estimation based on N =  10000, K =  20
Inf.Cr.AIC  =  27355.3 AIC/N =    2.736
Sample is 10 pds and   1000 individuals
TOBIT (censored) regression model
(Lower) censoring  limit is     .00
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
      X1|    1.05282***      .05571    18.90  .0000      .94363   1.16201
      X2|     .99677***      .06244    15.96  .0000      .87439   1.11914
      Z1|    -.14815**       .06345    -2.33  .0196     -.27251   -.02378
      Z2|    -.28628***      .10773    -2.66  .0079     -.49742   -.07514
Constant|    1.47990***      .11450    12.92  .0000     1.25549   1.70432
   Sigma|    2.04735***      .04916    41.64  .0000     1.95099   2.14371
        |Model parameters for latent class 2
      X1|     .98396***      .05065    19.42  .0000      .88468   1.08324
      X2|     .94552***      .05248    18.02  .0000      .84267   1.04838
      Z1|     .08412*        .04983     1.69  .0914     -.01355    .18179
      Z2|    -.02852         .09530     -.30  .7648     -.21530    .15827
Constant|     .10902         .14351      .76  .4475     -.17227    .39030
   Sigma|    1.94264***      .05097    38.11  .0000     1.84273   2.04255
        |Model parameters for latent class 3
      X1|     .97830***      .08225    11.89  .0000      .81709   1.13951
      X2|    1.10262***      .09133    12.07  .0000      .92362   1.28162
      Z1|    -.12957         .08407    -1.54  .1233     -.29435    .03521
      Z2|    -.01240         .15479     -.08  .9362     -.31577    .29098
Constant|   -1.31744***      .18876    -6.98  .0000    -1.68740   -.94748
   Sigma|    2.08057***      .08736    23.82  .0000     1.90935   2.25180
        |Estimated prior probabilities for class membership
Class1Pr|     .23211***      .03917     5.93  .0000      .15534    .30888
Class2Pr|     .47325***      .05732     8.26  .0000      .36090    .58560
Class3Pr|     .29464***      .06091     4.84  .0000      .17526    .41402
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Latent Class / Panel TruncReg Model
Dependent variable                    Y
Log likelihood function     -8030.07436
Estimation based on N =  10000, K =  20
Sample is 10 pds and   1000 individuals
Truncated regression model
(Lower) truncation limit is      .00
Model fit with  3 latent classes.

--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
      X1|     .92559***      .11991     7.72  .0000      .69057   1.16062
      X2|     .86473***      .10931     7.91  .0000      .65049   1.07897
      Z1|     .01226         .10300      .12  .9052     -.18961    .21414
      Z2|    -.25316         .20430    -1.24  .2153     -.65357    .14726
Constant|    1.75262***      .45167     3.88  .0001      .86736   2.63788
   Sigma|    2.00265***      .13186    15.19  .0000     1.74420   2.26109
        |Model parameters for latent class 2
      X1|    1.58660***      .24745     6.41  .0000     1.10160   2.07159
      X2|     .92187***      .21586     4.27  .0000      .49878   1.34495
      Z1|     .09177         .21943      .42  .6758     -.33830    .52184
      Z2|    -.43398         .40013    -1.08  .2781    -1.21822    .35026
Constant|     .95832**       .41865     2.29  .0221      .13777   1.77887
   Sigma|    1.26725***      .27840     4.55  .0000      .72159   1.81290
        |Model parameters for latent class 3
      X1|    1.03265***      .11910     8.67  .0000      .79921   1.26608
      X2|     .93355***      .10109     9.23  .0000      .73542   1.13169
      Z1|    -.01100         .08030     -.14  .8911     -.16839    .14639
      Z2|    -.04304         .17101     -.25  .8013     -.37822    .29213
Constant|    -.52724*        .27910    -1.89  .0589    -1.07427    .01978
   Sigma|    2.17506***      .11824    18.40  .0000     1.94332   2.40680
        |Estimated prior probabilities for class membership
Class1Pr|     .22782**       .10282     2.22  .0267      .02630    .42935
Class2Pr|     .06220         .05608     1.11  .2674     -.04772    .17213
Class3Pr|     .70997***      .11734     6.05  .0000      .47999    .93996
--------+--------------------------------------------------------------------
```

The following are the latent class estimates with variation only in the constant term. This model is comparable to the random effects model with continuous variation in the constant. The random effects model estimated earlier is shown below the latent class model. The estimates of the model parameters are strikingly similar. The similarity goes beyond that, however. After the results, we compute the standard deviation of the estimated random effects for the latent class model. Based on the three observations, the estimate is 0.9119797. The counterpart in the random effects model is 0.94842.

```
-----------------------------------------------------------------------------
Latent Class / Panel Tobit    Model
Dependent variable                    Y
Log likelihood function    -13662.84487
Estimation based on N =  10000, K =  10
Sample is 10 pds and   1000 individuals
TOBIT (censored) regression model
(Lower) censoring  limit is     .00
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
      X1|    1.00236***      .02494    40.19  .0000      .95348   1.05124
      X2|     .98747***      .02671    36.96  .0000      .93511   1.03983
      Z1|    -.02450         .02439    -1.00  .3151     -.07231    .02331
      Z2|    -.10470**       .04799    -2.18  .0291     -.19876   -.01065
Constant|    2.21773***      .26620     8.33  .0000     1.69598   2.73947
   Sigma|    2.00600***      .02309    86.87  .0000     1.96074   2.05126
        |Model parameters for latent class 2
      X1|    1.00236***      .02494    40.19  .0000      .95348   1.05124
      X2|     .98747***      .02671    36.96  .0000      .93511   1.03983
      Z1|    -.02450         .02439    -1.00  .3151     -.07231    .02331
      Z2|    -.10470**       .04799    -2.18  .0291     -.19876   -.01065
Constant|     .72316***      .09146     7.91  .0000      .54390    .90242
   Sigma|    2.00600***      .02309    86.87  .0000     1.96074   2.05126
        |Model parameters for latent class 3
      X1|    1.00236***      .02494    40.19  .0000      .95348   1.05124
      X2|     .98747***      .02671    36.96  .0000      .93511   1.03983
      Z1|    -.02450         .02439    -1.00  .3151     -.07231    .02331
      Z2|    -.10470**       .04799    -2.18  .0291     -.19876   -.01065
Constant|    -.86188***      .07386   -11.67  .0000    -1.00665   -.71712
   Sigma|    2.00600***      .02309    86.87  .0000     1.96074   2.05126
        |Estimated prior probabilities for class membership
Class1Pr|     .04677**       .02052     2.28  .0226      .00656    .08699
Class2Pr|     .47728***      .03267    14.61  .0000      .41325    .54130
Class3Pr|     .47595***      .03669    12.97  .0000      .40405    .54785
--------+--------------------------------------------------------------------
```

```
MATRIX    ; aj = b_class(1:3,5:5) $
MATRIX    ; aj2 = Dirp(aj,aj) $
MATRIX    ; pj = class_pr $
CALC      ; List ; meana = aj'pj ; sda = Sqr(aj2'pj - (aj'pj)^2) $
```

```
[CALC] MEANA   =       .0386644
[CALC] SDA     =       .9119797
```

E47: Limited Dependent Variable Models

E47.1 Introduction

The models and estimators described in this chapter are (numerous) variations on the following general structure:

Latent Underlying Regression: $y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \varepsilon_i \sim N[0,\sigma^2]$.

Observed Dependent Variable: if $y_i^* \leq L_i$, then $y_i = L_i$ (lower tail censoring)

if $y_i^* \geq U_i$, then $y_i = U_i$ (upper tail censoring)

if $L_i < y_i^* < U_i$, then $y_i = y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$.

Truncation, in which only data in the third group are observed, is a related case which is discussed in Section E47.4. In practice, most of the received applications involve censoring, rather than truncation. The thresholds, L_i and U_i, may be constants or variables. We accommodate censoring in the upper or lower (or both) tails of the distribution. The most familiar case of this model in the literature is the 'tobit' model, in which $U_i = +\infty$ and $L_i = 0$, i.e., the case in which the observed data contain a cluster of zeros. In the standard 'censored regression,' or tobit model, the censored range of y_i^* is the half of the line below zero. (For convenience, we will drop the observation subscript at this point.) If y^* is not positive, a zero is observed for y, otherwise the observation is y^*. Models of expenditure are typical. We also allow censoring of the upper tail ('on the right'). A model of the demand for tickets to sporting events might be an application, since the actual demand is only observed if it is not more than the capacity of the facility (stadium, etc.). A somewhat more elaborate specification is obtained when the range of y^* is censored in both tails. This is the 'two limit probit' model. An application might be a model of weekly hours worked, in which less than half time is reported as 20 and more than 40 is reported as 'full time,' i.e., 40 or more.

The preceding gives the basic model. We also allow for several variations, including a model with heteroscedasticity and models for panel data.

E47.2 Tobit Model

The basic tobit model corresponds to the specification in the Introduction. This model is developed in Chapter E45. The sections to follow show some extensions of the tobit model, including heteroscedasticity and two bivariate models.

E47.2.1 Heteroscedastic Tobit Model

The disturbance in the tobit model may be heteroscedastic so that the variance term is

$$\sigma_i = \sigma e^{\boldsymbol{\gamma}'\mathbf{z}_i}.$$

This is the model of multiplicative heteroscedasticity used in several earlier models. This model is requested with

```
TOBIT        ; Lhs  = y ; Rhs = list for x
             ; Het ; Hfn = list for z $
```

Limit specifications are as usual, upper (**; Upper**) or lower (default) censoring, and the limit value may be supplied with **; Limit = value** and all other parts of the command and options are the same as for the basic model.

> **NOTE:** Do not include *one* in the Hfn list. Since σ is a free parameter, including *one* will put a redundant constant in the variance model. This will cause a singular covariance matrix. (Previous versions of *LIMDEP* used Rh2 instead of Hfn in this specification. You may continue to use that syntax.)

The full parameter vector is now [**β**,**γ**,σ]. Use this setup if you are providing starting values with **; Start = list** or imposing restrictions with **; Rst = list** or **; CML: restrictions**. The results saved are: log likelihood, identification of limit values, configuration of parameter vector, estimates of [**β**,**γ**,σ], etc. The matrices *b* and *varb* will include the estimates of **γ**. As before, σ is the ancillary parameter. The specification **; Par** adds σ to the retained parameter vector. Finally, the *Last Model* parameters are [*b_variables_in_x*, *c_variables_in_z*].

Testing for Heteroscedasticity

The three familiar testing procedures are available for testing for heteroscedasticity in the tobit model. The following template shows how to apply the three procedures: We first set up the variables that appear in the model

```
NAMELIST    ; x = the full Rhs for the mean in the model $
NAMELIST    ; z = the variables in the variance function, does not include one $
CREATE      ; y = the dependent variable $
```

The dimensions will be needed for degrees of freedom and matrix manipulations.

```
CALC        ; kx = Col(x) ; kz = Col(z) $
```

This is the restricted, homoscedastic model.

```
TOBIT       ; Lhs = y ; Rhs = x ; (if necessary, set up limits specification) $
CALC        ; Lr = logl ; vr = s $
```

This does the LM test. The command sends in restricted estimates and does no iterations.

```
TOBIT       ; Lhs = y ; Rhs = x ; Rh2 = z ; Het
            ; Start = b, kz_0, vr ; Maxit = 0 $
```

This does the likelihood ratio test.

```
TOBIT       ; Lhs = y ; Rhs = x ; Rh2 = z ; Het ; Par $
CALC        ; lu = logl ; List ; chisq = 2*(lu - lr) ; signif = 1 - Chi(chisq,kz) $
```

We now do the Wald test.

```
CALC        ; kx1 = kx+1 ; kxz = kx+kz $
MATRIX      ; ghet = b(kx1:kxz) ; vghet = Varb(kx1:kxz, kx1:kxz)
            ; List ; waldstat = ghet'<vghet>ghet $
CALC        ; List ; signif = 1 - Chi(waldstat,kz) $
```

Partial Effects

Let w_i be a variable which can appear in either $\mathbf{x}_i$ or $\mathbf{z}_i$ or both. The marginal effects for $\mathbf{x}_i$ were given earlier;

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial w_i} = [\Phi_{Ui} - \Phi_{Li}]\beta_{\mathrm{w}} = [\text{Prob(uncensored region)}] \times \text{coefficient},$$

where Φ_{ji} is the probability associated with the censored regions, lower or upper. (See the technical details below.) For the terms in the variance function, we have the result

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial w_i} = \sigma_i[\phi_{Li} - \phi_{Ui}]\,\gamma_{w\,=}\,\sigma_i\,[\text{difference in densities at censoring points}]\times\text{coefficient}.$$

Now, let w_i be a variable which is assumed to appear both in $\mathbf{x}_i$ and $\mathbf{z}_i$. Then,

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial w_i} = [\Phi_{Ui} - \Phi_{Li}]\beta_w + (\phi_{Li} - \phi_{Ui})\sigma_i\gamma_w.$$

This provides a decomposition of the marginal effect. The decomposition is reported in the table of marginal effects. (See the example below and the derivations in the technical details below.) As always, marginal effects are requested with

```
            ; Partial Effects
```

Partial Effect for a Dummy Variable

The partial effect for a binary variable will be more involved than that for a continuous variable. Define the following components, where C is the dummy variable

$$E[y^*|\mathbf{x},C{=}1] = \boldsymbol{\beta}'\mathbf{x} + \delta,\ \mathrm{Var}[\varepsilon|\mathbf{z},C{=}1] = \sigma\exp(\boldsymbol{\gamma}'\mathbf{z} + \lambda) = \sigma_1$$

$$E[y^*|\mathbf{x},C{=}0] = \boldsymbol{\beta}'\mathbf{x}\ ,\ \mathrm{Var}[\varepsilon|\mathbf{z},C{=}0] = \sigma\exp(\boldsymbol{\gamma}'\mathbf{z}) = \sigma_0$$

We assume that C can enter either the mean function, the standard deviation, or both. To accommodate all cases, either δ or λ may be zero, but neither need be.

Now, let the censoring limits be L and U,

$$\alpha_L^1 = (L - \boldsymbol{\beta}'\mathbf{x} - \delta)/\sigma_1,\ \alpha_L^0 = (L - \boldsymbol{\beta}'\mathbf{x})/\sigma_0$$

$$\alpha_U^1 = (U - \boldsymbol{\beta}'\mathbf{x} - \delta)/\sigma_1,\ \alpha_U^0 = (U - \boldsymbol{\beta}'\mathbf{x})/\sigma_0$$

$$\Phi_j^m = \Phi(\alpha_j^m),\ j = L,U,\ m = 0,1$$

$$\phi_j^m = \phi(\alpha_j^m),\ j = L,U,\ m = 0,1$$

The conditional mean functions for the two cases, $C = 1$ and $C = 0$, are

$$E[y|\mathbf{x},\mathbf{z},C] = \Phi_L^m L + (1 - \Phi_U^m)U + (\Phi_U^m - \Phi_L^m)(\boldsymbol{\beta}'\mathbf{x} + m) + \sigma_m(\phi_L^m - \phi_U^m),\ m = 0,1$$

This does not simplify in any convenient way. Taking the difference,

$$\begin{aligned}\text{E}[y|\mathbf{x},\mathbf{z},C{=}1] - \text{E}[y|\mathbf{x},\mathbf{z},C{=}0] = &\ (\Phi_L^1 - \Phi_L^0)L + (\Phi_U^0 - \Phi_U^1)U \\ &+ \beta'\text{x}\,(\Phi_U^1 - \Phi_L^1 - \Phi_U^0 + \Phi_L^0) + \delta(\Phi_U^1 - \Phi_L^1) \\ &+ \sigma_1(\phi_L^1 - \phi_U^1) - \sigma_0(\phi_L^0 - \phi_U^0)\end{aligned}$$

The internal program invoked with **; Partials** in the command computes effects for dummy variables using the scaled coefficients, as if the variable were continuous. To obtain partial effects (at the means, or averaged over the data), you should use the **PARTIALS** command instead. The example below illustrates.

The following illustrates the testing procedures and computation of partial effects for the heteroscedasticity model. It computes the LM, LR and Wald tests, respectively.

```
CREATE      ; kids = (kl6+k618)>0 ; y = whrs $
NAMELIST    ; x = one,kl6,k618,wa,we  ; z = wa,kids $
CALC        ; kx = Col(x) ; kz = Col(z) $
TOBIT       ; Lhs = y ; Rhs = x  $
CALC        ; lr = logl ; vr = s $
TOBIT       ; Lhs = y ; Rhs = x ; Hfn = z ; Het
            ; Start = b, kz_0, vr ; Maxit = 0 $
CALC        ; List  ; signif = 1 - Chi(lmstat,(Col(z))) $
TOBIT       ; Lhs = y ; Rhs = x ; Hfn = z ; Het ; Par ; Partials $
NAMELIST    ; xz = x,kids $  (wa already appears in x)
PARTIALS    ; Effects : xz ; Summary $
CALC        ; lu = logl ; kx1 = kx+1 ; kxz = kx+kz
            ; List  ; lrstat = 2*(lu - lr) ; signif = 1 - Chi(lrstat,kz) $
MATRIX      ; ghet = b(kx1:kxz) ; vghet = varb(kx1:kxz, kx1:kxz) $
MATRIX      ; List ; waldstat = ghet’ <vghet> ghet $
CALC        ; List ; signif = 1 - Chi(waldstat,kz) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                        Y
Log likelihood function     -3904.16871
Estimation based on N =    753, K =   6
Threshold values for the model:
Lower =       .0000    Upper = +infinity
LM test [df] for tobit=     32.311[  5]
Normality Test, LM    =     10.355[  2]
ANOVA  based fit measure =      .049046
DECOMP based fit measure =      .165396
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Primary Index Equation for Model
Constant|    1320.87***    482.9241     2.74  .0062      374.36   2267.39
     KL6|   -1077.45***    126.2053    -8.54  .0000    -1324.81   -830.09
    K618|   -128.258***    42.74783    -3.00  .0027    -212.043   -44.474
      WA|   -41.5052***     7.70256    -5.39  .0000    -56.6019  -26.4084
      WE|    95.5038***    22.86314     4.18  .0000     50.6928  140.3147
        | Disturbance standard deviation
   Sigma|    1281.18***    48.18563    26.59  .0000     1186.74   1375.62
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Maximum of     0 iterations. Exit iterations with status=1.
Maxit = 0. Computing LM statistic at starting values.
No iterations computed and no parameter update done.
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                        Y
LM Stat. at start values        5.65988
LM statistic kept as scalar    LMSTAT
Log likelihood function     -3904.16871
Estimation based on N =    753, K =   8
Inf.Cr.AIC  =   7824.3 AIC/N =   10.391
Model estimated: Aug 02, 2011, 07:51:54
Threshold values for the model:
Lower =       .0000    Upper = +infinity
LM test [df] for tobit=     32.311[  5]
ANOVA  based fit measure =      .049046
DECOMP based fit measure =      .165396
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Primary Index Equation for Model
Constant|    1320.87***    493.6328     2.68  .0075      353.37   2288.38
     KL6|   -1077.45***    132.6381    -8.12  .0000    -1337.41   -817.48
    K618|   -128.258***    45.15824    -2.84  .0045    -216.767   -39.750
      WA|   -41.5052***     8.06388    -5.15  .0000    -57.3101  -25.7003
      WE|    95.5038***    24.45763     3.90  .0001     47.5677  143.4398
        | Heteroscedasticity Term
      WA|        0.0          .00664      .00 1.0000 -.13009D-01  .13009D-01
    KIDS|        0.0          .11419      .00 1.0000 -.22380D+00  .22380D+00
        | Disturbance standard deviation
   Sigma|    1281.18***    426.1212     3.01  .0026      446.00   2116.36
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

```
[CALC] SIGNIF  =       .0590164
--------+-------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                      Y
Log likelihood function     -3901.57772
LM test [df] for tobit=    191.958[  5]
ANOVA  based fit measure =      .039876
DECOMP based fit measure =      .070291
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        | Primary Index Equation for Model
Constant|    1436.37***    484.4913     2.96  .0030      486.79   2385.96
     KL6|   -1034.95***    119.7063    -8.65  .0000    -1269.57   -800.33
    K618|   -113.243***    43.13381    -2.63  .0087    -197.783   -28.702
      WA|   -44.2385***     7.90943    -5.59  .0000    -59.7406  -28.7363
      WE|    92.8179***    24.87012     3.73  .0002     44.0734  141.5625
        | Heteroscedasticity Term
      WA|     .00772         .00689     1.12  .2625     -.00578    .02122
    KIDS|    -.07185         .11555     -.62  .5341     -.29833    .15462
        | Disturbance standard deviation
   Sigma|    968.183***    325.5848     2.97  .0029     330.048  1606.317
--------+-------------------------------------------------------------------
Partial derivatives of expected value
--------+-------------------------------------------------------------------
        |     Partial     Standard            Prob.      95% Confidence
       Y|      Effect        Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        | Effect of variables in Xbeta (mean)
     KL6|   -612.113***    69.69413    -8.78  .0000    -748.711  -475.515
    K618|   -66.9762***    25.57280    -2.62  .0088   -117.0980  -16.8544
      WA|   -26.1644***     4.58136    -5.71  .0000    -35.1437  -17.1851
      WE|    54.8962***    14.80223     3.71  .0002     25.8844   83.9080
        | Effect of variables in exp(Zgamma) (variance)
      WA|    3.83349        3.48858     1.10  .2718    -3.00400  10.67099
    KIDS|   -35.6906       57.22086     -.62  .5328   -147.8414   76.4602
--------+-------------------------------------------------------------------

Partial Effects for Tobit (Censored) Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                     Partial    Standard
(Delta method)       Effect       Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      KL6        -579.06888   98.76835    5.86  -772.65129  -385.48647
      K618        -63.36060   25.57620    2.48  -113.48904   -13.23216
      WA          -22.20567    5.65882    3.92   -33.29675   -11.11460
      WE           51.93273   15.47267    3.36    21.60686    82.25860
   *  KIDS        -23.49918   50.69046     .46  -122.85065    75.85229
---------------------------------------------------------------------

[CALC] LRSTAT  =      5.1819841
[CALC] SIGNIF  =       .0749457
WALDSTAT|             1
--------+--------------
       1|       3.90215
[CALC] SIGNIF  =       .1421215
```

Technical Details for the Tobit Model with Heteroscedasticity

The parameters are not normalized by the Olsen transformation for this model. We let:

$$\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i,$$

$$\theta_i = \sigma\varepsilon^{\boldsymbol{\gamma}'\mathbf{z}_i},$$

$$d_{0i} = 1 \text{ if } y_i < L \text{ or } y_i > U, 0 \text{ otherwise},$$

$$d_{1i} = 1 - \mathrm{d}_0,$$

$$r_i = 1 \text{ and } w_i = (\boldsymbol{\beta}'\mathbf{x}_i - L_i)/\theta_i \text{ if } y_i \le L_i,$$

$$r_i = -1 \text{ and } w_i = (U_i - \boldsymbol{\beta}'\mathbf{x}_i)/\theta_\mathrm{i} \text{ if } y_i \ge U_i,$$

$$P_i = \Phi(-w_i)$$

$$\phi_i = \phi(-w_i).$$

Then,

$$\log L_i = d_{0i}\log \mathrm{P}_i + d_{1i}(-\log\theta_i - (\varepsilon_i/\theta_i)^2/2 - \log(2\pi)/2),$$

$$\partial\log L_i/\partial\boldsymbol{\beta} = [(d_{1i}\varepsilon_i/\theta_i - d_{0i}\, r_i\phi_i/\Phi_i)/\theta_i]\mathbf{x}_i,$$

$$\partial\log L_i/\partial\boldsymbol{\gamma} = [\mathrm{d}_{1i}((\varepsilon_i/\theta_i)^2 - 1) + d_{0i}w_i\phi_i/\Phi_i]\mathbf{z}_\mathrm{i},$$

$$\partial\log L_i/\partial\sigma = [\mathrm{d}_{1i}((\varepsilon_i/\theta_i)^2 - 1) + d_{0i}w_i\phi_i/\Phi_i]/\sigma.$$

The BHHH estimator, using the outer product of the gradients, is used to estimate the asymptotic covariance matrix of the estimates.

The marginal effects in this model are complicated a bit by the fact that variables may appear in both the mean and the variance. The conditional mean function in the fully general model is

$$E[y_i|\mathbf{x}_i,\mathbf{z}_i] = \Phi_{Li}\times L_i + (1-\Phi_{Ui})Ui + (\Phi_{Ui} - \Phi_{Li})\left(\boldsymbol{\beta}'\mathbf{x}_i + \sigma_i\frac{\phi_{Li} - \phi_{Ui}}{\Phi_{Ui} - \Phi_{Li}}\right)$$

where

$$\Phi_{Li} = \Phi\left(\frac{L_i - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right) = \Phi(\alpha_{Li}), \text{ and let } a_{Li} = \partial\alpha_{Li}/\partial\sigma_i = -\alpha_{Li}/\sigma_i$$

$$\Phi_{Ui} = \Phi\left(\frac{U_i - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right) = \Phi(\alpha_{Ui}), \text{ and let } a_{Ui} = \partial\alpha_{Ui}/\partial\sigma_i = -\alpha_{Ui}/\sigma_i$$

$$\sigma_i = \sigma e^{\boldsymbol{\gamma}'\mathbf{z}_i}.$$

As derived in Greene (1999), marginal effects for the variables in the mean are simple;

$$\partial\mathrm{E}[y_i|\mathbf{x}_i,\mathbf{z}_i]/\partial\mathbf{x}_i = (\Phi_{Ui} - \Phi_{Li})\boldsymbol{\beta}$$

But, $\partial E[y_i|\mathbf{x}_i,\mathbf{z}_i]/\partial\sigma_i$ is considerably more involved (at least it appears so). The desired result is

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial\sigma_i} = \phi_{Li}L_i a_{Li} - \phi_{ui}U_i a_{Ui} + \boldsymbol{\beta}'\mathbf{x}_i\phi_{Ui}a_{Ui} - \boldsymbol{\beta}'\mathbf{x}_i\phi_{Li}a_{Li} + \phi_{Li} - \phi_{Ui}$$
$$- \sigma_i\alpha_{Li}\phi_{Li}\times(-\alpha_{Li})/\sigma_i + \sigma_i\alpha_{Ui}\phi_{Ui}\times(-\alpha_{Ui})/\sigma_i.$$

Collecting terms, and recalling the definitions of α_{Li} and α_{Ui} produces the striking result

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial \sigma_i} = \phi_{Li} - \phi_{Ui}$$

Now, let w_i be a variable which is assumed to appear both in $\mathbf{x}_i$ and $\mathbf{z}_i$. Then,

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{z}_i]}{\partial w_i} = [\Phi_{Ui} - \Phi_{Li}]\beta_w + (\phi_{Li} - \phi_{Ui})\sigma_i\gamma_w$$

Analytic results for the standard errors of the marginal effects are complicated considerably by the presence of the estimated ancillary parameter, σ. To simplify matters, let $\gamma_0 = \log\sigma$ and add a constant (one) to $\mathbf{z}_i$. Include this parameter in γ so that now, $\sigma_i = \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$. The full parameter vector is now $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\gamma}']'$. In preparation for this, the entire last row of the $(K + L + 1)\times(K + L + 1)$ asymptotic covariance matrix for the directly estimated parameters, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ (without γ_0), and σ, is multiplied by $1/\sigma$, then the last diagonal element is multiplied by $1/\sigma$ again. This gives us a parameter vector and asymptotic covariance matrix that are conveniently partitioned into two parts. With all this in place, we now obtain the estimates of the asymptotic standard errors for the marginal effects using the delta method. The marginal effects are

$$\boldsymbol{\delta}_{\mathbf{x}} = [\Phi_U - \Phi_L] \times \boldsymbol{\beta}$$

$$\boldsymbol{\delta}_{\mathbf{z}} = [\phi_L - \phi_U]\sigma \times \boldsymbol{\gamma}$$

(Observation subscripts are dropped for the moment. Recall, $\sigma = \exp(\boldsymbol{\gamma}'\mathbf{z})$. In order to use the delta method, we will require the following derivatives - the tedious algebra is omitted;

$$\frac{\partial \boldsymbol{\delta}_{\mathbf{x}}}{\partial \boldsymbol{\beta}'} = (\Phi_U - \Phi_L)\mathbf{I} - \frac{1}{\sigma}(\phi_U - \phi_L)\boldsymbol{\beta}\mathbf{x}'$$

$$\frac{\partial \boldsymbol{\delta}_{\mathbf{x}}}{\partial \boldsymbol{\gamma}'} = (\phi_L \alpha_L - \phi_U \alpha_U)\boldsymbol{\beta}\mathbf{z}'$$

$$\frac{\partial \boldsymbol{\delta}_{\mathbf{z}}}{\partial \boldsymbol{\beta}'} = (\phi_L \alpha_L - \phi_U \alpha_U)\boldsymbol{\gamma}\mathbf{x}'$$

$$\frac{\partial \boldsymbol{\delta}_{\mathbf{z}}}{\partial \boldsymbol{\gamma}'} = \sigma(\phi_L - \phi_U)(\mathbf{I} + \boldsymbol{\gamma}\mathbf{z}') + \sigma\left(\alpha_L^2\phi_L - \alpha_U^2\phi_U\right)\boldsymbol{\gamma}\mathbf{z}'$$

Collect these in the matrix $\mathbf{G}$ which is now partitioned conformably with the estimated asymptotic covariance matrix for the parameter estimates, $\mathbf{V}$. The estimated asymptotic covariance matrix for the marginal effects is then $\mathbf{GVG}'$. At completion, the last row and column, corresponding to the scale parameter, σ, are discarded. Finally, for variables which appear in both $\mathbf{x}$ and $\mathbf{z}$, the marginal effect is the sum. The estimated asymptotic variance for such a variable is simply the sum of the two estimated variances plus twice the estimated covariance.

E47.2.2 Bivariate and Nested Tobit Models

The equations of a bivariate tobit model would be

$$y_1^* = \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1$$

$$y_1 = \text{Maximum}(y_1^*, 0) \quad \text{(the usual tobit specification)}$$

$$y_2^* = \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2$$

$$y_2 = \text{Maximum}(y_2^*, 0) \quad \text{(the usual tobit specification)}$$

$$\varepsilon_1,\varepsilon_2 \sim N[0,0,\sigma_1^2,\sigma_2^2,\rho], \text{ covariance is } \sigma_{12} = \rho\sigma_1\sigma_2.$$

The parameters of the bivariate model may be estimated by full information maximum likelihood (FIML). The nested tobit variant of this model, as specified in Lee (1992) and Howe, et al. (1994) is another form of sample selection model: The model is defined by the additional specification

$$y_2,\mathbf{x}_2 \text{ observed only when } y_1 > 0.$$

The command for the bivariate tobit model is

BTOBIT ; Lhs = y1,y2 ; Rh1 = ... x1 ... ; Rh2 = ... x2 ... $

and for nested tobit model, it is

NTOBIT ; Lhs = y1,y2 ; Rh1 = ... x1 ... ; Rh2 = ... x2 ... $

The parameter vector for both models is $\boldsymbol{\theta} = [\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma_1,\sigma_2,\rho]$. The default starting values for the iterations are OLS as usual for the tobit model, and zero for ρ. You may provide your own starting values with **; Start = list** and impose within equation restrictions on the parameters with **; Rst = list**. The limit points for both equations in this model must be zero. The **; Limits = ...** specification is not used, and is ignored if present.

The usual output and optimization options are available for this model, however, neither fitted values (**; Keep**, **; List**, **; Res**) nor marginal effects (**; Partial Effects**) are computed. The retrievable results are $b = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ and *varb*. The specification **; Par** adds (σ_1,σ_2,ρ) to b and *varb*. The scalars are $kreg = k_1+k_2+3$, *nreg*, *logl*, *sigma*1, *sigma*2, *rho*, and *exitcode*. The *Last Model* labels for **WALD** are *b*1*_variables*, *b*2*_variables*, *sigma*1, *sigma*2, *r*12.

Technical Details

We use the Olsen normalization, $\boldsymbol{\gamma}_1 = \boldsymbol{\beta}_1/\sigma_1$, $\eta_1 = 1/\sigma_1$, $\boldsymbol{\gamma}_2 = \boldsymbol{\beta}_2/\sigma_2$, $\eta_2 = 1/\sigma_2$. The remaining parameter is $\rho = \text{Corr}[\eta_1\varepsilon_1,\eta_2\varepsilon_2]$. During estimation, we use the transformation of ρ, $\tau = \log((1+\rho)/(1-\rho))$. This transformed parameter ranges over the entire real line, so the parameter cannot go out of bounds during the iterations. Internally, ρ is obtained as $\rho = [\exp(\tau)-1]/][\exp(\tau)+1]$. Derivatives are modified accordingly. However, note that technical output shown during iterations will display τ, nor ρ.

Let

$$\delta = 1/(1 - \rho^2)^{1/2},$$

$$\varepsilon_1 = \eta_1 y_1 - \boldsymbol{\gamma}_1'\mathbf{x}_1,$$

$$\varepsilon_2 = \eta_2 y_2 - \boldsymbol{\gamma}_2'\mathbf{x}_2$$

Then, the log likelihood function for the nested tobit model is

$$\begin{aligned}\log L \quad = \quad & \Sigma_{y1=0}\log[1 - \Phi(\boldsymbol{\gamma}_1'\mathbf{x}_1)] \\ & + \Sigma_{y1=1,y2=0} \log\{\eta_1\phi(\eta_1 y_1 - \boldsymbol{\gamma}_1'\mathbf{x}_1)[1 - \Phi(\delta(\boldsymbol{\gamma}_2'\mathbf{x}_2 + \rho\varepsilon_1))]\} \\ & + \Sigma_{y1=1,y2=1} -\log 2\pi + \log(\eta_1\eta_2\delta - 2\delta^2(\varepsilon_1^2 + \varepsilon_2^2 - 2\rho\varepsilon_1\varepsilon_2).\end{aligned}$$

The third term is the log of the density of the bivariate normal distribution. Derivatives can be obtained from results above. The BHHH estimator is used for the asymptotic covariance matrix. For the bivariate tobit model, there are four cells. The second term above is accompanied by a counterpart which reverses the role of the two variables while the first is replaced with the joint probability of two limit observations. The result is

$$\begin{aligned}\log L \quad = \quad & \Sigma_{y1=0,y2=0}\log[\Phi_2(-\boldsymbol{\gamma}_1'\mathbf{x}_1, -\boldsymbol{\gamma}_2'\mathbf{x}_2, \rho)] \\ & + \Sigma_{y1>0,y2=0} \log\{\eta_1\phi(\eta_1 y_1 - \boldsymbol{\gamma}_1'\mathbf{x}_1)[1 - \Phi(\delta(\boldsymbol{\gamma}_2'\mathbf{x}_2 + \rho\varepsilon_1))]\} \\ & + \Sigma_{y1=0,y2>0} \log\{\eta_2\phi(\eta_2 y_2 - \boldsymbol{\gamma}_2'\mathbf{x}_2)[1 - \Phi(\delta(\boldsymbol{\gamma}_1'\mathbf{x}_1 + \rho\varepsilon_2))]\} \\ & + \Sigma_{y1=1,y2=1} -\log 2\pi + \log(\eta_1\eta_2\delta - 2\delta^2(\varepsilon_1^2 + \varepsilon_2^2 - 2\rho\varepsilon_1\varepsilon_2),\end{aligned}$$

where Φ_2 denotes the CDF for the bivariate standard normal distribution.

NOTE: Because of the use of the Olsen transformation, it is not possible to impose cross equation equality restrictions in this model. In principle they may be imposed, but equality of scaled (by σ_j) coefficients does not imply equality of the original coefficients.

Application

To illustrate this model, we have fit a bivariate tobit model for the wife's and husband's hours in the labor supply data. The command is

BTOBIT ; Lhs = whrs,hhrs ; Rh1 = one,kl6,k618
; Rh2 = one,ha,he,faminc,kids $

```
-----------------------------------------------------------------------------
Maximum likelihood ests.: Bivariate Tobit
First  equation LHS variable: Y1 = WHRS
Second equation LHS variable: Y2 = HHRS
Estimation based on N =    753, K =  11
Inf.Cr.AIC  =  19611.7 AIC/N =   26.045
Nonlimit observations: WHRS    --   428.0
Nonlimit observations: HHRS    --   753.0
--------+--------------------------------------------------------------------
    WHRS|                  Standard            Prob.      95% Confidence
    HHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Equation (RHS) for WHRS
Constant|    545.371***    74.42661     7.33  .0000     399.498   691.245
     KL6|   -775.299***   121.8240     -6.36  .0000   -1014.069  -536.528
    K618|   -40.9340       42.04329     -.97  .3302   -123.3374   41.4693
        |Equation (RHS) for HHRS
Constant|    2300.79***   211.2934     10.89  .0000     1886.66   2714.92
      HA|   -6.88365**      3.39381    -2.03  .0425   -13.53540   -.23190
      HE|    8.61231        7.69781     1.12  .2632    -6.47512  23.69975
  FAMINC|     .00619***      .00193     3.21  .0013      .00241    .00997
    KIDS|    38.2482       57.43590      .67  .5055    -74.3241  150.8205
        |          Disturbance Variances and Correlation
Sigma(1)|    1325.83***    56.71038    23.38  .0000     1214.68   1436.98
Sigma(2)|    586.483***    11.51530    50.93  .0000     563.913   609.052
RHO(1,2)|    -.10319**      .04170    -2.47  .0133     -.18491   -.02146
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E47.3 Categorical (Grouped) Data

A special case of the censored data regression model arises when the range of the dependent variable is completely censored. This is the case when data are reported only by interval category. For example, income data might be reported only by range. We assume that the finite (internal) terminal points are known. The dependent variable is coded $y = 1, 2, ..., J$ (not 0,..., as in the case of the probability models). For example, consider a survey of incomes, which reports ranges:

$$y = \begin{cases} 1 \text{ if } & y^* < \$\,15{,}000, \\ 2 \text{ if } \$\,15{,}000 \le & y^* < \$\,30{,}000, \\ 3 \text{ if } \$\,30{,}000 \le & y^* < \$\,50{,}000, \\ 4 \text{ if } \$\,50{,}000 \le & y^* < \$\,75{,}000, \\ 5 \text{ if } & y^* \ge \$\,75{,}000. \end{cases}$$

Formally, the model is

$$\text{(unobserved)} \quad y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \ \varepsilon \sim N[0,\sigma^2],$$

$$\text{(observed)} \quad y = j \ \text{if} \ A_{j-1} \le y^* < A_j, j = 1,...,J, A_0 = -\infty, A_J = +\infty.$$

The difference between this and the ordered probit model of Chapter E34 is that the threshold values are known here. Since that is true, there is information on the scale of y^* in the data. Hence an estimate of σ is produced. It is not necessary to normalize it to 1.0.

The command for the grouped data model is

```
GROUPED DATA        ; Lhs = dependent variable
(or just GROUPED)   ; Rhs = regressors
                    ; Limits = a1,a2,...,aJ-1 $
```

The limit points may be constants or variables. If your data are in J groups, there will always be exactly J-1 interior limit values, which must be given in increasing order. If the match is not found, the estimator aborts. The data will be inspected to determine the value of J. In addition, if there are any empty cells (i.e., intermediate values of y which are never observed), a diagnostic is given and the estimation is discontinued. The limits are also checked. If they are not in ascending order for every observation, it is necessary to stop the estimation. For the earlier example, the command would be

```
GROUPED DATA        ; Lhs = y
                    ; Rhs = ... regressors
                    ; Limits = 5000,7000,10000,15000 $
```

In this case, there are five values of the dependent variable, so four limit values are given.

The command is otherwise identical to the **TOBIT** command, and the other options (fitted values, restrictions, starting values, iteration controls, and so on) are the same. Output is likewise the same. Since the dependent variable is not observed, there is no obvious conditional mean function. As such, there are no marginal effects for this model. You can request a listing of predictions of a sort with **; List**. Let L_i and U_i denote the lower and upper limits of the range indicated by the observed y_i. Thus, if y_i equals one, L_i is -∞ and U_i is A_1, the first limit value given. The conditional mean function is then the expected value of y^* in this range, which is the same as that for the truncated regression model,

$$E[y^*|\mathbf{x}_i, L_i < y^* < U_i] = \boldsymbol{\beta}'\mathbf{x}_i + \sigma_i \frac{\phi_L - \phi_U}{\Phi_U - \Phi_L},$$

where

$$\alpha_j = (j - \boldsymbol{\beta}'\mathbf{x})/\sigma, j = L,U$$

$$\Phi_j = \Phi(\alpha_j)$$

$$\phi_j = \phi(\alpha_j).$$

The results displayed for this model are the same as for the tobit model including OLS results if you request them with **; OLS**, the iterations, then, the log likelihood, endpoints of all intervals, estimates of [$\boldsymbol{\beta}$,σ], and so on. The retrievable results (matrices *b* and *varb*, scalars, and *Last Model* labels) for this estimator are also the same as for the tobit model with the creation of an additional matrix,

limits = limit values, including large values for the outside limits ($-\infty$ and $+\infty$).

This matrix can be used in subsequent **GROUPED DATA** commands, for example, if you are using the same Lhs variable and just changing the specification on the right hand side.

NOTE: The OLS starting values are obtained by a crude transformation of the dependent variable. For $y = 1$, $y' = A_1$; if $y = J$, $y' = A_{J-1}$. For other values, y' is the average of the two bracketing limit values. Then, y' is regressed on the Rhs variables. This will produce a coefficient vector that has the same order of magnitude as the MLE.

E47.3.1 Grouped (Categorical) Panel Data

LIMDEP's full menu of panel data estimators is available for the categorical data regression model. (Full documentation on the modeling frameworks appears in Chapter R24 and below for the tobit model.) To estimate the model, you must provide the starting values, which you should do, in all cases, by first fitting the model with no individual effects. Thus, your command for this model will appear as

GROUPED ; Lhs = ... ; Rhs = ... ; Limits = the set of limits as described above $
GROUPED ; Lhs = ... ; Rhs = ... ; Limits = the set of limits as described above
; Pds = ... the specification of the panel structure

plus exactly one of

; FEM for the fixed effects model
or **; RPM ; Fcn = ... specification for the random parameters model**
or **; LCM ; Pts = J for the latent class model $**

Other parts of the specification for the categorical data model are the same as for other models of this type, e.g., tobit and truncation, that are documented elsewhere in this chapter.

E47.3.2 Heteroscedasticity

LIMDEP's generic formulation for heteroscedasticity,

$$\sigma_i = \sigma \times \exp(\boldsymbol{\delta}'\mathbf{z}_i)$$

is supported for the grouped data (interval censored) regression model. The option is requested with

; Het ; Hfn = list

Since the basic scale parameter σ is maintained, it plays the role of the constant term in the variance model, so your **; Hfn** list should not contain *one*.

E47.3.3 Grouped Data and Sample Selection

The grouped data model is also extended to the sample selection treatment. (This model is developed in Bhat (1994).) The model is as above with the added feature that data for the primary model are observed (or not) nonrandomly via a Heckman style selection equation. The model is as follows:

$$
\begin{aligned}
y^* &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim N[0,\sigma^2],\\
y &= j \text{ if } A_{j-1} \le y^* < A_j,\ j = 1,\ldots,J,\ A_0 = -\infty,\ A_J = +\infty,\\
d^* &= \boldsymbol{\alpha}'\mathbf{z} + u,\\
d &= 1 \text{ if } d^* > 0 \text{ and } 0 \text{ otherwise},\\
[\varepsilon,u] &\sim N_2[0,0,\sigma^2,1,\rho],\\
[y,\mathbf{x}] &\quad \text{are observed only when } d = 1.
\end{aligned}
$$

The correlation between ε and u is ρ. The selection aspect of the model arises when ρ is not equal to zero. Note that this extension is the same as its counterpart discussed below for the tobit model.

The command is

```
GROUPED     ; Lhs = y,d
            ; Rh1 = variables in x
            ; Rh2 = variables in z
            ; Limits = a1, a2,...,aJ-1 $
```

The **GROUPED DATA** command is exactly the same as in the nonselected case. As before, you give only the interior limit points. The difference is the specification of the probit equation by the second Lhs variable and the Rh2 list. (Since this model proceeds directly to the MLE, we do not begin with a separate **PROBIT** command, as we do with most other sample selection models.)

The usual options are available, including fitted values, residuals, optimization controls, etc., with two exceptions. First, the **; Partial Effects** option is not supported for this model. Second, the default algorithm is BFGS, and this cannot be changed. In addition, you may impose within equations restrictions with the **; Rst = list** option.

> **NOTE:** Cross equation restrictions in this model are problematic, because the model fits the transformed parameters, not the original ones. Unless you force the two standard deviations to be equal, it is not possible to force coefficients in the two equations to be equal.

The fitted values for this model are computed using Bhat's results: Let

$$\delta = 1/(1 - \rho^2)^{1/2},$$

$$\eta = 1/\sigma,$$

$$q = \boldsymbol{\alpha}'\mathbf{z}$$

$$W_m = \eta A_m - \boldsymbol{\gamma}'\mathbf{x},\ m = L,U \text{ (limits for the range in which } y \text{ falls)}$$

$$V_m = \delta(q + \rho W_m),\ m = L,U$$

$$T_m = \delta(W_m + \rho q),\ m = L,U.$$

Then

$$E[y^*|\mathbf{x},d=1] = \boldsymbol{\beta}'\mathbf{x} + \sigma \frac{\phi(W_L)\Phi(V_L) - \phi(WU)\Phi(V_U) + \rho\phi(q)[\Phi(T_U) - \Phi(T_L)]}{\Phi_2(W_U, q, -\rho) - \Phi_2(W_L, q, -\rho)}$$

The retrievable results from this model are

Matrices: *b*, *varb*; use **; Par** to add (σ,ρ) to the parameter vector

Scalars: *s*, *rho*, *logl*, *kreg*, *nreg*, *ybar*, *sy*, *exitcode*

Last Model: *b_variables* = elements of $\boldsymbol{\beta}$
a_variables = elements of $\boldsymbol{\alpha}$, *sigma*, *r*12

The grouped data model with sample selection is developed further in Section E47.3.3. The mathematical background and an application are presented there as well.

E47.3.4 Application

To illustrate the model, the dependent variable in the tobit hours equation was recoded as follows:

```
CREATE     ; whrss = whrs $
RECODE     ; whrss ; 0/600 = 1 ; 600.1/1000 = 2 ; 1000.1/1500 = 3
           ; 1500.1/2000 = 4 ; * = 5 $
NAMELIST   ; x = one,kl6,k618,wa,we $
```

Then,

```
GROUPED    ; Lhs = whrss
           ; Rhs = x
           ; List
           ; Limits = 600,1000,1500,2000 $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                  WHRSS
Log likelihood function        -935.28872
Estimation based on N =    753, K =   6
Inf.Cr.AIC  =   1882.6 AIC/N =    2.500
Censoring Thresholds for the  5 cells:
 y   Lower    Upper y    Lower     Upper
 1 *******  600.00  2  600.00 1000.00
 3 1000.00 1500.00  4 1500.00 2000.00
 5 2000.00 *******
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   WHRSS|  Coefficient       Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    1588.32***    478.3766     3.32  .0009     650.72   2525.92
     KL6|   -1022.09***    139.5848    -7.32  .0000   -1295.67   -748.51
    K618|   -160.894***    43.99771    -3.66  .0003   -247.128   -74.660
      WA|   -35.3833***     7.70697    -4.59  .0000   -50.4887  -20.2779
      WE|    63.2289***    22.58890     2.80  .0051    18.9554  107.5023
        |Disturbance standard deviation
   Sigma|    1180.16***    61.29222    19.25  .0000    1060.03   1300.29
--------+--------------------------------------------------------------------
```

E47.3.5 Technical Details for the Grouped Data Regression Models

Optimization is the same as for **TOBIT**. All options, including **; Maxit**, **; Tlf**, **; Start**, **; Rst**, etc. operate the same. Olsen's transformation is used during the iterations. The log likelihood function for the grouped data model is

$$\log L \;=\; \Sigma_i\{\log[\Phi(\eta U - \boldsymbol{\gamma}'\mathbf{x}_i) - \Phi(\eta L - \boldsymbol{\gamma}'\mathbf{x}_i)]\}$$

where $\boldsymbol{\gamma} \;=\; \boldsymbol{\beta}/\sigma$ and $\eta \;=\; 1/\sigma$.

For this case, U is the upper limit of the range in which y_i falls, and L is the lower limit. Gradients and Hessians for these can be derived using the results shown earlier for the tobit model, as the terms are identical. The second derivatives are used in estimating the asymptotic covariance matrix for the estimates.

$$\partial \text{logL}/\partial(\boldsymbol{\gamma},\eta) \;=\; \sum_{i=1}^{n} \frac{1}{\Phi_U - \Phi_L}\left[\phi_U\begin{pmatrix}-\mathbf{x}_i\\ U\end{pmatrix} - \phi_L\begin{pmatrix}-\mathbf{x}_i\\ L\end{pmatrix}\right]$$

Let $\lambda_m \;=\; \phi_{\text{m}} \,/\, [\Phi_{\text{U}} - \Phi_{\text{L}}],\; m = L,U$

and $w_m \;=\; [-\mathbf{x},\, m]',\; m = L,U$

Then, $$\frac{\partial^2 \log L}{\partial(\boldsymbol{\gamma},\eta)\partial(\boldsymbol{\gamma},\eta)'} = \sum_{i=1}^{n} \{\lambda_U\mathbf{w}_U\,[(-\alpha_U - \lambda_U)\mathbf{w}_U' + \lambda_L\mathbf{w}_L']\} - \{\lambda_L\mathbf{w}_L\,[(-\alpha_L + \lambda_L)\mathbf{w}_L' - \lambda_U\mathbf{w}_U']\}.$$

E47.4 The Truncated Regression Model

The truncated regression model applies to the nonlimit observations in the tobit formulation. For the basic model with lower truncation at zero, the density is

$$f(y \mid \mathbf{x}) \quad = f(y^* \mid \mathbf{x}) \text{ if } y^* > 0.$$

But, $\quad \text{Prob}(y^* > 0 \mid \mathbf{x}) = \text{Prob}(\boldsymbol{\beta}'\mathbf{x} + \varepsilon > 0),\ \varepsilon \sim N[0,\sigma^2]$

$$= \text{Prob}(\varepsilon > -\boldsymbol{\beta}'\mathbf{x})$$

$$= \Phi[\boldsymbol{\beta}'\mathbf{x}/\sigma].$$

Therefore, $\quad f[y \mid \mathbf{x}] = (1/\sigma)\phi[(y - \boldsymbol{\beta}'\mathbf{x})/\sigma] \,/\, \Phi[\boldsymbol{\beta}'\mathbf{x}/\sigma].$

This is not a linear regression model. For a general specification in which the range of the variable is truncated in (possibly) both tails, the conditional mean function is

$$E[y_i|\mathbf{x}_i\,, L_i \le y_i \le U_i\,] = \boldsymbol{\beta}'\mathbf{x}_i + \sigma\frac{\phi_L - \phi_U}{\Phi_U - \Phi_L},$$

where the specification of the limits is the same as that for the tobit model and

$$\alpha_{limit} = (limit - \boldsymbol{\beta}'\mathbf{x})/\sigma,\ limit = U_i \text{ or } L_i$$

$$\Phi_{limit} = \Phi(\alpha_{limit})$$

$$\phi_{limit} = \phi(\alpha_{limit})$$

Least squares is inconsistent. The estimator used here is maximum likelihood. As with the tobit model, we allow truncation in either or both tails of the distribution, and the truncation points may be constants or variables. The specifications are identical to that for the tobit model described in Section E47.2 save for the command name:

```
TRUNCATION ; Lhs = y
           ; Rhs = covariates $
```

This specifies the default case of lower truncation at zero. Alternative specifications of the truncation limits use

; Limit = value or name for a different lower truncation limit
; Upper (only) for *upper* truncation at 0 (all observed *y*s will be negative)
; Limit = value ; Upper for a different, upper truncation limit
; Limits = lower, upper specification for truncation in both tails

E47.4.1 Truncated Regression for Panel Data

LIMDEP's full menu of panel data estimators is available for the categorical data regression model. (Full documentation on the modeling frameworks appears in Chapter E46 and below for the tobit model.) To estimate the model, you must provide the starting values, which you should do, in all cases, by first fitting the model with no individual effects. Thus, your command for this model will appear as

GROUPED ; Lhs = ... ; Rhs = ... ; Limits = the set of limits as described above $
GROUPED ; Lhs = ... ; Rhs = ... ; Limits = the set of limits as described above
; Pds = ... the specification of the panel structure

plus exactly one of

; FEM for the fixed effects model
or **; RPM ; Fcn = ... specification for the random parameters model**
or **; LCM ; Pts = J for the latent class model $**

Other parts of the specification for the categorical data model are the same as for other models of this type, e.g., tobit and truncation, that are documented elsewhere in this chapter.

Standard Model Specifications for the Truncated Regression Model

This is the full list of general specifications that are applicable to this model estimator

Controlling Output from Model Commands

; Par	keeps ancillary parameter σ with main parameter β vector in *b*.
; Margin	displays marginal effects.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = name	cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E47.4.2 Displayed and Retained Results

The display includes the log likelihood, the values or identity of the lower and upper bounds, and the estimates of [β,σ], once again, the same as the tobit model. For the truncated regression, the prediction displayed by **; List** and retained with **; Keep = name** is the conditional mean function listed above. We emphasize, the prediction is not $\beta'\mathbf{x}$. The residual that is kept with **; Res** is just the difference between actual and predicted values. Results saved by this estimator are:

Matrices: *b* and *varb* which contain the estimates for β. **; Par** adds σ to these.

Scalars:	*s*	=	σ
	ybar, *sy*		for the dependent variable
	kreg, and *nreg*		for the size of the estimation problem
	logl	=	log likelihood
	nonlimts	=	number of nonlimit observations for truncated regression.

Last Model: *b_variables, sigma*

Last Function: Conditional mean function

E47.4.3 Application

In the following results, a tobit model is reestimated as a truncated regression model. The results are based only on the nonlimit observations. A comparison of the coefficients and marginal effects to those of the tobit model is included in the results. Note that both the truncated regression and tobit models are estimating the same parameters, consistently, though results reveal substantial differences. The partial effects are quite different as well. These are finite sample effects.

```
NAMELIST    ; x = one,kl6,k618,wa,we $
TRUNCATION  ; Lhs = whrs ; Rhs = x ; Partial Effects $
TOBIT       ; Lhs = whrs ; Rhs = x ; Partial Effects $
```

```
----------------------------------------------------------------------------
Limited Dependent Variable Model - TRUNCATE
Dependent variable                  WHRS
Log likelihood function     -3408.77245
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        | Primary Index Equation for Model
Constant|    2355.46***    482.4378     4.88  .0000     1409.90   3301.02
     KL6|   -570.113***    170.0824    -3.35  .0008    -903.468  -236.757
    K618|   -179.301***    47.05707    -3.81  .0001    -271.531   -87.071
      WA|   -13.1034*       7.79318    -1.68  .0927    -28.3778    2.1710
      WE|   -30.8952       23.38742    -1.32  .1865    -76.7337   14.9433
        | Disturbance standard deviation
   Sigma|    902.808***    48.72942    18.53  .0000     807.300   998.316
--------+-------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                  WHRS
Log likelihood function     -3904.16871
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        | Primary Index Equation for Model
Constant|    1320.87***    482.9241     2.74  .0062      374.36   2267.39
     KL6|   -1077.45***    126.2053    -8.54  .0000    -1324.81   -830.09
    K618|   -128.258***    42.74783    -3.00  .0027    -212.043   -44.474
      WA|   -41.5052***     7.70256    -5.39  .0000    -56.6019  -26.4084
      WE|    95.5038***    22.86314     4.18  .0000     50.6928  140.3147
        | Disturbance standard deviation
   Sigma|    1281.18***    48.18563    26.59  .0000     1186.74   1375.62
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are All Obs.
Conditional Mean at Sample Point ********
Scale Factor for Marginal Effects   .6914
--------+-------------------------------------------------------------------
        |     Partial     Standard            Prob.      95% Confidence
    WHRS|      Effect        Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
 (Truncated regression)
     KL6|   -394.180***    115.7743    -3.40  .0007    -621.093  -167.267
    K618|   -123.970***    32.01708    -3.87  .0001    -186.723   -61.218
      WA|   -9.05979*       5.37207    -1.69  .0917   -19.58885   1.46927
      WE|   -21.3612       16.14521    -1.32  .1858    -53.0052   10.2829
--------+-------------------------------------------------------------------
 (Tobit)
     KL6|   -638.163***    73.38291    -8.70  .0000    -781.991  -494.335
    K618|   -75.9663***    25.35200    -3.00  .0027   -125.6553  -26.2773
      WA|   -24.5831***     4.54659    -5.41  .0000    -33.4943  -15.6720
      WE|    56.5660***    13.46930     4.20  .0000     30.1667   82.9653
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E47.4.4 Technical Details

The log likelihood for the truncated regression model is maximized easily using Olsen's transformation of the parameters. After transformation, the log likelihood is

$$\Sigma_i \log L_i = \Sigma_i \log\theta - ½\log 2\pi - ½\, \varepsilon_i^2 - \log[\Phi(\theta U_i - \boldsymbol{\gamma}'\mathbf{x}_i) - \Phi(\theta L_i - \boldsymbol{\gamma}'\mathbf{x}_i)].$$

where $\theta = 1/\sigma$ and $\gamma = \beta/\sigma$. Derivatives are obtained from results given earlier for the tobit model;

$$\frac{\partial \log L_i}{\partial \begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}} = \varepsilon_i \begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix} + \frac{1}{\Phi_U - \Phi_L}\left[\phi_U \begin{pmatrix}\mathbf{x}_i\\ -U_i\end{pmatrix} - \phi_L \begin{pmatrix}\mathbf{x}_i\\ -L_i\end{pmatrix}\right] + \begin{pmatrix}0\\ 1/\theta\end{pmatrix}$$

The Hessian is tedious, but obtained along similar lines. Newton's method is used for estimation, and the actual Hessian is used to estimate the asymptotic covariance matrix of the estimates.

HINT: The truncated regression model is sometimes less well behaved than the tobit model. If the data are clustered far from the assumed truncation point, the model will attempt to mimic ordinary least squares, but in so doing, the iterations can fail to converge. Thus, in an hours equation, if the range of yearly hours is 1,000 to 2,500, and you assume a truncation point of zero, in some data sets, the model (as specified) can be difficult or impossible to estimate.

HINT: Because the truncated regression is a bit volatile, Newton's method will occasionally break down. One of the symptoms is that the estimated variance goes out of range, and the function cannot be computed. If you are having trouble getting estimates, try the BFGS algorithm; add **; Alg = BFGS** to your command.

HINT: Discarding censored observations in the tobit setting does not legitimize OLS on the remaining observations. It produces the truncated regression model.

The conditional mean function in the truncated regression model is

$$E[y_i|\mathbf{x}_i, L_i \le y_i \le U_i] = \boldsymbol{\beta}'\mathbf{x}_i + \sigma\left(\frac{\phi_L - \phi_U}{\Phi_U - \Phi_L}\right),$$

As usual for nonlinear models, therefore, the coefficients are not the marginal effects. Differentiation of this function with respect to $\mathbf{x}_i$ produces the vector of slopes

$$\frac{\partial E[y_i \mid \mathbf{x}_i, L_i \le y_i \le U_i]}{\partial \mathbf{x}_i} = \beta\left[1 - \left(\frac{\alpha_L\phi_L - \alpha_U\phi_U}{\Phi_U - \Phi_L}\right) - \left(\frac{\phi_L - \phi_U}{\Phi_U - \Phi_L}\right)^2\right].$$

The term in brackets is the scale factor for the marginal effects shown in the examples above. Standard errors for the marginal effects are obtained by using the delta method. The effects are evaluated at the data means. The effects are written in terms of the parameters and the preceding scale factor as

$$\boldsymbol{\delta}(\boldsymbol{\beta},\sigma) \;=\; \boldsymbol{\beta} \times \mathrm{h}(\boldsymbol{\beta},\sigma)$$

Now, let $\boldsymbol{\Gamma} = \partial\boldsymbol{\delta}(\boldsymbol{\beta},\sigma)/\partial(\boldsymbol{\beta}',\sigma)$. The asymptotic covariance matrix is the sample estimate of

$$\mathbf{V} \;=\; \boldsymbol{\Gamma} \times \text{Asy.Var}\left[\hat{\boldsymbol{\beta}},\hat{\sigma}\right] \times \boldsymbol{\Gamma}'$$

The matrix $\boldsymbol{\Gamma}$ is extremely tedious. (For brevity, derivatives are omitted). The general form is

$$\partial\boldsymbol{\delta}(\boldsymbol{\beta},\sigma)/\partial(\boldsymbol{\beta}',\sigma) \;=\; [\mathbf{I} : \mathbf{0}] \;+\; [\partial \mathrm{h}/\partial(\boldsymbol{\beta}'\mathbf{x}) \times \boldsymbol{\beta}\mathbf{x}' \;:\; \boldsymbol{\beta}\partial \mathrm{h}/\partial\sigma]$$

Derivation makes repeated use of the rule

$$\partial[f(t)/F(t)]/\partial t \;=\; [1/F(t)]\partial f(t)/\partial t - [f(t)\,/\,F^2(t)]\,[\partial F(t)/\partial t]$$

and the template results for the standard normal distribution

$$\partial\Phi(t)/\partial \mathrm{t} \;=\; \phi(t),$$

$$\partial\phi(t)/\partial t \;=\; -t\phi(t),$$

$$\partial\alpha_{limit}/\partial\boldsymbol{\beta} \;=\; -\mathbf{x}/\sigma,$$

and

$$\partial\alpha_{limit}/\partial\sigma \;=\; -\alpha_{limit}/\sigma.$$

E48: Multiple Equation LDV Models

E48.1 Introduction

The models and estimators described in this chapter are variations on the following general simultaneous equations structure suggested in Maddala (1983) that encompasses most of the cases we wish to consider.

$$y_1^* = \gamma_1 y_2^* + \mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon_1$$

$$y_2^* = \gamma_2 y_1^* + \mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon_2$$

$$[\varepsilon_1,\varepsilon_2] \sim \text{BVN}\,[(0,0), (\sigma_{11},\sigma_{22}), \sigma_{12}], \text{ with correlation} = \rho.$$

The simultaneous equations model is stated in terms of the latent, continuous dependent variables, *before* censoring. The (assumed) existence of a reduced form ($\gamma_1\gamma_2 \neq 1$) is crucial. The literature is occasionally a bit ambiguous on this point. If, for example, it is assumed that the simultaneous equations model applies to observed variables, y_1 and y_2, while either or both are simple censored variables, for example, $y_{1i} = \text{Max}(L_i,y_{1i}^*)$, then restrictions are needed to insure that the model is even internally consistent, and, worse yet, for most formulations, that will not even be possible. (A large amount of useful discussion appears in Amemiya (1984) and Maddala (1983).) .

We consider two groups of models. The first is estimated by full information maximum likelihood. The estimators in the models in Sections E48.3 and E48.4 below use the two step maximum likelihood method. Details of this method for general applications are given in Section E48.6.

E48.2 Simultaneous Equations Model

The tobit model may be embedded in a recursive simultaneous equations model:

$$y_1 = \text{tobit as formulated above with } y_1^* = \boldsymbol{\beta}'\mathbf{x}_1 + \gamma y_2 + \varepsilon_1,$$

$$y_2 = \boldsymbol{\pi}_2'\mathbf{x}_2 + \varepsilon_2 \text{ in which Corr}[\varepsilon_1 ,\varepsilon_2] = \rho_{12}.$$

The estimator is full information maximum likelihood. (The second equation is a linear regression with observed dependent variable.) This model requires specification of a two equation model. As such, you must give both dependent variables and the Rhs for each equation. The command is

```
TOBIT    ; Lhs = y1,y2
         ; Rh1 = Rhs for tobit, including y2
         ; Rh2 = Rhs of regression model $
```

The primary object of estimation in this model is the tobit model. As such, the model output will show the regression results, but other statistics will be primarily related to the tobit equation, not the regression. Also, the retrievable results and fitted values will be for the tobit model only. An example appears below.

The results displayed include: log likelihood; identification of the model; description of the displayed parameter vector; separate estimates of σ_1^2, σ_2^2, ρ; then the parameter vector, $\boldsymbol{\beta}$ including γ the coefficient on y_2, $\boldsymbol{\pi}_2$, σ_{12}/σ_2^2, and $\sigma_{1.2} = [\sigma_1^2\,(1\text{-}\rho^2)]^{1/2}$.

The full set of model parameters is $[\boldsymbol{\beta},\boldsymbol{\pi}_2,\rho,\sigma_1^2,\sigma_2^2]$. But, for purposes of starting values and restrictions, base your specification of **; Start** and **; Rst** on the vector

$$\boldsymbol{\theta} = [\boldsymbol{\beta},\boldsymbol{\pi}_2,\rho,\sigma_{1.2}]$$

in which $\boldsymbol{\beta}$ includes the coefficient on y_2. (The variables need not be in such order that y_2 is the last variable in Rh1. Subscript '2' refers to the 2nd equation.); $\boldsymbol{\beta}$ in the first equation includes γ, the coefficient on y_2. The second disturbance variance, σ_2^2, is estimated separately as the mean squared residual. The estimated parameters are $\boldsymbol{\beta},\boldsymbol{\pi}_2$, $\psi = \sigma_{12}/\sigma_2^2$, and $\sigma_{11.2} = [\sigma_1^2(1 - \rho^2)]^{1/2}$.

Output is as usual for the tobit model. The initial OLS output will not include the second equation. These initial estimates will be inconsistent both because of the censoring and because of the endogeneity of y_2. OLS starting values are used for the second equation as well, but these are not displayed.

All other options for this model are the same as for the basic tobit model, including fitted values, iteration controls, marginal effects, and so on. The fitted values must be modified slightly for the simultaneous equations model. We condition on $\varepsilon_2 = (y_2 - \boldsymbol{\pi}_2'\mathbf{x}_2)$, so

$$E[y_i^* \,|\, y_2,\mathbf{x}_1, \varepsilon_2] = \boldsymbol{\beta}'\mathbf{x}_1 + \gamma y_2 + (\sigma_{12}/\sigma_2^2)\varepsilon_2.$$

Other computations are the same. Retrievable results, are also the same as for the tobit model of Section E45.2. For this model, the matrix *b* includes only the slopes in the tobit equation, including the coefficient on y_2. The specification **; Par** adds $\sigma_{1.2}$ to the parameter vector. Since σ_1 is saved in *s*, the estimate of ρ can be computed from these values. An additional matrix named *pi*2 is saved and contains the estimates of $\boldsymbol{\pi}_2$. The scalars and *Last Model* labels and coefficients saved are those of the tobit model.

E48.2.1 Application

To illustrate the model, we use the Mroz data and a contrived example in which

$$whrs = f_1(k618,wa,we,kl6)$$
and
$$kl6 = f_2(faminc,cit)$$

(that is, the number of small children is assumed to be endogenous). As might be expected, the data lend little support to the specification. (See the next subsection.)

```
TOBIT    ; Lhs = whrs,kl6
         ; Rh1 = one,k618,wa,we,kl6
         ; Rh2 = one,faminc,cit $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                   WHRS
Log likelihood function     -3022.86186
Estimation based on N =    753, K =  10
Inf.Cr.AIC  =   6065.7 AIC/N =    8.055
Threshold values for the model:
Lower =       .0000    Upper = +infinity
LM test [df] for tobit=     37.160[  5]
Tobit fit jointly with model for KL6
Variance estimates:
          Sigma-squared(1) =1660519.2145
          Sigma-squared(2) =       .2736
          Rho              =      -.1075
First   5 slopes are for WHRS
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|     1246.85       842.5984     1.48  .1389     -404.61   2898.32
    K618|   -128.196***     44.95896    -2.85  .0044    -216.314   -40.078
      WA|   -41.4272***      8.10705    -5.11  .0000    -57.3167  -25.5376
      WE|    96.1374***     25.49679     3.77  .0002     46.1646  146.1101
     KL6|   -813.287        2244.597     -.36  .7171   -5212.616  3586.042
        |Regression Equation
Constant|      .27648***      .04052     6.82  .0000      .19706    .35591
  FAMINC|-.36504D-06      .1215D-05      -.30  .7638 -.27458D-05  .20157D-05
     CIT|     -.04720         .03894    -1.21  .2255     -.12351    .02912
        |Variance parameters
 s12/s22|   -264.941        2235.637     -.12  .9057   -4646.709  4116.828
s[e1:e2]|    1281.14***     52.76083    24.28  .0000     1177.73   1384.55
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E48.2.2 Simultaneous Equations and Testing Exogeneity

In estimating the simultaneous equations model,

$$y_1^* \;=\; \boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma y_2 \;+ \varepsilon_1 \;\; \text{(tobit)}$$

$$y_2 \;\;=\; \boldsymbol{\pi}_2'\mathbf{x}_2 \;+\; \varepsilon_2,$$

LIMDEP estimates $\boldsymbol{\beta}_1$, γ, $\boldsymbol{\pi}_2$, $\psi=\sigma_{12}/\sigma_2^2$, and $\sigma_{11.2} \;=\; [\sigma_1^2(1 - \rho^2)]^{1/2}$. Exogeneity of y_2 can be tested by a simple t test of the hypothesis that ψ equals zero. (I.e., that $\rho[\varepsilon_1, \varepsilon_2] = 0$.) This is just the second to last coefficient reported in the model output. Note in the application above that the t ratio is quite close to zero.

E48.2.3 Models with More than Two Equations

For models with more than one regression equation, a similar maximum likelihood procedure could be constructed (see Blundell and Smith (1986)). But, the authors show that there is a much simpler way to proceed. The model is:

$$y_1^* = \boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_2 y_2 + \gamma_3 y_3 + ... + \varepsilon_1 \text{ (tobit)},$$

$$y_2 = \boldsymbol{\pi}_2'\mathbf{x}_2 + \varepsilon_2,$$

$$y_3 = \boldsymbol{\pi}_3'\mathbf{x}_3 + \varepsilon_3,$$

and so on.

The authors show that under the null hypothesis of no simultaneity, the following procedure is asymptotically equivalent to a score, or Lagrange multiplier test of weak exogeneity, i.e., $\text{Cov}[\varepsilon_1, \varepsilon_j] = 0$, $j = 2,...$:

Step 1. Use OLS to regress y_j on $\mathbf{x}_j$ for $j = 2,...$ (the regression equations) and keep the residuals (as v_j, say).

Step 2. Estimate the tobit model as specified above by maximum likelihood, but include these residual vectors as additional Rhs variables.

Step 3. The hypothesis is tested by testing the joint hypotheses that the slopes on the residuals jointly equal zero.

E48.2.4 Technical Details

This model is examined in Blundell and Smith (1986). FIML is a straightforward method of estimation. The strategy is to factor the joint distribution of ε_{i1} and ε_{i2} as $f(\varepsilon_{i1},\varepsilon_{i2}) = f(\varepsilon_{i2})f(\varepsilon_{i1}|\varepsilon_{i2})$. The log likelihood function factors likewise. The second equation is a classical regression model, so the log likelihood function can be concentrated over σ_{22}. Regardless of how $\boldsymbol{\beta}_2$ is ultimately estimated, the estimator of σ_{22} will be

$$\hat{\sigma}_{22} = (1/n)\sum_{i=1}^{n}\left(y_{i2} - \mathbf{x}_{i2}'\hat{\boldsymbol{\beta}}_2\right)^2 .$$

To construct the remaining part of the log likelihood, define

$$v_{i2} = (y_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta}_2),$$

$$v_{i1} = (y_{i1} - \mathbf{x}_{i1}'\boldsymbol{\beta}_1 - \gamma_1 y_{i2} - \psi v_{i2})$$

$$\psi = \sigma_{12}/\sigma_{22}$$

$$\omega = [\sigma_{11}(1 - \rho^2)]^{1/2}$$

Then, the log likelihood consists of the sum of the concentrated log likelihood for the regression and the part for the censored regression based on conditional distribution:

$$\log L = \text{-}(n/2)\log(1/n)\sum_{i=1}^{n}\left(y_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta}_2\right)^2 + \sum_{d_{i1}=1}\log\left[\frac{1}{\psi}\phi\left(\frac{v_{i1}}{\psi}\right)\right]$$

$$+\sum_{d_{i1}=0}\log\Phi\left[\frac{L_{i1}\text{-}\,\mathbf{x}_{i1}'\boldsymbol{\beta}_1 - \gamma_1 y_{i2} - \psi v_{i2}}{\omega}\right]$$

The function is maximized over $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_1$, ψ, and ω. The second variance parameter is estimated residually, and then, σ_{12} and σ_{11} are recovered from the estimated parameters.

E48.3 Tobit Model with Sample Selection

The sample selection model detailed in Chapter E52 is extended to the tobit model. That is,

y = tobit as formulated earlier with $\mathbf{x}$ on the Rhs,

d = a probit model based on $z^* = \boldsymbol{\alpha}'\mathbf{z} + u$,

Corr[ε,u] = ρ,

[y ,$\mathbf{x}$] observed only when d = 1.

This model is a mixture of censoring and a type of truncation. The procedure for estimating this model follows the standard set of steps for selectivity models given in Section E52.2. Complete details are given there, so we will just sketch the procedure here. The procedure for estimating a sample selectivity model in *LIMDEP* is:

Step 1. Estimate the parameters of the probit model first and **; Hold** them aside for the next step in the procedure.

Step 2. Using the probit results from Step 1, fit the sample selection model.

The estimator to be described here is a full information maximum likelihood estimator. Nonetheless, at the beginning of Step 2, a second step least squares regression is computed in order to obtain the starting values for the MLE. These are corrected for selection, to a degree, *but they are still inconsistent*. The results given at this point are obtained by least squares, and, as such, are inconsistent in the same manner as the OLS coefficients are in the basic tobit model. As noted, these are just starting values for the iterations. The MLE is consistent and efficient.

The commands are:

PROBIT ; Lhs = d ; Rhs = list for z ; Hold $
SELECT ; Tobit ; MLE ; Lhs = ... ; Rhs = ... $

Note that the command for the tobit model in this case is **SELECT**, not **TOBIT**.

NOTE: As in the MLE for the selection model, there is no 'lambda' variable computed for this model. The estimator is not least squares. When a sample selection model is fit by maximum likelihood, there is no selection 'correction' variable added to the model.

The model parameters estimated by MLE are $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, σ, ρ. These are also the estimation parameters. The probit coefficients precede the regression parameters in the parameter vector. You may provide your own starting values for the iterations with

; Start = ... list

Fixed value and equality restrictions may be imposed with

; Rst = ... list

as well. Note that constraining σ and/or ρ will likely produce unsatisfactory results. In addition, cross equation restrictions that equate elements of $\boldsymbol{\alpha}$ to elements of $\boldsymbol{\beta}$ will be problematic because of the different scaling of the two dependent variables.

The first set of output from the **SELECT** command is the standard output from the two step least squares estimation of this model. The final output includes the log likelihood and an indication of the parts of the parameter vector. The parameter vector shown is $[\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_1,\rho]$. Remaining output is the same as for the selection model. The retrievable results from this estimator are as follows:

Matrices: *b* and *varb* as usual. These contain $[\boldsymbol{\alpha},\boldsymbol{\beta},\sigma,\rho]$. Do not use **; Par**.
*bsr*1 = all of *b* except α.

Scalars: *logl, nreg, rho, varrho, s, ybar, sy, sigma*1,

Last Model: *a_variables*, *b_variables*, *r*12, *sigma*

Last Function: None

The tobit model with sample selection is developed further in Section E54.7, where derivation of the mathematical framework, an application, and further technical details are presented.

E48.4 Two Step Estimation of Censored Regression Models

The following will describe several multiple equations models. In principle, they can be generalized to an arbitrary numbers of equations. But, practical limitations, primarily the difficulty of computing multivariate normal integrals, have usually limited the applications to two equations. We will focus on this case. Some of the original sources noted suggest multivariate extensions.

E48.4.1 Recursive Simultaneous Equations Model

If $\gamma_2 = 0$, and only y_1 is censored, the resulting equation system is

$$y_{i1}^* = \gamma_1 y_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\; y_{i1} = \text{Max}(L_{i1}, y_{i1}^*)$$

$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2}.$$

This estimator is programmed directly in *LIMDEP*, so FIML estimation is a preprogrammed procedure. (See Section E48.2.) But, this model is a candidate for the two step procedure, and offers a good illustration of the technique.

Since y_{i2} is directly observed without censoring, it can be inserted into the first equation, to obtain

$$y_{i1}^* = \gamma_1(\mathbf{x}_{i2}'\boldsymbol{\beta}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + (\varepsilon_{i1} + \gamma_1\varepsilon_{i2}),\; y_{i1} = \text{Max}(L_{i1}, y_{i1}^*)$$

$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2}$$

Since the variance of ε_{i1} and the covariance of ε_{i1} and ε_{i2} are both free parameters, no generality is lost by writing the first equation as

$$y_{i1}^* = \gamma_1(\mathbf{x}_{i2}'\boldsymbol{\beta}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + u_i,\; y_{i1} = \text{Max}(L_{i1}, y_{i1}^*),\; \text{Var}[u_i] = \sigma_u^2.$$

So, in the two equation model, the second equation is a classical normal linear regression model and the first is a censored regression model with a nonlinear index function. We propose to fit this in two steps, then adjust the estimated asymptotic covariance matrix at the second step with the Murphy and Topel estimator. (Note, we will be reversing subscripts 1 and 2 in this presentation.)

Step 1. Step 1 is ordinary least squares, and $\hat{\boldsymbol{\beta}}_2$ is simply $\mathbf{b}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\,\mathbf{X}_2'\mathbf{y}_2$. The estimated asymptotic covariance matrix would normally be $\mathbf{V}_2 = [\mathbf{e}_2'\mathbf{e}_2/(n - K_2)](\mathbf{X}_2'\mathbf{X}_2)^{-1}$. The rows of the matrix $\mathbf{D}$ are the derivatives of the log likelihood for the classical regression, which would be $\mathbf{d}_i = (1/s_2^2)e_{i2}\mathbf{x}_{i2}$.

Step 2. Step 2 is maximum likelihood estimation of the censored regression model, in which the index function is $\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1$ where $z_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta}_2$. The parameter vector $[\gamma_1, \boldsymbol{\beta}_1, \sigma_u]$ is estimated exactly as we estimated the censored regression earlier. The derivatives of the log likelihood will be

$$\mathbf{g}_i = \frac{1}{\sigma_u}\left[(1-d_{i1})\lambda_{i1}^0\begin{pmatrix} z_{i2} \\ x_{i1} \\ \alpha_{i1}\end{pmatrix} + d_{i1}\begin{pmatrix} z_{i2}e_{i1}/\sigma_u \\ x_{i1}e_{i1}/\sigma_u \\ (e_{i1}/\sigma_u)^2 - 1\end{pmatrix}\right].$$

Note there is an extra element for the new term γ_1,

$$e_{i1} = y_{i1} - \gamma_1 z_{i2} - \mathbf{x}_{i1}'\boldsymbol{\beta}_1,$$

$$\alpha_{i1} = (L_{i1} - \gamma_1 z_{i2} - \mathbf{x}_{i1}'\boldsymbol{\beta}_1)/\sigma_u,$$

and

$$\lambda_{i1}^0 = -\phi(\alpha_{i1})/\Phi(\alpha_{i1}).$$

These K_1+2 element vectors are stacked in the matrix **G**. Finally, we require the matrix **M** which is embodied in **G**. The rows of **M** would be

$$\mathbf{m}_i = \frac{1}{\sigma_u}\left[(1-d_{i1})\lambda_{i1}^0\gamma_1 + d_{i1}e_{i1}/\sigma_u\right]\mathbf{x}_{i2}.$$

The matrices **D**, **G**, and **M** are used to correct the asymptotic covariance matrix computed by the censored regression estimator. Note that the variance parameters estimated are σ_{22} by s_2^2 in the second equation, computed by OLS, and $\sigma_u^2 = \sigma_{11} + \gamma_1^2\sigma_{22} + 2\gamma_1\sigma_{12}$ by squaring the estimate of σ_u from the censored regression. The covariance, σ_{12} is not estimable by this method – we are using a limited information (LIML) estimator, not a FIML one. By construction, σ_{22} is unidentified as well. (In fact, it is possible to construct an estimator of the covariance parameter. We will return to this possibility in discussion of models of sample selection.)

This program computes Blundell and Smith's two step estimator of a two equation recursive simultaneous equations model with censoring in one equation. The structure is:

$$y_1^* = \gamma_1 y_2^* + \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1 \qquad y_1^* \text{ censored at lower limit } L_i$$
$$y_2^* = \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2, \qquad y_2^* \text{ observed directly}$$

Define the variable lists, $x1$ and $x2$, dependent variables, $y1$, $y2$, and censoring limit, *li*.

```
NAMELIST    ; x1 = ... ; x2 = ... $
CREATE      ; y1 = ... ; y2 = ... $
CREATE      ; li = ... $ (May be a variable or a constant. Use a variable for both.)
```

Estimate the second equation first by OLS, and retain fitted values, s-squared, and covariance matrix.

```
REGRESS     ; Quietly ; Lhs = y2 ; Rhs = x2 ; Keep = z2 ; Res = e2 $
CALC        ; s22 = ssqrd $
MATRIX      ; v2 = varb $
```

For the second step, use the censored regression for $y1$ with $z2$. Keep all parameters including σ.

```
TOBIT       ; Quietly ; Lhs = y1 ; Rhs = z2,x1 ; Limit = li ; par $
MATRIX      ; v1 = varb $
```

Now construct **G**, **M** and **D**. In the matrix products, **G**, **M** etc. are scalars times products of variables.

```
NAMELIST   ; z2x1= z2,x1 $                                 (all variables in censored regression)
CREATE     ; d1 = y1 > li $                                (censoring indicator)
CALC       ; k1 = Col(z2X1) $                              (number of slope parameters)
MATRIX     ; cb = b(1:k1) $                                (parameters without sigma)
CREATE     ; alpha = (li - z2x1'cb)/s
           ; lambdai0 = -N01(alpha)/Phi(alpha)
           ; e1 = (y1 - z2x1'cb)/s
           ; h1 = e1*e1 - 1
           ; gcb = ((1-d1)*lambdai0+d1*e1)/s               ? partial wrt slopes
           ; gs = ((1-d1)*lambdai0*alpha+d1*h1)/s          ? partial wrt s
           ; d2b = e2/s22                                  ? use for d2i
           ; mb = ((1-d1)*lambdai0+d1*e1)*b(1)/s
           ; gcbmb = gcb*mb
           ; gsmb = gs*mb
           ; gcbd = gcb*d2b
           ; gsd = gs*d2b  $
```

The matrix assembly is done here. **G**,**M** and **G**,**D** have two parts.

```
MATRIX     ; gmb = z2x1'[gcbmb]x2
           ; gms = gsmb'x2
           ; gm = [gmb/gms]
           ; gdb = z2x1'[gcbd]x2
           ; gds = gsd'x2
           ; gd = [gdb/gds] $
```

Now compute the revised covariance matrix and display the results.

```
MATRIX     ; q = gm * v2 * gm' - gd*v2*gm' - gm*v2*gd'
           ; v1star = v1 + v1*q*v1 $
CLIST      ; twostep = z2x1,sigma $
DISPLAY    ; Parameters = b ; Covariance = v1star ; Labels = twostep $
```

The results below illustrate with the following data setup using the Mroz labor supply data:

```
CREATE     ; numkids = kl6 + k618 $
NAMELIST   ; x1 = one,wa,we ; x2 = one,faminc $
CREATE     ; y1 = whrs ; y2 = numkids $
CREATE     ; li = 0 $
```

The intermediate results are suppressed. The final computations are shown below.

```
--------------------------------------------------------------------------------
User Specified Model
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Matrix|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
      Z2|   -3850.45       5347.283     -.72  .4715   -14330.93    6630.03
Constant|    6027.32       8730.194      .69  .4899   -11083.55   23138.18
      WA|   -9.17475        6.96638    -1.32  .1878   -22.82860    4.47910
      WE|    65.0048**     29.86126     2.18  .0295      6.4778   123.5318
   SIGMA|    1353.74***    68.97469    19.63  .0000     1218.55    1488.93
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E48.4.2 Simultaneous Equations Model with Censoring

For this case, we retain the original structure,

$$y_{i1}^* = \gamma_1 y_{i2}^* + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_1$$

$$y_{i2}^* = \gamma_2 y_{i1}^* + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_2$$

$$[\varepsilon_1,\varepsilon_2] \sim BVN[(0,0),(\sigma_{11},\sigma_{22}),\rho\sigma_{11}\sigma_{22}]$$

and suppose that either or both variables are censored at lower limit L_i. For the first case, this is the model considered earlier without the restriction that γ_2 equals zero, while for the second, we assume that both variables are censored, so that least squares is inappropriate for both equations.

One Variable Censored

Once again, we depart from the reduced form of the equation system,

$$y_{i1}^* = \mathbf{x}_i'\boldsymbol{\pi}_1 + v_1$$

$$y_{i2}^* = \mathbf{x}_i'\boldsymbol{\pi}_2 + v_2$$

where $\mathbf{x}_i = \mathbf{x}_{i1} \cup \mathbf{x}_{i2}$ (all exogenous variables in the model). Suppose that variable y_{i1}^* is censored at L_i, but y_{i2}^* is observed without censoring. (This model is due to Nelson and Olsen (1978).) Insert the reduced form for y_{i2}^* into the structure for y_1^* to obtain the censored regression and linear regression model

$$y_{i1}^* = \gamma_1(\mathbf{x}_i'\boldsymbol{\pi}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + (\varepsilon_{i1} + \gamma_1 v_{i2}),\ y_{i1} = \text{Max}(L_{i1}, y_{i1}^*)$$

$$y_{i2} = \mathbf{x}_i'\boldsymbol{\pi}_2 + v_{i2}.$$

We propose the following two step estimation strategy for estimation of $(\gamma_1,\boldsymbol{\beta}_1)$:

Step 1. Estimate $\boldsymbol{\pi}_2$ by ordinary least squares regression of y_2 on all exogenous variables in the model.

Step 2. Estimate $\gamma_1,\boldsymbol{\beta}_1,\sigma_{11}$ by maximum likelihood in the censored regression model in the first equation. Then, use the Murphy and Topel correction for the asymptotic covariance matrix.

This procedure is exactly the one described in the previous section, as the partially reduced form shown above is precisely the recursive simultaneous equations model shown there. Thus, there is no need to develop the estimator in detail.

However, it remains to estimate $(\gamma_2,\boldsymbol{\beta}_2)$. We propose the following strategy: Insert the reduced form equation for y_{i1}* in the second equation to obtain the equation system

$$y_{i1}^* = \mathbf{x}_i'\boldsymbol{\pi}_1 + v_1, \; y_{i1} = \text{Max}(L_{i1}, y_{i1}^*)$$

$$y_{i2} = \gamma_2(\mathbf{x}_i'\boldsymbol{\pi}_1) + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + (\varepsilon_{i2} + \gamma_2 v_{i1}).$$

This is similar to the previous system, but is actually simpler. The two step estimation strategy is:

Step 1. Estimate $\boldsymbol{\pi}_1$ by maximum likelihood using the first equation, which is a censored regression model.

Step 2. Estimate $(\gamma_2,\boldsymbol{\beta}_2)$ in the second equation by ordinary least squares regression of y_{i2} on $z_{i1} = \mathbf{x}_i'\hat{\boldsymbol{\pi}}_1$ and $\mathbf{x}_{i2}$. Adjust the asymptotic covariance matrix at the second step using the Murphy and Topel results.

For this case, the various components of the estimator are as follows:

$$\mathbf{V}_2 = s_2^2[(\mathbf{z}_1,\mathbf{X}_2)'(\mathbf{z}_1,\mathbf{X}_2)]^{-1},$$

where s_2^2 is the usual residual variance estimator from this regression. From the Step 1 censored regression model, $\mathbf{V}_1$ is the $K\times K$ submatrix of the full estimated asymptotic covariance matrix, omitting the row and column that correspond to the MLE of σ_{v1}. The derivatives from this estimation are

$$\mathbf{d}_i = \frac{1}{\hat{\sigma}_{v1}}\left[(1-d_{i1})\lambda_{i1}^0 + d_{i1}\left(e_{i1}/\hat{\sigma}_{v1}\right)\right]\mathbf{x}_i$$

where $\quad \hat{\alpha}_{i1} = (L_{i1} - \mathbf{x}_i'\hat{\boldsymbol{\pi}}_1)/\hat{\sigma}_{v1}, \; e_{i1} = y_{i1} - \mathbf{x}_i'\hat{\boldsymbol{\pi}}_1,$

and $\quad \lambda_{i1}^0 = -\phi(\alpha_{i1})/\Phi(\alpha_{i1}).$

The remaining components are

$$\mathbf{g}_i = \frac{e_{i2}}{s_2^2}\begin{pmatrix} z_{i1} \\ \mathbf{x}_{i2} \end{pmatrix} \text{ and } \mathbf{m}_i = \frac{e_{i2}}{s_2^2}(\hat{\gamma}_2\mathbf{x}_{1i}).$$

The full set of variance parameters is not estimated by this procedure. Since we are using single equation ML techniques, the estimator uses no information about σ_{12} and, therefore, only σ_{22} is estimable. The first equation produces only an estimate of $\sigma_{v1} = (\sigma_{11}+\gamma_1^2\sigma_{22}+2\gamma_1\sigma_{12})^{1/2}$. Maddala (1983) proposes an alternative estimation technique for this model which appears to require an estimate of σ_{12}, but does not provide the necessary expression for obtaining one. Greene (1997, p. 735) observes this omission, and suggests an approach based on a two step estimator of Heckman's (1979). Greene's result, albeit correct, is unnecessary. As noted above, the simple functions of sample moments provides all the information needed to compute the asymptotic covariance matrix.

The program below completes the Blundell and Smith's two step estimator of a two equation recursive simultaneous equations model with censoring in one equation:

$$y_1^* = \boldsymbol{\pi}_1'\mathbf{x}_1 + \varepsilon_1 \qquad y_1^* \text{ censored at lower limit } L_i$$

$$y_2^* = \gamma_2(\boldsymbol{\pi}_1'\mathbf{x}_1) + \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2, \qquad y_2^* \text{ observed directly}$$

We begin with the usual data setup for the specific application.

```
NAMELIST   ; x1 = ... ; x2 = ...
           ; x = OR (x1,x2) $
CREATE     ; y1 = ... ; y2 = ... $
CREATE     ; li = ... (May be a variable or a constant. Use for both) $
```

The initial tobit estimator fits the reduced form for the first equation. The variance estimator picks up only the part for $\boldsymbol{\pi}_1$.

```
TOBIT      ; Quietly ; Lhs = y1 ; Rhs = x ; Limit = li $
MATRIX     ; pi1 = b ; v1 = varb $
```

This obtains the scale factor for the slope derivatives from the tobit equation.

```
CREATE     ; zi1 = pi1'x
           ; alphai1 = (li - zi1)/s
           ; ei = y1 - zi1
           ; lambdai0 = -N01(alphai1)/Phi(alphai1)
           ; di = (y1 <= li)*lambdai0/s + (y1 > li)*ei/s^2 $
```

This is the second step linear regression. Variable *gi* is the slope derivatives.

```
NAMELIST   ; z1x2 = zi1,x2 $
REGRESS    ; Lhs = y2 ; Rhs = z1x2 ; Res = ei2 $
CREATE     ; gi = ei2/s^2 ; mi = gi*b(1) $
```

Compute the corrected covariance matrix, then report results.

```
CREATE     ; gimi =  gi*mi
           ; gidi  =  gi*di $
MATRIX     ; a =  z1x2'[gimi] x * v1 * x'[gimi]z1x2
                - z1x2'[gidi ] x * v1 * x'[gimi]z1x2
                - z1x2'[gimi] x * v1 * x'[gidi]z1x2
           ; v2star = varb + varb * a * varb $
DISPLAY    ; Parameters = b ; Covariance = v2star ; Labels = z1x2 $
```

The results below continue the analysis of the model above with the following data setup using the Mroz labor supply data. This estimates the second equation in the system and corrects the asymptotic covariance matrix.

```
CREATE      ; numkids = kl6 + k618 $
NAMELIST    ; x1 = one,wa,we
            ; x2 = one,faminc  $
NAMELIST    ; x = OR (x1,x2) $
CREATE      ; y1 = whrs
            ; y2 = numkids  $
CREATE      ; li = 0 $
```

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=Y2       Mean                 =       1.59097
             Standard deviation   =       1.46048
             No. of observations  =           753  Degrees of freedom
Regression   Sum of Squares       =       294.090           2
Residual     Sum of Squares       =       1309.93         750
Total        Sum of Squares       =       1604.02         752
             Standard error of e  =       1.32158
Fit          R-squared            =        .18335  R-bar squared =   .18117
Model test   F[  2,   750]        =      84.19082  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =   -1276.91459  Akaike I.C.   =   .56163
             Restricted (b=0)     =   -1353.17084  Bayes  I.C.   =   .58005
             Chi squared [  2]    =     152.51250  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     ZI1|    -.00113***   .8741D-04   -12.95  .0000     -.00130   -.00096
Constant|    1.83416***      .10398    17.64  .0000     1.63037   2.03796
  FAMINC| .41300D-05      .3995D-05     1.03  .3012 -.36999D-05  .11960D-04
--------+--------------------------------------------------------------------
User Specified Model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      Y2|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     ZI1|    -.00113***      .00017    -6.77  .0000     -.00146   -.00080
Constant|    1.83416***      .17129    10.71  .0000     1.49845   2.16988
  FAMINC| .41300D-05      .4401D-05      .94  .3481 -.44964D-05  .12756D-04
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Both Variables Censored

Finally, suppose both variables are censored. The structural equation system is

$$y_{i1}^* = \gamma_1 y_{i2}^* + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1}$$

$$y_{i2}^* = \gamma_2 y_{i1}^* + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2}$$

$$[\varepsilon_{i1},\varepsilon_{i2}] \sim \text{BVN}[(0,0),(\sigma_{11},\sigma_{22}),\rho\sigma_{11}\sigma_{22}].$$

The reduced form is, as before,

$$y_{i1}^* = \mathbf{x}_i'\boldsymbol{\pi}_1 + v_{i1},\ y_{i1} = \text{Max}(L_{i1},y_{i1}^*)$$

$$y_{i2}^* = \mathbf{x}_i'\boldsymbol{\pi}_2 + v_{i2},\ y_{i2} = \text{Max}(L_{i2},y_{i2}^*).$$

For the first equation, the partially reduced form is

$$y_{i1}^* = \gamma_1(\mathbf{x}_i'\boldsymbol{\pi}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1\varepsilon_{i2}$$

and likewise for the second.

A two step estimator can be obtained by combining the elements of the previous procedures:

Step 1. Estimate ($\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$) separately by maximum likelihood estimation of the two reduced form censored regression models. Retain for each estimation, the part of the asymptotic covariance matrix corresponding to the slope parameters and the variables needed to compute the derivatives.

Step 2. Estimate $\gamma_1,\boldsymbol{\beta}_1,\sigma_{11}$ by maximum likelihood in the first equation, with variables $z_{i2} = \mathbf{x}_i'\boldsymbol{\pi}_2$ and $\mathbf{x}_{i2}$ on the right hand side. Correct the asymptotic covariance matrix.

The estimator is symmetric for estimation of $\gamma_2,\boldsymbol{\beta}_2,\sigma_{22}$

The program below shows the full set of computations for estimation of the parameters of the first equation. The procedure is symmetric in the two variables, so the counterpart for the second equation would be obtained by repeating the computations with subscripts reversed. The structure for the program is

$$y_1^* = \gamma_1(\boldsymbol{\pi}_2'\mathbf{x}) + \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1 \quad y_1^* \text{ censored at lower limit } L_{i1}$$

$$y_2^* = \gamma_2(\boldsymbol{\pi}_1'\mathbf{x}) + \boldsymbol{\beta}_2'\mathbf{x}_2 + \varepsilon_2 \quad y_2^* \text{ censored at lower limit } L_{i2}$$

As usual, we begin by setting up the specific application.

```
NAMELIST    ; x1 = ... ; x2 = ...
            ; x = OR (x1,x2) $
CREATE      ; y1 = ... ; y2 = ... $
CREATE      ; l1 = ... limits for first equation
            ; l2 = ... limits for second equation $
```

This initial reduced form tobit will estimate π_2. Pick up variance estimator from tobit model. This picks up only the π_2 part; the σ_{22} part is not needed. The derivatives from the tobit estimation are also picked up here.

```
TOBIT       ; Quietly ; Lhs = y2 ; Rhs = x ; Limit = l2 $
MATRIX      ; v1 = varb $
CREATE      ; q1 = y1 > l1  ;  q2 = y2 > l2 $
CREATE      ; z2 = b'x
            ; alpha2 = (l2 - z2)/s
            ; e2 = y2 - z2
            ; lamda2 = -N01(alpha2)/Phi(alpha2)
            ; d2i = (1-q2) * lamda2/s + q2*e2/s^2 $
```

This is the second step tobit with corrected asymptotic covariance matrix for the first equation. The correction must pick up the estimated sigma now, as in tobit, Cov(**b**,*s*) is not zero.

```
NAMELIST    ; z2x1= z2 , x1 $
TOBIT       ; Lhs = y1 ; Rhs = z2x1  ; Limit = l1 ; Par $
CALC        ; k21 = Col(z2x1) $
MATRIX      ; bz = b(1:k21) $
CREATE      ; v = bz'z2x1 ; alpha = (l1 - v)/s
            ; e = y1 - v ; lambda = -N01(alpha)/Phi(alpha)
            ; gbi = (1-q1)* lambda/s + q1*e/s^2
            ; gsi = (1-q1)* lambda*alpha/s +  q1*(1/s)*(((e/s)^2 - 1)/s - 1)
            ; mi = gbi*b(1) $
CREATE      ; gmi = gbi * mi   $
MATRIX      ; gbm= z2x1'[gmi]x
            ; gsm = gsi'[mi]x
            ; gm = [gbm / gsm]  $
CREATE      ; gdi = gbi * d2i $
MATRIX      ; gbd = z2x1'[gdi] x
            ; gsd = gsi   '[d2i] x
            ; gd = [gbd /gsd ]  $
MATRIX      ; a = gm*v1*gm' - gd*v1*gm' - gm*v1*gd' $
MATRIX      ; v1star = varb + varb * a * varb  $
CLIST       ; twostep = z2x1,sigma $
DISPLAY     ; Parameters = b ; Covariance = v1star ; Labels = twostep $
```

To illustrate use of the program, we have fit hours equations for husband and wife using the labor supply data. The data setup is

```
NAMELIST    ; x1 = one,kl6,k618,wa,we
            ; x2 = one,ha,he,faminc,cit
            ; x = OR (x1,x2) $
CREATE      ; y1 = whrs ; y2 = hhrs $
CREATE      ; l1 = 0 ?...  limits for first equation
            ; l2 = 0 $     limits for second equation
```

We have omitted the intermediate results. The output below shows the original tobit estimates for the wife's hours followed by the estimates with the corrected covariance matrix. The counterpart for the second equation would be obtained from the proceeding by reversing the subscripts in the commands or, perhaps more simply, by reversing the definitions of $y1$,$x1$ and $y2$,$x2$ in the data setup.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                   Y1
Log likelihood function     -3902.61379
Estimation based on N =    753, K =   7
Inf.Cr.AIC  =   7819.2 AIC/N =   10.384
Threshold values for the model:
Lower = L1            Upper = +infinity
ANOVA  based fit measure =      .047441
DECOMP based fit measure =      .164508
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      Y1|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
      Z2|     .76571*         .43327     1.77  .0772     -.08349   1.61491
Constant|    -289.097       1030.810     -.28  .7791   -2309.448  1731.253
     KL6|   -1057.28***     125.6007    -8.42  .0000    -1303.46   -811.11
    K618|   -159.159***     46.12655    -3.45  .0006    -249.565   -68.752
      WA|   -38.8022***      7.81149    -4.97  .0000    -54.1125  -23.4920
      WE|    79.0480***     24.52455     3.22  .0013     30.9807  127.1152
        |Disturbance standard deviation
   Sigma|    1276.57***     48.01167    26.59  .0000     1182.47   1370.67
--------+--------------------------------------------------------------------
User Specified Model
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
 LHSVar.|  Coefficient       Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
      Z2|     .76571*         .45896     1.67  .0952     -.13383   1.66525
Constant|    -289.097       1088.379     -.27  .7905   -2422.281  1844.087
     KL6|   -1057.28***     142.9767    -7.39  .0000    -1337.51   -777.06
    K618|   -159.159***     49.67412    -3.20  .0014    -256.518   -61.799
      WA|   -38.8022***      8.29942    -4.68  .0000    -55.0688  -22.5357
      WE|    79.0480***     26.02201     3.04  .0024     28.0458  130.0502
   SIGMA|    1276.57***     51.93209    24.58  .0000     1174.79   1378.36
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E48.5 Models with Binary Variables

The two step methods used for censored variables in the previous section can also be used when the endogenous variables are binary. We examine two models.

E48.5.1 Simultaneous Equations Model with Binary Variables

In principle, this is a fairly straightforward extension of the earlier results that combines a censored regression with the probit model discussed in Chapter E27. Suppose the model is formulated as

$$y_{i1}^* = \gamma_1 y_{i2}^* + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$

$$y_{i2}^* = \gamma_2 y_{i1}^* + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},\ y_{i2} = \text{Max}(L_{i2}, y_{i2}^*).$$

Thus, the first equation is a probit model and the second is a censored regression. As before, we can manipulate the reduced form to obtain the needed two step estimator. We should note before proceeding, that the sample data provide no information about the scale of y_{i1}^*, so there is no point to carrying the parameter σ_{11} through the analysis – nothing is lost by assuming $\sigma_{11} = 1$ at the outset. The reduced form of the equation system in the latent variables is, as before,

$$y_{i1}^* = \mathbf{x}_i'\boldsymbol{\pi}_1 + v_{i1},\ \ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$

$$y_{i2}^* = \mathbf{x}_i'\boldsymbol{\pi}_2 + v_{i2},\ \ y_{i2} = \text{Max}(L_{i2}, y_{i2}^*).$$

The reduced form parameters can be estimated by applying maximum likelihood to the probit model in the first equation and the censored regression in the second. We may now insert the estimated equations in the partial reduced forms

$$y_{i1}^* = \gamma_1(\mathbf{x}_i'\boldsymbol{\pi}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1 v_{i2},\ \ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$

$$y_{i2}^* = \gamma_2(\mathbf{x}_i'\boldsymbol{\pi}_1) + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2} + \gamma_2 v_{i1},\ \ y_{i2} = \text{Max}(L_i, y_{i2}^*).$$

The first equation can now be estimated as a probit model and the second as a censored regression. As a consequence of the loss of scale information, the probit estimator in the first equation estimates γ_1 and $\boldsymbol{\beta}_1$ scaled down by $\text{Var}[\varepsilon_{i1} + \gamma_1 v_{i2}]$. This is the fundamental indeterminacy in the model. The second equation can be estimated as a censored regression model. In both cases, we then adjust the estimated asymptotic covariance matrix. Both of these can be programmed using the results already given in the previous sections.

E48.5.2 Two Binary Variables

Finally, suppose both observed variables are binary. The resulting model is

$$y_{i1}^* = \gamma_1 y_{i2}^* + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2}^* = \gamma_2 y_{i1}^* + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},\ y_{i2} = \mathbf{1}(y_{i2}^* > 0).$$

Estimation can be done using the strategy suggested earlier. Once again, the observed data contain no information on scaling of the latent variables, so we assume $\sigma_1 = \sigma_2 = 1$ at the outset, with no loss of generality. Thus, $\text{Cov}[\varepsilon_1,\varepsilon_2] = \text{Corr}[\varepsilon_1,\varepsilon_2] = \rho$. This would be a conventional simultaneous equations model but for the censoring. The reduced form system is

$$y_{i1}^* = \mathbf{x}_i'\boldsymbol{\pi}_1 + v_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2}^* = \mathbf{x}_i'\boldsymbol{\pi}_2 + v_{i2},\ y_{i2} = \mathbf{1}(y_{i2}^* > 0)$$

where $\mathbf{x}_i = \mathbf{x}_{i1} \cup \mathbf{x}_{i2}$. The partial reduced form system is

$$y_{i1}^* = \gamma_1(\mathbf{x}_i'\boldsymbol{\pi}_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1 v_{i2},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2}^* = \gamma_2(\mathbf{x}_i'\boldsymbol{\pi}_1) + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2} + \gamma_2 v_{i1},\ y_{i2} = \mathbf{1}(y_{i2}^* > 0).$$

The model can be estimated using the two step method described earlier. Note what can be estimated by this method, and what cannot. Since we have normalized σ_1 and σ_2 to one in the original structure, in the reduced form, the two variances are (after skipping a bit of algebra) $\text{Var}[v_{i1}] = \theta_1^2 = (1 + \gamma_1^2 + 2\rho\gamma_2)/(1-\gamma_1\gamma_2)^2$ and $\text{Var}[v_{i2}] = \theta_2^2 = (1 + \gamma_2^2 + 2\rho\gamma_2)/(1-\gamma_1\gamma_2)^2$. The full reduced form equations can both be estimated as probit models using maximum likelihood, but as always, the coefficients are implicitly scaled. Thus, probit estimation of the reduced form produces estimates of $(1/\theta_1)\boldsymbol{\pi}_1$ and $(1/\theta_2)\boldsymbol{\pi}_2$. The partial reduced form for y_1^* is, therefore,

$$\begin{aligned} y_{i1}^* &= \theta_2\gamma_1(\mathbf{x}_i'\boldsymbol{\pi}_2/\theta_2) + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1 v_{i2} \\ &= \theta_2\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1 v_{i2} \end{aligned}$$

Taking z_{i2} as known, we could now fit this as a probit model. Once again, however, there will be a scaling because of the lack of information about the scale of the latent variable. The algebra is a bit tedious, but it follows from the fact that the partial reduced form is the true reduced form that has $\varepsilon_{i1} + \gamma_1 v_{i2}, = v_{i1}$. Therefore, probit estimation of the parameters of the first partial reduced form equation by maximum likelihood produces estimates of $(\theta_2/\theta_1)\gamma_1$ and $(1/\theta_1)\boldsymbol{\beta}_1$. Likewise, the MLEs in the second equation are of $(\theta_1/\theta_2)\gamma_2$ and $(1/\theta_2)\boldsymbol{\beta}_2$. Therefore, this two step estimator produces estimates not of the original parameters, but of these scaled versions of them. Can the original parameters be recovered? No, because this method produces no estimate of the correlation coefficient between ε_{i1} and ε_{i2} or v_{i1} and v_{i2}. The product of γ_1 and γ_2 can be recovered, in an obvious way, but that is as far as one can go.

In fact, there is a way to estimate all of the parameters of the model. We return to the original structural equations. The reduced form for the system can be written

$$\begin{aligned}
y_{i1}^* &= \gamma_1\gamma_2 y_{i1}^* + \gamma_1\mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} + \gamma_1\varepsilon_{i2} \\
&= [1/(1-\gamma_1\gamma_2)]\mathbf{x}_{i1}'\boldsymbol{\beta}_1 + [\gamma_1/(1-\gamma_1\gamma_2)]\mathbf{x}_{i2}'\boldsymbol{\beta}_2 + [1/(1-\gamma_1\gamma_2)](\varepsilon_{i1} + \gamma_1\varepsilon_{i2}) \\
&= \mathbf{x}_{i1}'\boldsymbol{\delta}_1 + \gamma_1(\mathbf{x}_{i2}'\boldsymbol{\delta}_2) + v_{i1} \\
y_{i2}^* &= \gamma_2(\mathbf{x}_{i1}'\boldsymbol{\delta}_1) + \mathbf{x}_{i2}'\boldsymbol{\delta}_2 + v_{i2}.
\end{aligned}$$

This is a bivariate probit model, which can be fit by maximum likelihood. The estimation is fairly complicated, because all of the cross equation restrictions must be imposed, and the index parts of the equations are nonlinear. But, it is a conventional programming problem. The variances of v_{i1} and v_{i2} were given earlier. The maximum likelihood procedure will produce estimates of γ_1/θ_1, γ_2/θ_2, and Corr(v_{i1},v_{i2}), which are extremely complicated, but invertible functions of γ_1, γ_2, and ρ. Finally, with these in hand, $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ could be unscaled to recover $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Obviously, this is far more complicated than the two step approach. It does demonstrate the different quantities estimated by the two approaches. Finally, we note, that the primary difference between the FIML approach and the two step (LIML) approach is that the former estimates ρ while the latter does not.

The two step estimator is by far the simpler of the two procedures. For estimation of the first equation,

Step 1. Fit the reduced form for the second equation and compute $z_{i2} = \mathbf{x}_i'\boldsymbol{\pi}_2$ using the MLE. Let $\mathbf{V}_1$ denote the asymptotic covariance matrix computed at this step. At this step, also compute

$$\mathbf{d}_{i2} = (2y_{i2}-1)\phi(\mathbf{x}_i'\boldsymbol{\pi}_2)/\Phi[(2y_{i2}-1)(\mathbf{x}_i'\boldsymbol{\pi}_2)] \times \mathbf{x}_i = \lambda_{i1}\mathbf{x}_i$$

Step 2. Fit the structural probit equation on z_{i2} and $\mathbf{x}_{i1}$. Denote the asymptotic covariance matrix computed at this step as $\mathbf{V}_2$. The vectors needed at this step for the corrected asymptotic covariance matrix are

$$\mathbf{g}_{i1} = \{(2y_{i1}-1)\phi(\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1)/\Phi[(2y_{i1}-1)(\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1)]\}\begin{bmatrix} z_{i2} \\ \mathbf{x}_{i1}\end{bmatrix} = \lambda_{i2}\mathbf{x}_{i1}^*$$

$$\mathbf{m}_{i1} = \{(2y_{i1}-1)\phi(\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1)/\Phi[(2y_{i1}-1)(\gamma_1 z_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1)]\}\gamma_1\,\mathbf{x}_i = \lambda_{i2}\mathbf{x}_i.$$

(Note, $\mathbf{x}_{i1}^* = (z_{i2},\mathbf{x}_{i1}')'$) With these in hand, the corrected asymptotic covariance matrix can be computed as usual. In this case, the estimator has a particularly simple form. Let $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_{i1})$ and $\boldsymbol{\Lambda}_2 = \text{diag}(\lambda_{i2})$, and suppose we use the BHHH estimator for the asymptotic covariance matrix for both probit estimators. Then, $\mathbf{V}_1 = [\mathbf{X}'\boldsymbol{\Lambda}_1^2\mathbf{X}]^{-1}$, $\mathbf{V}_2 = [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*]^{-1}$, $\mathbf{G}'\mathbf{M} = \gamma_1[\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}]$, and $\mathbf{G}'\mathbf{D} = [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_1\mathbf{X}]$. Multiplying out the parts produces

$$\begin{aligned}
\mathbf{V}_2^* = [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*]^{-1} + \gamma_1[\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*]^{-1}\{&\gamma_1 [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}][\mathbf{X}'\boldsymbol{\Lambda}_1^2\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*] \\
&- [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_1\mathbf{X}] [\mathbf{X}'\boldsymbol{\Lambda}_1^2\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*] \\
&- [\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}][\mathbf{X}'\boldsymbol{\Lambda}_1^2\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_1\mathbf{X}_2^*]\}[\mathbf{X}_2^{*\prime}\boldsymbol{\Lambda}_2^2\mathbf{X}_2^*]^{-1}
\end{aligned}$$

The roles of equations 1 and 2 are reversed to obtain the counterparts for the second equation.

The following program can be used for estimation of an equation system with two binary variables as the observed data. Data setup is the usual. Note in this application, *y*1 and *y*2 are binary, and there is no need to define the censoring limits. The right hand sides of the two equations and the union are defined first.

```
NAMELIST    ; x1 = ... ; x2 = ... $
CREATE      ; y1 = ... ; y2 = ... $
NAMELIST    ; x = OR (x1,x2) $
```

We now do the estimation for the first equation. To repeat for the second equation, it is simplest just to reverse the subscripts in the data setup above. Fit the reduced form for the second equation first.

```
PROBIT      ; Lhs = y2 ; Rhs = x ; Hold(IMR = di) $
CREATE      ; z2 = b'x $
MATRIX      ; v1 = varb $
```

Estimate the structure for the first equation, for *y*1.

```
NAMELIST    ; z2x1= z2,x1 $
PROBIT      ; Lhs = y1 ; Rhs = z2x1 ; Hold(IMR = gi) $
```

This is all that is needed to compute the corrected covariance matrix.

```
CREATE      ; migi = gi*gi*b(1) ; digi = di*gi $
MATRIX      ; a = z2x1'[migi]x * v1 * x'[migi]z2x1
                - z2x1'[digi] x * v1 * x'[migi]z2x1
                - z2x1'[migi]x * v1 * x'[digi]z2x1
            ; v2star = varb + varb * a * varb $
DISPLAY     ; Parameters = b ; Covariance = v2star ; Labels = z2x1 $
```

Second equation. Same procedure.

```
PROBIT      ; Lhs = y1 ; Rhs = x ; Hold(IMR = di) $
CREATE      ; z1 = b'x$
MATRIX      ; v1 = varb $
NAMELIST    ; z1x2 = z1,x2 $
PROBIT      ; Lhs = y2 ; Rhs = z1x2 ; Hold(IMR = gi) $
CREATE      ; migi = gi*gi*b(1) ; digi = di*gi $
MATRIX      ; a = z1x2'[migi]x * v1 * x'[migi]z1x2
                - z1x2'[digi] x * v1 * x'[migi]z1x2
                - z1x2'[migi]x * v1 * x'[digi]z1x2
            ; v2star = varb + varb * a * varb $
DISPLAY     ; Parameters = b ; Covariance = v2star ; Labels = z1x2 $
```

To illustrate the procedure, we have fit a model for the wife's labor force participation, and for the husband, whether they are working 'full time,' that is, at least 2,000 hours for the year.

```
NAMELIST    ; x1 = one,kl6,k618,wa,we
            ; x2 = one,ha,he,faminc,cit
            ; x = OR(x1,x2) $
CREATE      ; y1 = whrs > 0
            ; y2 = hhrs >= 2000  $
```

These are the two step structural estimates for the wife's labor force participation.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y1
Log likelihood function      -464.51061
Restricted log likelihood    -514.87320
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      Y1|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
      Z2|      .19357         .17690     1.09  .2738     -.15314     .54029
Constant|      .55374         .47081     1.18  .2395     -.36904    1.47651
     KL6|     -.87276***      .11258    -7.75  .0000    -1.09342    -.65210
    K618|     -.07098*        .04254    -1.67  .0952     -.15435     .01239
      WA|     -.03565***      .00782    -4.56  .0000     -.05098    -.02032
      WE|      .11058***      .02379     4.65  .0000      .06396     .15720
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
User Specified Model
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      Y1|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      Z2|      .19357         .18279     1.06  .2896     -.16469     .55184
Constant|      .55374         .47897     1.16  .2476     -.38503    1.49250
     KL6|     -.87276***      .11639    -7.50  .0000    -1.10088    -.64464
    K618|     -.07098         .04395    -1.62  .1063     -.15712     .01516
      WA|     -.03565***      .00791    -4.51  .0000     -.05114    -.02015
      WE|      .11058***      .02477     4.46  .0000      .06204     .15913
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E48.5.3 Endogenous Binary Variables

A problem with the approach of the previous section is that unlike censored regressions, in which arguably the latent regressands, y_j^*, might be of interest, researchers would not ordinarily analyze a model involving a binary variable in terms of the latent regression. The dummy variable is generally viewed as a shift parameter in the equation, and, as such, y_{i1}, not y_{i1}^*, is what would typically appear in the second equation. The reduced form analysis we have done above greatly simplifies the derivations, but they may substitute a simpler estimation process for the one really of interest.

Results have been obtained for cases of 'endogenous' dummy variables – the extensive literature of Maddala (1983), Heckman (1979), Terza (1998), and others will apply. Many of these results are based on placing endogenous binary variables in linear regression models – that is, models without censoring. In this case, the conditional mean function, rather than the likelihood function turns out to be the platform on which consistent estimation can be performed. Consider, first, a regression model with a binary variable, but no censoring:

$$y_{i1}^* = \gamma_1 y_{i2} + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\; y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2} = \gamma_2 y_{i1} + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},$$

This model cannot be internally consistent unless γ_1 equals zero. (A sketch of a proof appears in Maddala (1983).) With $\gamma_1 = 0$, a fairly common specification emerges – this is often employed as a 'treatment effects' model. In this formulation, the binary variable y_{i1} indicates presence or absence of some treatment (such as participation in a program or experiment), and y_{i2} measures the outcome variable of interest, such as income, grade improvement, health improvement, and so on. There are at least three approaches to estimation, FIML, instrumental variables, and the two step estimator pioneered by Heckman (1979). We will analyze this model in some detail in the chapter on the sample selection model, so we consider it only briefly here.

An important part of the development will be that y_{i2} is fully observed. One approach to estimation can be based on constructing the conditional mean functions. From the first equation, $E[y_{i1}|\mathbf{x}_{i1}] = \Phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)$ – recall, $\gamma_1 = 0$ - so $y_{i1} = \Phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1) + u_{i1}$ where $E[u_{i1}|\mathbf{x}_{i1}] = 0$ and, as usual for conditional mean functions, $\text{Cov}[u_{i1},\Phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)] = 0$. Inserting these into the second equation, we obtain

$$\begin{aligned} y_{i2} &= \gamma_2\Phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1) + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + (\varepsilon_{i2} + \gamma_2 u_{i1}) \\ &= \gamma_2 z_{i1} + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + v_{i2}. \end{aligned}$$

At least in terms of the population values, this is a classical regression model. It is interesting to note that the conventional rules for identification in simultaneous equations models do not apply here. Even though this is a recursive model, consistent estimation will not require that ε_1 and ε_2 be uncorrelated. Moreover, because of the nonlinearity of the conditional mean function, it is not necessary for there to be variables excluded from either equation – the standard rank and order conditions do not apply to nonlinear systems.

This is a natural candidate for the two step estimator. The first step would be consistent estimation of $\boldsymbol{\beta}_1$ by treating the first equation as a probit model. Estimates of the variable z_{i1} are computed using this maximum likelihood estimator of $\boldsymbol{\beta}_1$. At the second step, γ_2 and $\boldsymbol{\beta}_2$ are consistently estimated by least squares regression of y_{i2} on $[z_{i1},\mathbf{x}_{i2}]$. The asymptotic covariance matrix at the second step is adjusted by the Murphy and Topel estimator. Let $\mathbf{V}_1$ denote the asymptotic covariance matrix computed for the probit estimator at the first step, and let $\mathbf{V}_2$ denote the estimated asymptotic covariance matrix computed at step 2. Then, as in the other cases,

$$\mathbf{V}_2^* = \mathbf{V}_2 + \mathbf{V}_2[\mathbf{G}'\mathbf{M}\,\mathbf{V}_1\mathbf{M}'\mathbf{G} - \mathbf{G}'\mathbf{D}\mathbf{V}_1\mathbf{M}'\mathbf{G} - \mathbf{G}'\mathbf{M}\,\mathbf{V}_1\,\mathbf{D}'\mathbf{G}]\mathbf{V}_2$$

where

$$\mathbf{g}_i = e_i[z_{i1,}\mathbf{x}_{i2}]/\sigma^2,$$

$$\mathbf{m}_i = e_i\{\gamma_2\phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)]\mathbf{x}_{i2}\}/\sigma^2, \text{ and}$$

$$\mathbf{d}_i = (2y_{i1}\text{-}1)\phi(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)/\Phi[(2y_{i1}\text{-}1)(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)]\mathbf{x}_{i1}.$$

Our interest at this juncture is in a model which includes censoring of the dependent variable in the second equation, so we consider that case now. A fully operational estimator for the simultaneous equations model

$$y_{i1}^* = \gamma_1 y_{i2}^* + \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2}^* = \gamma_2 y_{i1} + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},\ y_{i2} = \text{Max}(L_i, y_{i2}^*) \text{ (note, } y_{i1}, \text{ not } y_{i1}^*)$$

remains to be derived. Estimation is not the only difficulty. It is unclear whether the model is even internally consistent. (Maddala shows that the model cannot be internally consistent if γ_1 is nonzero.) This restricts us to recursive models. Consider, then, the model with $\gamma_1 = 0$

$$y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0)$$
$$y_{i2}^* = \gamma_2 y_{i1} + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},\ y_{i2} = \text{Max}(L_i, y_{i2}^*).$$

The two step estimator is problematic in this case. The partial reduced form for y_{i2}^* is not available – the precise prediction that should be inserted for y_{i1} is unclear. But, a full information maximum likelihood estimator is quite feasible, so we will develop that.

We continue to assume that ε_1 and ε_2 are bivariate normally distributed with zero means, variances one and σ_1^2, and correlation ρ. Consider, first, the cases in which y_{i2}^* is censored. The probabilities associated with these outcomes are the probabilities of the joint events,

$$\text{Prob}[y_{i2} = L_{i2}\,,\, y_{i1} = 0] \text{ and } \text{Prob}[y_{i2} = L_{i2}\,,\, y_{i1} = 1].$$

These are simply the bivariate standard normal integrals,

$$\text{Prob}[y_{i2} = L_{i2}, y_{i1} = 0] = \text{Prob}[\varepsilon_{i1} \le \text{-}\mathbf{x}_{i1}'\boldsymbol{\beta}_1, (\varepsilon_{i2}/\sigma_2) \le (L_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta}_2)/\sigma_2 \mid \rho]$$

and

$$\text{Prob}[y_{i2} = L_{i2}, y_{i1} = 1] = \text{Prob}[\varepsilon_{i1} > \text{-}\gamma - \mathbf{x}_{i1}'\boldsymbol{\beta}_1, (\varepsilon_{i2}/\sigma_2) \le (L_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta}_2)/\sigma_2 \mid \rho]$$
$$= \text{Prob}[\varepsilon_{i1} \le \gamma + \mathbf{x}_{i1}'\boldsymbol{\beta}_1, (\varepsilon_{i2}/\sigma_2) \le (L_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta}_2)/\sigma_2 \mid \text{-}\rho],$$

so these are the terms in the likelihood function for the fully censored data. It might seem odd that this has ignored the simultaneity. However, the result can be obtained trivially by writing Prob[$y_{i2} = L_{i2}$, $y_{i1} = 1$] = Prob[$y_{i2} = L_{i2}| \, y_{i1} = 1$]Prob[$y_{i1} = 1$]. The former probability is just the joint probability divided by the marginal, which then cancels out of the product, and, of course, conditioned on y_{i1}, we are free to treat y_{i1} as a constant. The 'simultaneity' only becomes an issue in regression because of the use of covariances and moments. In this instance, we are using the probabilities directly.

For the uncensored observations, we require the mixed distributions, $f(y_{i1}=0,y_{i2})$ and $f(y_{i1}=1,y_{i2})$. The first of these, $f(y_{i1}=0,y_{i2})$, is derived from

$$
\begin{aligned}
f(\varepsilon_{i1} \le -\mathbf{x}_{i1}'\boldsymbol{\beta}_1, \varepsilon_{i2}) &= \int_{-\infty}^{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1} f(\varepsilon_{i1},\varepsilon_{i2})d\varepsilon_{i1} \\
&= \int_{-\infty}^{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1} f(\varepsilon_{i2})f(\varepsilon_{i1}\,|\,\varepsilon_{i2})d\varepsilon_{i1} \\
&= f(\varepsilon_{i2})\int_{-\infty}^{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1} f(\varepsilon_{i1}\,|\,\varepsilon_{i2})d\varepsilon_{i1} \\
&= \frac{1}{\sigma_2}\phi\left(\frac{y_{i2}-\mathbf{x}_{i2}'\boldsymbol{\beta}_2}{\sigma_2}\right)\int_{-\infty}^{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1}\frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{\varepsilon_{i1}-(\rho/\sigma_2)\varepsilon_{i2}}{\sqrt{1-\rho^2}}\right)d\varepsilon_{i1} \\
&= \frac{1}{\sigma_2}\phi\left(\frac{y_{i2}-\mathbf{x}_{i2}'\boldsymbol{\beta}_2}{\sigma_2}\right)\int_{-\infty}^{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1}\frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{\varepsilon_{i1}-(\rho/\sigma_2)\left(y_{i2}-\mathbf{x}_{i2}'\boldsymbol{\beta}_2\right)}{\sqrt{1-\rho^2}}\right)d\varepsilon_{i1} \\
&= \frac{1}{\sigma_2}\phi\left(\frac{y_{i2}-\mathbf{x}_{i2}'\boldsymbol{\beta}_2}{\sigma_2}\right)\Phi\left(\frac{-\mathbf{x}_{i1}'\boldsymbol{\beta}_1+(\rho/\sigma_2)(y_{i2}-\mathbf{x}_{i2}'\boldsymbol{\beta}_2)}{\sqrt{1-\rho^2}}\right).
\end{aligned}
$$

For $f(y_{i1}=1,y_{i2})$, the numerator inside the CDF is changed to $\mathbf{x}_{i1}'\boldsymbol{\beta}_1+(\rho/\sigma_2)(y_{i2} - \gamma - \mathbf{x}_{i2}'\boldsymbol{\beta}_2)$, and other parts remain the same. These four results, then, give the parts of the likelihood function, which can then be maximized to estimate the parameters.

The following estimation program is for a censored regression with an endogenous binary variable. The equation system is

$$
\begin{aligned}
y_{i1}^* &= \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},\ y_{i1} = \mathbf{1}(y_{i1}^* > 0) \\
y_{i2}^* &= \gamma_2 y_{i1} + \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2},\ y_{i2} = \text{Max}(L_i, y_{i2}^*).
\end{aligned}
$$

The left and right hand sides of the two equations are defined for the specific problem. The censoring limit for the second equation will typically be zero, but can be nonzero. That is defined here as well. The rest of the command set is generic, and can be used without modification.

```
NAMELIST   ; x1  = the Rhs of the probit model $
NAMELIST   ; x2  = exogenous variables in the censored regression $
CREATE     ; y1  = binary dependent variable $
CREATE     ; y2  = censored dependent variable $
CREATE     ; li  = censoring limit $
```

Obtain the dimensions of the problem, and pointers to partition the parameter vector.

```
CALC        ; k2 = Col(x2) ; k21=k2+1 ; k1=Col(x1) $
```

Get the starting values for the probit model. These are consistent, but LIML, so they are inefficient.

```
PROBIT      ; Lhs = y1 ; Rhs = x1 $
MATRIX      ; beta10 = b $
```

Obtain the starting values for censored regression. These are inconsistent, but better than zero.

```
TOBIT       ; Lhs = y2 ; Rhs = y1,x2 $
CALC        ; gamma0 = b(1) $
MATRIX      ; beta20 = b(2:k21) $
```

Compute a starting value for σ in the tobit equation, then use the Olsen transformation.

```
CALC        ; s20 = s ; h20 = 1/s $
CREATE      ; d = y2 > li ; q1 = 2*y1 - 1 $
```

Finally, compute the FIML estimator of all model parameters using maximum likelihood.

```
MAXIMIZE    ; Quietly ; Labels = k1_b1,c,k2_b2,h2,r
            ; Start = beta10,gamma0,beta20,h20,0
            ; Fcn = x1b1 = b11'x1                 |
                    x2b2 = b21'x2                 |
                    a2   = (li - x2b2)* h2        |
                    e2   = (y2 - c*y1 - x2b2) * h2 |
                    dr   = 1/sqr(1 - r*r)         |
                    u1   = q1*(c*y1 + x1b1)       |
                    u2   = -q1*r                  |
              Log((1-d) * BVN(u1, a2, u2) + d * h2*N01(e2) * Phi(dr*(q1*x1b1 + r*e2))) $
CLIST       ; fiml = x1,gamma0,x2,sigma2,corr $
DISPLAY     ; Parameters = b ; Covariance = varb ; Labels = FIML $
```

We used the procedure to fit a model for joint determination of the wife's labor force participation and husband's hours for full time (hours greater than 2000).

```
NAMELIST    ; x1 = one,wa,we,kl6,k618,cit $
NAMELIST    ; x2 = one,ww,ha,he $
CREATE      ; y1 = lfp $
CREATE      ; y2 = hhrs $
CREATE      ; li = 2000 $
```

```
--------------------------------------------------------------------------------
User Specified Model
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
Constant|      .39909        .30801     1.30  .1951     -.20460   1.00278
      WA|    -.02963***      .00529    -5.60  .0000     -.04000   -.01927
      WE|     .09423***      .01942     4.85  .0000      .05616    .13230
     KL6|    -.51115***      .07963    -6.42  .0000     -.66722   -.35507
    K618|    -.01623         .02689     -.60  .5461     -.06893    .03647
     CIT|    -.22019***      .07158    -3.08  .0021     -.36049   -.07988
  GAMMA0|   -56.1770       69.06464     -.81  .4160   -191.5412   79.1872
Constant|    2341.94***    96.29192    24.32  .0000     2153.21   2530.67
      WW|   -8.28686       10.03896     -.83  .4091   -27.96285  11.38914
      HA|   -7.00215***      .28790   -24.32  .0000    -7.56643  -6.43787
      HE|    19.3696***     6.21548     3.12  .0018      7.1875   31.5517
  SIGMA2|      .00160***  .4756D-04    33.74  .0000      .00151    .00170
    CORR|      .88659***     .03027    29.29  .0000      .82727    .94591
--------+-----------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

E48.6 Murphy and Topel's Two Step Estimator

We consider limited information maximum likelihood (LIML) estimation of a model which can be formulated in terms of two marginal distributions, $f_1(y_1|\mathbf{x}_1,\boldsymbol{\theta}_1)$ and $f_2(y_2|\mathbf{x}_1,\mathbf{x}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$. We propose to estimate this model in two steps: First, estimate $\boldsymbol{\theta}_1$ by maximum likelihood estimation based on $f_1(y_1|\mathbf{x}_1,\boldsymbol{\theta}_1)$. Second, estimate $\boldsymbol{\theta}_2$ by maximum likelihood based on $f_2(y_2|\mathbf{x}_1,\mathbf{x}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, after inserting the estimate of $\boldsymbol{\theta}_1$ obtained at Step 1, and treating it as known. The consistency of the MLE of $\boldsymbol{\theta}_1$ implies that this strategy will produce a consistent estimator at the second step. However, the conventional asymptotic covariance matrix computed at Step 2 will be inappropriate because of the variation introduced by the estimated value of $\boldsymbol{\theta}_1$. The Murphy and Topel (2002) result provides a strategy for computing an appropriate covariance matrix at the second step. Let

$$\log L_1 \;=\; \sum\nolimits_{i=1}^{n} \log f_1(y_{i1}\,|\,x_{i1},\boldsymbol{\theta}_1)$$

The first step MLE of $\boldsymbol{\theta}_1$ is obtained by maximizing log L_1. Let $\mathbf{V}_1$ denote an appropriate estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_1$, however computed – this might be based on the actual Hessian (Newton's method), the expected Hessian (scoring), or the BHHH estimator. Let

$$\log L_2^c \;=\; \sum\nolimits_{i=1}^{n} \log f_2\left[\, y_{i2}|\mathbf{x}_{i2},\boldsymbol{\theta}_2,\left(\mathbf{x}_{i1},\hat{\boldsymbol{\theta}}_1\right)\right]$$

denote the conditional log likelihood for y_2 with the first step MLE of $\boldsymbol{\theta}_1$ inserted as if it were the known $\boldsymbol{\theta}_1$. We obtain the two step MLE of $\boldsymbol{\theta}_2$ by maximizing log L_2^c with respect to $\boldsymbol{\theta}_2$. Let $\mathbf{V}_2$ denote the estimated asymptotic covariance matrix for this estimator, however computed,

assuming(incorrectly) that $\hat{\boldsymbol{\theta}}_1$ is the true value of $\boldsymbol{\theta}_1$. The matrix $\mathbf{V}_2$ underestimates the asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}_2$. Murphy and Topel show that an appropriate estimator is found as follows:
Let $\mathbf{g}_i$ be the vector of partial derivatives of the ith term in log L_2^c

$$\mathbf{g}_i = \partial \log f_2(\mathrm{y}_2|\mathbf{x}_1,\mathbf{x}_2, \hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2$$

The matrix $\mathbf{G}$ constructed by stacking the rows $\mathbf{g}_i$ contains the derivatives of the log likelihood for $\boldsymbol{\theta}_2$, so $\partial \log L_2/\partial\boldsymbol{\theta}_2 = \mathbf{G}'\mathbf{i}$, which is $\mathbf{0}$ at the MLE. The BHHH estimator would be $[\mathbf{G}'\mathbf{G}]^{-1}$ when $\mathbf{G}$ is computed using $\hat{\boldsymbol{\theta}}_2$. Let

$$\mathbf{m}_i = \partial \log f_2(y_2|\mathbf{x}_1,\mathbf{x}_2, \hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_1.$$

The matrix $\mathbf{M}$ contains the derivatives of log L_2^c with respect to $\boldsymbol{\theta}_1$ (the first step parameter vector), so $\partial \log L_2^c/\partial\theta_1 = \mathbf{M}'\mathbf{i}$ – this is not necessarily $\mathbf{0}$. Finally, return to the first step maximum likelihood estimation procedure, and define

$$\mathbf{d}_i = \partial \log f_1(y_1|\mathbf{x}_1,\boldsymbol{\theta}_1)/\partial\boldsymbol{\theta}_1.$$

The matrix $\mathbf{D}$ contains the derivatives of log L_1 with respect to $\boldsymbol{\theta}_1$, so $\partial \log L_1/\partial\boldsymbol{\theta}_1 = \mathbf{D}'\mathbf{i}$. This vector does equal $\mathbf{0}$ when evaluated at $\hat{\boldsymbol{\theta}}_1$. With these in place, the Murphy and Topel estimator of the appropriate estimator for the two step maximum likelihood estimator, $\hat{\boldsymbol{\theta}}_2 \,|\, \hat{\boldsymbol{\theta}}_1$ is

$$\mathbf{V}_2^* = \mathbf{V}_2 + \mathbf{V}_2[(\mathbf{G}'\mathbf{M})\mathbf{V}_1(\mathbf{M}'\mathbf{G}) - (\mathbf{G}'\mathbf{D})\mathbf{V}_1(\mathbf{M}'\mathbf{G}) - (\mathbf{G}'\mathbf{M})\mathbf{V}_1(\mathbf{D}'\mathbf{G})]\mathbf{V}_2$$

where $\mathbf{G}$, $\mathbf{M}$, and $\mathbf{D}$ are computed using the two sets of maximum likelihood estimates.

For most of the familiar econometric models, including the ones we will consider here, the variables, $\mathbf{x}_1$ and $\mathbf{x}_2$ enter the log likelihoods through linear index functions, $\mathbf{x}_1'\boldsymbol{\theta}_1$ and $\mathbf{x}_2'\boldsymbol{\theta}_2$. This means that frequently, we will find $\mathbf{g}_i = w_{i22}(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \times \mathbf{x}_2$ for some scalar function $w_i(.)$, and likewise for a w_{i21} for $\mathbf{m}_i$ and w_{i11} for $\mathbf{d}_i$. This would make, for example,

$$\mathbf{G}'\mathbf{M} = \sum_{i=1}^{n} w_{i22}(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \times w_{i12}(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \times \mathbf{x}_2 \times \mathbf{x}_1'.$$

If we denote the product of scalars as simply w_i, arrange these scalars in an $n\times n$ diagonal matrix, $\mathbf{W}$, and define data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ in the obvious way, then this computation will simplify to

$$\mathbf{G}'\mathbf{M} = \sum_{i=1}^{n} w_i(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \times \mathbf{x}_2 \times \mathbf{x}_1' = \mathbf{X}_2'\mathbf{W}\mathbf{X}_1.$$

This is a pattern that occurs often enough that the Murphy and Topel results are usually far simpler than first appearances would suggest. We made use of it in several of the applications described earlier.

E49: Generalized Linear Models – 1: Discrete

E49.1 Introduction

This chapter and Chapters E50 and E51 present a group of 'generalized linear models' (GLMs) that can be used for dependent variables whose range is generally restricted, either because they are discrete (such as a binary variable) or because they naturally vary over only a restricted range (such as variables that are only nonnegative). The class of generalized linear models was defined in the pioneering works of Nelder and Wedderburn (1972) and McCullagh and Nelder (1983). As shown below, many of these are models that *LIMDEP* fits under a different heading, but it is convenient to group them here. Formally, the class of models is a group in which the conditional mean function is of the form $E[y|\mathbf{x}] = h(\boldsymbol{\beta}'\mathbf{x})$ for some continuous function $h(.)$. (McCullagh and Nelder and others since have focused on 'exponential families,' but we take some license here, and broaden their class.) This class includes most of the single index function models already considered, such as the binary choice models, censored regression, truncated regression, and all of the count models considered in Chapters E41-E44. These chapters will present a group of models not already considered and also organize several from earlier chapters for the convenience of the user interested in this class of models.

The basic command for estimation of the models described in this chapter is

```
GLIM          ; Lhs = dependent variable
              ; Rhs = independent variables
              ; Model = type of model $
```

where '**type of model**' is one of the 25 generalized linear models presented here.

The most convenient way to organize these models is by type of dependent variable. This chapter will describe several models for discrete dependent variables, such as the probit and logit model. Chapter E50 will describe models for continuous variables, such as some for variables constrained to lie in the interval (0,1). Chapter E51 will extend both types of models to several panel data settings.

In each framework, the estimation procedure is maximum likelihood, based on the formal specification of the distribution of the observed random variable. We begin the development with some methodological points about GLIM models.

E49.2 Estimating Generalized Linear Models

The following are the central features of Nelder and Wedderburn's (1972) and McCullagh and Nelder's (1983) GLM approach to specification. (We present this as an application to panel data to simplify the presentation in Chapter E51.) The generalized linear model is specified by a 'link' to the conditional mean function,

$$f(E[y_{it} \mid \mathbf{x}_{it}]) = \boldsymbol{\beta}'\mathbf{x}_{it},$$

and a 'family' of distributions,

$$y_{it} \mid \mathbf{x}_{it} \sim g(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it}, \boldsymbol{\theta})$$

where $\boldsymbol{\beta}$ and $\mathbf{x}_{it}$ are as already defined and $\boldsymbol{\theta}$ is zero or more ancillary parameters, such as the dispersion parameter in the negative binomial model (which is a GLM). Many of the models already discussed fit into this framework, such as the standard probit model which has link function $f(.) = \Phi^{-1}(P)$ and Bernoulli distribution family and the classical normal linear regression which has link function equal to the identity function and normal distribution family. More generally, for the single index binary choice models estimated by *LIMDEP*, if $\text{Prob}(y_{it} = 1) = F(\boldsymbol{\beta}'\mathbf{x}_{it})$, then this is the conditional mean function, and the link function is simply (by definition)

$$f(E[y_{it} \mid \mathbf{x}_{it}]) = F^{-1}[F(\boldsymbol{\beta}'\mathbf{x}_{it})] = \boldsymbol{\beta}'\mathbf{x}_{it}.$$

This includes the probit, logit, Gompertz, complementary log log, arctangent and Burr (scobit) models described in Chapter E27. A like result holds for the count models, Poisson, negative binomial, etc. presented in Chapter E41 (and the extensions in Chapters E42-E44) for which the link is simply the log function.

E49.2.1 Internally Consistent Generalized Linear Models

One can create a vast array of models by crossing a menu of link functions with a second menu of distributional families. (As shown below, *LIMDEP* offers at least 25 different distributional families.) Consider, for example, the following matrix of a few possibilities.

		Link Functions				
Kind of r.v.	**Family**	Identity	Logit	Probit	Log	Reciprocal
Binary	Bernoulli	X	•	•	X	X
Continuous	normal	•	•	•	•	•
Count	Poisson	X	X	X	•	X
Nonnegative	gamma	X	X	X	•	X

Table E49.1 Families and Link Functions for Generalized Linear Models

We choose four distributional families to provide models for the indicated kinds of random variables and five link functions. There is no theoretical restriction on the mesh between links and families. But, in fact, most of the combinations are internally inconsistent. For example, for the binary dependent variable, only the probit and logit links make sense; the others imply a conditional mean that is not bounded by zero and one. For the continuous random variable, any link could be chosen; this just defines a linear or nonlinear regression model. For the count variable, only the log transformation insures an appropriate nonnegative mean. The logit and probit transformations also imply a positive mean, but one would not want to formulate a model for counts that forces the conditional mean function to be a probability between zero and one, so these make no sense either. The exact same considerations rule out all but the log transformation for the gamma family. The preceding lists most of the commonly used link functions. More than half of our table is null. Of the nine combinations that are internally consistent, five are just nonlinear regressions. But, the nonlinear regression model is a much broader class than this, and one would unduly restrict the model if they limited it to the GLIM framework for nonlinear regression analysis. The end result of this development is that typically, only one link function is appropriate for most of the distributional families. (Similar analyses appear in other popular programs such as *SAS* and *Stata*. In general, the matrix of model combinations is usually about one third full, with most cells containing unusable or inconsistent combinations such as the ones noted above.)

The upshot of all this is that you can fit nearly all of the internally consistent 'generalized linear models in common use' – partly because in the end, the set of them is surprisingly small. The width of the class is deceptive because of this consideration of consistency of the model and the specification of the conditional mean function.

E49.2.2 The Similarity of Different Link Functions

The generic form of the GLIM implies that

$$E[y \mid \mathbf{x}] = h(\boldsymbol{\beta}'\mathbf{x}).$$

As noted in many previous applications, the implication of this is that while coefficients in different forms of the models for a given dependent variable may differ substantially, the differences often disappear (or nearly so), when one computes the partial effects. Generally, for index function models, the partial effects are scaled versions of the structural coefficient vector;

$$\boldsymbol{\delta} = \frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = h'(\boldsymbol{\beta}'\mathbf{x})\boldsymbol{\beta}\ .$$

The different scale factors tend to eliminate the differences in the associated parameter vectors. The effect is strikingly persistent in binary choice modeling, but it is likewise prevalent more generally in the analysis of generalized linear models. Consider an example, based on the German health care data analyzed earlier and again in the sections to follow. Suppose y is income, which we model with an exponential regression model,

$$p(y_{it} \mid \mathbf{x}_{\mathrm{it}}) = \lambda_{it} \exp(-y_{it}\, \lambda_{it}).$$

Then, this constitutes the 'family' of distributions. For this model, $E[y_{it}|\mathbf{x}_{it}] = 1/\lambda_{it}$. We consider two possible 'link' functions, the log function, for which $\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it})$ and the identity function, $\lambda_{it} = \boldsymbol{\beta}'\mathbf{x}_{it}$. For the first of these, $\boldsymbol{\delta}_{it} = (-1/\lambda_{it})\boldsymbol{\beta}$, while for the second, $\boldsymbol{\delta}_{it} = (-1/\lambda_{it}^2)\boldsymbol{\beta}$. The following program does these computations for a model of incomes and displays the coefficients and the marginal effects.

```
SAMPLE      ; All $
REJECT      ; _groupti <  7 $
NAMELIST    ; x = one,age,educ,hhkids,female,married $
CALC        ; k = Col(x) $
MAXIMIZE    ; Labels = k_blog ; Start = k_0
            ; Fcn = bx = blog1'x | ti = Exp(bx) | Log(ti) - hhninc*ti  $
MATRIX      ; xb = Mean(x) $
CALC        ; eb = -Exp(-b'xb) $
MATRIX      ; deltae = eb*b $
CALC        ; k1 = k-1 ; yb1 = 1/Xbr(hhninc) $
MAXIMIZE    ; Labels = k_biden ; Start = yb1,k1_0
            ; Fcn = bx = b1'x | ti = bx | Log(ti) - hhninc*ti $
CALC        ; eb = -1/(b'xb)^2 $
MATRIX      ; deltal = eb*b ; List ; deltas = [deltae,deltal] $
```

Though the coefficient vectors appear to be quite different, the marginal effects are, in fact, close to the same. Moreover, the pattern of significance in the coefficients is the same as well. The upshot, as illustrated in this example, is that there is generally little impact of the choice of the link function on quantities usually of interest in the model. However, there is a cost to imposing the restriction of an internally inconsistent conditional mean on a model, for example, in forcing the mean of a Poisson variable to be a probability.

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable                 Function
Log likelihood function           395.15949
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
   BLOG1|    2.06290***      .19758    10.44  .0000     1.67564   2.45016
   BLOG2|    -.00595*        .00339    -1.76  .0787     -.01259    .00068
   BLOG3|    -.05285***      .01384    -3.82  .0001     -.07996   -.02573
   BLOG4|     .06976         .05980     1.17  .2434     -.04744    .18697
   BLOG5|     .01509         .05818      .26  .7954     -.09895    .12913
   BLOG6|    -.23094***      .06558    -3.52  .0004     -.35947   -.10241
--------+--------------------------------------------------------------------
User Defined Optimization
Dependent variable                 Function
Log likelihood function           396.18120
```

```
--------+-------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
UserFunc|  Coefficient       Error        z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
  BIDEN1|     5.73701***      .62302     9.21  .0000     4.51591    6.95811
  BIDEN2|     -.01789*        .01053    -1.70  .0892     -.03852     .00274
  BIDEN3|     -.13524***      .03641    -3.71  .0002     -.20660    -.06388
  BIDEN4|      .16315         .18231      .89  .3709     -.19418     .52047
  BIDEN5|      .03030         .17374      .17  .8615     -.31022     .37082
  BIDEN6|     -.73576***      .21886    -3.36  .0008    -1.16471    -.30680
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
  DELTAS|              1              2
--------+-----------------------------
       1|       -.712105       -.668612
       2|      .00205501      .00208502
       3|       .0182427       .0157610
       4|      -.0240819      -.0190135
       5|     -.00520873     -.00353141
       6|       .0797201       .0857480
```

The similarity of these effects seems to be little noted in the literature and on websites that discuss the generalized linear models. For example, displays such as that in Figure E49.1 are meant to suggest the difference between the identity link E[y|**x**] = $\boldsymbol{\beta}'\mathbf{x}$ and log link E[y|**x**] = exp($\boldsymbol{\beta}'\mathbf{x}$).

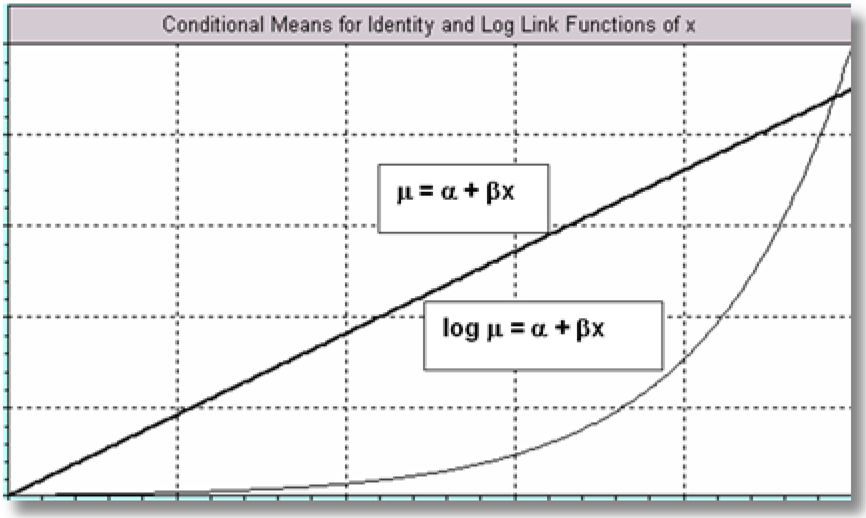


Figure E49.1 Link Functions for GLMs

But, these figures vastly exaggerate what will occur when the methods are applied to a model and a data set. The scaling by the coefficients, and the inherent relationships among the variables will obscure the differences in functional form. Consider the earlier example. We now refit this model as a normal family regression model with identity and log link functions, and the same regressors. We then hold the other variables constant at their means, and plot the conditional mean functions as a function of *age*. The figure shows that the impact of the choice of link function is minor.

The commands are:

```
NAMELIST   ; x0 = one,educ,hhkids,female,married $
REGRESS    ; Quietly ; Lhs = hhninc ; Rhs = x0,age $
MATRIX     ; b0i = b(1:5) ; xb0 = Mean(x0) $
CALC       ; b6i = b(6) ; ai = b0i'xb0 $
GLIM       ; Quietly ; Lhs = hhninc ; Rhs = x0,age ; Model = Normal $
MATRIX     ; b0l = b(1:5) $
CALC       ; b6l = b(6) ; al = b0l'xb0 $
SAMPLE     ; 1-40 $
CREATE     ; years = Trn(25,1)$
CREATE     ; yf_iden = ai+b6i*years $
CREATE     ; yf_log = Exp(al+b6l*years) $
PLOT       ; Lhs  = years ; Rhs = yf_iden,yf_log ; Fill
           ; Title = Conditional Means for Log and Identity Links
           ; Grid ; Yaxis = E[y|x] $
```

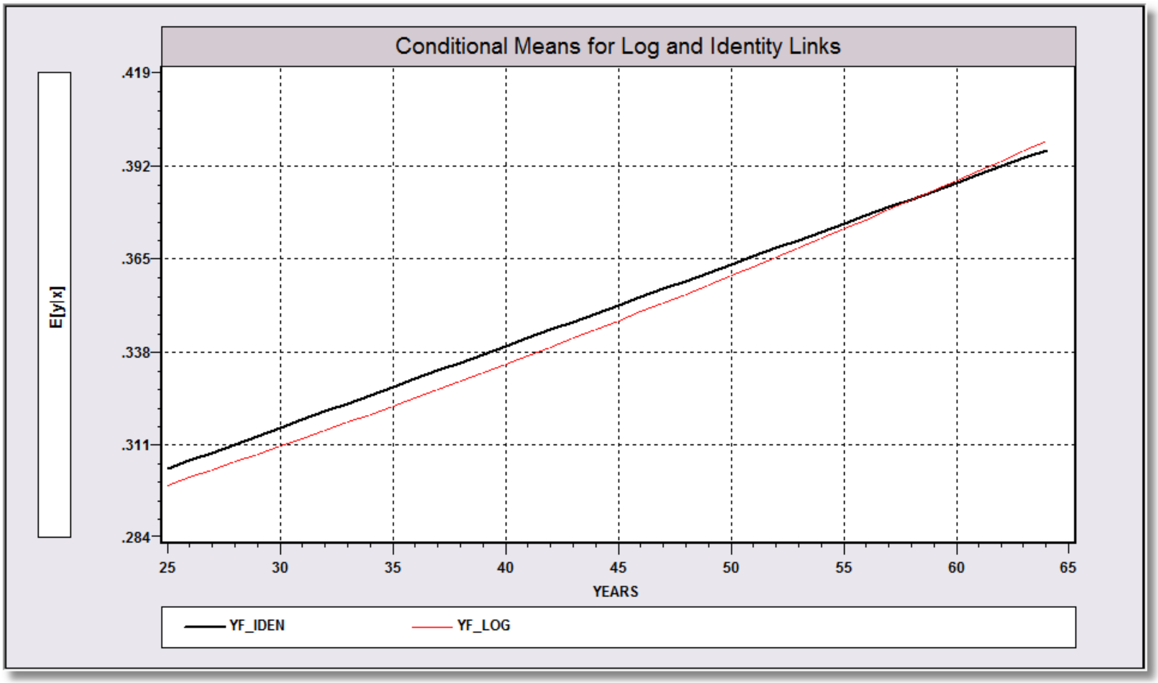


Figure E49.2 Estimated Conditional Means for Two Link Functions

E49.2.3 Estimation Methods

One of the useful byproducts of the early development of the generalized linear models methodology was an implied estimation technique, which has been labeled 'iteratively reweighted least squares,' or IRWLS. Since the conditional mean function is identified, it implies a kind of nonlinear weighted least squares estimator. Among the virtues of the GLM specifications is that this weighted least squares procedure is quite simple to compute – at least simple in that it should be easy to program. In most cases, at convergence, IRWLS will produce the maximum likelihood estimator. Since the GLIM estimator is based on a fully specified parametric model of the density for the observed random variable, the maximum likelihood estimator is fully efficient, and, where it differs from the MLE, the IRWLS estimator is not. (This will be the case in models that contain ancillary parameters, such as the overdispersion parameter in the negative binomial model.) All of *LIMDEP*'s estimators for these models are MLEs not weighted least squares estimators. (You may, as in other models, provide any observation specific weights you wish, but these are applied to terms in the log likelihood, not in any form of least squares.)

E49.2.4 Generalized Linear Models

As noted earlier, many of the single index function models described in the preceding chapters fall under the definition of GLMs in that the conditional mean function is a function of the linear index. However, in some of these cases, such as the censored regression, the inverse transformation to the index will be difficult or impossible to obtain. Table E49.2 lists most of *LIMDEP*'s GLMs – some not listed here are documented elsewhere in the manual, though perhaps not specifically identified as GLMs. The table also indicates the subset of the indicated families that are typically analyzed in the formal literature on GLMs. As can be seen, we have extended the class a bit. The third column defines the one 'link function' used in each of these cases. As discussed earlier, typical tabulations in the literature provide a menu of link functions. However, in most cases, only a single link function makes sense in any particular context. Moreover, in actual practical terms, except where they impose a strong inappropriate restrictions on the model – such as using a probability as the conditional mean in a count or regression model – different link functions will produce similar empirical results.

As shown in the top half of Table E49.2, many of the models have already been documented in earlier chapters. All of the models above may be requested with the command

GLIM ; Lhs = ... ; Rhs = ... ; Model = the model name given above $

For those models with command names given in the top half of the table, the command will be the same as if you had used the earlier command. That is, for example, the following two commands are identical:

GLIM ; Lhs = y ; Rhs = x ; Model = Logit $
LOGIT ; Lhs = y ; Rhs = x $

The remaining models may all be invoked with the **GLIM** command. For these ten models (only), the command **GLIM** is also synonymous with **LOGLINEAR**. So, for example, the following two commands are identical:

GLIM ; Lhs = y ; Rhs = x ; Model = Beta $
LOGLINEAR ; Lhs = y ; Rhs = x ; Model = Beta $

In the developments below, the results for the first group of models will only be sketched. The reader is directed to the full, earlier chapters on the subjects. Most of this chapter will relate to the additional models detailed in the lower half of the table – the 'loglinear' models.

Model	**Dependent Variable**	**Conditional Mean**	**Command**
Models developed in preceding chapters			
Probit[b]	Binary	$\Phi(t)$	**PROBIT**
Logit[b]	Binary	$\Lambda(t)$	**LOGIT**
Gompertz[b]	Binary	$\exp(-\exp(-t)$	**GOMPERTZ**
Comp. log log[b]	Binary	$1 - \exp(- \exp(t))$	**LOG**
Arctangent	Binary	$2/\pi$ Arctan(exp(t))	**ARCTANGENT**
Burr	Binary	$\Lambda(t)^{\gamma}$	**BURR**
Poisson[b]	Count	$\exp(t)$	**POISSON**
NB1 Neg. Bin.	Count	$\exp(t)$	**NEGBIN**
NB2 Neg. Bin[b].	Count	$\exp(t)$	**NEGBIN**
NBP Neg. Bin.	Count	$\exp(t)$	**NEGBIN**
Polya-Aeppli	Count	$\exp(t)/(1-\theta)$	**POISSON**
GP, Generalized Poisson	Count	$\exp(t)$	**POISSON**
PGamma, Poisson/Gamma	Count	$\exp(t)/\alpha$ (approx.)	**POISSON**
Linear[b]	Continuous	t	**REGRESS**
Loglinear models developed in this chapter			
Lognormal	Nonnegative	t	**LOGNORMAL**
Binomial[b]	Count of successes	$K\Lambda(t)$	
Geometric[b]	Count until success	$\exp(t)$	
Beta	Bounded in (0,1)	$\exp(t_a)/[\exp(t_a)+\exp(t_b)]$	
Power	Bounded in (0,1)	$[\exp(t)+1]/[\exp(t)+2]$	
Normal (Loglinear)	Continuous	$\exp(t)$	
Gamma	Nonnegative	$P\exp(-t)$	
Weibull	Nonnegative	$[\exp(-t)]^{1/P}\Gamma[(P+1)/P]$	
Exponential[b]	Nonnegative	$\exp(-t)$	
Rayleigh	Nonnegative	$[\pi\exp(-t)/2)]^{1/2}$	
Inverse Gaussian[b]	Nonnegative	$P\exp(-t)$	

[a] In all models, $t = \boldsymbol{\beta}'\mathbf{x}$.
[b] Exponential families typically included in analysis of 'Generalized Linear Models'

Table E49.2 Generalized Linear Models[a]

E49.2.5 Residual Analysis

Many types of 'residuals' are suggested for model assessment of GLMs. Three that have some useful characteristics are the 'Pearson residual,' the 'deviance residual' and Chesher and Irish's generalized residuals. Unfortunately, none of the three are useful for all of the models considered here, though they do come close. The deviance residuals are based on models for which the log likelihood can be written in terms of the conditional mean function. For this computation, the estimated model is compared to one in which y_i is predicted perfectly at every observation. Thus, in computing the log likelihood for the 'saturated' model, we replace the estimator of the conditional mean with the actual value of y_i. Thus, the deviance measures the extent to which the model fails to predict perfectly. The 'deviance residual' is

$$e_{D,i} = \log L_i(y_i) \ - \ \log L_i(\hat{y}_i)$$

where $\hat{y}_i$ is the model prediction of y_i using the estimated parameters to compute the conditional mean function. The 'deviance' for the model is

$$D \ = 2\ \Sigma_i\ d_i \ = \ -2\times [\Sigma_i \log L_i(\hat{y}_i) - \Sigma_i \log L_i(y_i)].$$

Consider two examples. For a binary choice model, the conditional mean is $\hat{P}_i$ based on whatever model is used to estimate the probability model. The model that produces a perfect fit would have $\hat{P}_i = y_i$. Therefore, the deviance residual would be

$$e_{D,i} \ = \ [(1-y_i)\log(1-y_i) + y_i\log y_i] - [(1-y_i)\log(1-\hat{P}_i) + y_i\log\hat{P}_i]$$

(where 0log0 = 0). The first term in square brackets is zero. The model deviance would be

$$D \ = \ -2\ \Sigma_i\ [(1-y_i)\log(1-\hat{P}_i) + y_i\log\hat{P}_i]$$

which is just -2 times the log likelihood for the model. Second, consider a Poisson model, in which $\hat{y}_i = \hat{\lambda}_i$. The deviance residual would be

$$\begin{aligned} e_{D,i} \ &= \ [-y_i + y_i\log y_i - \log\Gamma(y_i+1)] - [-\hat{\lambda}_i + y_i\log\hat{\lambda}_i - \log\Gamma(y_i+1)] \\ &= \ -(y_i - \hat{\lambda}_i) + y_i\log(y_i/\hat{\lambda}_i) \end{aligned}$$

(where as before 0log0 = 0). The deviance for the model is

$$D \ = \ -2\Sigma_i\ [(y_i - \hat{\lambda}_i) - y_i\log(y_i/\hat{\lambda}_i)].$$

These measures are not computed internally for the models. However, they are easily computed using the predictions from the models. Catalogs of formulas for many generalized linear models can be found in the vast literature on GLMs.

Deviance measures of 'fit' compute in the opposite direction from familiar measures of fit. For example, in linear models, the R^2 compares the estimated model to a model that provides no fit. Likewise, the so called 'pseudo R^2' for maximum likelihood estimation, 1 - log$L$/logL_0 compares the estimated model to one which has no coefficients other than a constant term. Again, the intent is to compare the estimated model to one which provides no fit. The deviance measure, in contrast, compares an estimated model to one which predicts the dependent variable perfectly. The scale of the measure is unclear. For example, for a binary choice model, the deviance is simply -2 times the log likelihood. For the Poisson (and other) models, the measure is not a simple function of the log likelihood. Moreover, it should be noted that the model itself, is not estimated in order to predict the dependent variable well with the estimated conditional mean function. For example, for a binary choice model, the maximum score estimator will outperform any MLE. Thus, it remains ambiguous what is being computed by the deviance measures.

The second residual of interest is the 'Pearson residual,'

$$e_{P,i} = \frac{y_i - \hat{y}_i}{\sqrt{V\hat{a}r(y_i)}}$$

In many treatments, the denominator is assumed to be a function of $\hat{y}_i$. We leave it in the more general form to accommodate those cases in which the conditional variance is not a simple function of the conditional mean. These can also be computed easily with the model results. The predictions are all available after estimation with **; Keep = variable name**. To complete the computation, the conditional variances are required. These are given in Table E49.3. These are saved for the models listed in the table when the commands for these models (only) contain

; Pres = variable name

Note that *LIMDEP* supports many variants of these models for which these residuals are not computed (and sometimes not computable). For example, by the various constructions in this chapter, the censored, truncated, zero inflated and hurdle versions of the Poisson and negative binomial models are all GLMs, however, they are not included in the set of models analyzed here.

A third useful quantity in some analyses is Chesher and Irish's (1987) 'generalized residual,' which for the models in which they are useful, can be computed as the derivative of the log likelihood with respect to the constant term. (For the normal linear regression model, it coincides with the Pearson residual above.) The quantity is useful for specification testing in latent regression models based on the normal distribution. Applications appear in Chapter E29 for the probit model and Chapter E47 with the development of the tobit model.

Finally, there are an array of variations on the Pearson and deviance residuals for the GLMs, such as the Anscombe residuals and variations thereon.

Model	Conditional Mean	Conditional Variance
Probit	$\Phi(t) = F$	$F(1 - F)$
Logit	$\Lambda(t) = \lambda/(1+\lambda) = F$	$F(1 - F)$
Gompertz	$\exp(-1/\lambda) = F$	$F(1 - F)$
Comp. log log	$1 - \exp(-\lambda) = F$	$F(1 - F)$
Arctangent	$1/\pi \text{ Arctan}(\lambda)$	$F(1 - F)$
Burr	$\Lambda(t)^{\gamma} = F$	$F(1 - F)$
Poisson	λ	λ
NB1 Neg. Bin.	λ	$\lambda(1 + \lambda)$
NB2 Neg. Bin.	λ	$\lambda(1 + \theta\lambda)$
NBP Neg. Bin.	λ	$\lambda(1 + \theta^{P-1}\lambda)$
Polya-Aeppli	$\lambda/(1-\theta)$	$\lambda(1 + \theta)/(1 - \theta)^2$
GP, Generalized Poisson	λ	$\lambda(1 + \theta\lambda)^2$
PGamma, Poisson/Gamma[b]	λ/α	λ/α^2
Linear	t	σ^2
Lognormal	t	$\sigma^2[t]^2$
Binomial	$K\Lambda(t) = K F$	$K F (1 - F)$
Geometric	λ	$\lambda(1 + \lambda)$
Beta	$\lambda_a/(\lambda_a + \lambda_b)$	$\lambda_a\lambda_b/[(\lambda_a+\lambda_b+1)(\lambda_a+\lambda_b)^2]$
Power	$(\lambda+1)/(\lambda+2)$	$(\lambda+1)/(\lambda+3) - [(\lambda+1)/(\lambda+2)]^2$
Normal (Loglinear)	λ	σ^2
Gamma	P/λ	P/λ^2
Weibull	$(1/\lambda)^{1/P}\Gamma[(P+1)/P)]$	$(1/\lambda)^{2/P}\{\Gamma[(P+2)/P)] - \Gamma^2[(P+1)/P]\}$
Exponential	$1/\lambda$	$1/\lambda^2$
Rayleigh	$[\pi/(2\lambda)]^{1/2}$	$(4 - \pi)/(2\lambda)$
Inverse Gaussian	P/λ	P/λ^3

[a] In all results given, $t = \boldsymbol{\beta}'\mathbf{x}$ and $\lambda = \exp(t) = \exp(\boldsymbol{\beta}'\mathbf{x})$.
[b] Mean and variance for the gamma count model are approximate based on increasing event window. See Winkelmann (2003, p. 55).

Table E49.3 Means and Variances of Variables in Generalized Linear Models[a]

E49.2.6 Standard Model Specifications for the Loglinear Regression Models

This is the full list of general specifications that are applicable to this group of model estimators.

Controlling Output from Model Commands

; Par	keeps ancillary parameter p in main results vector b.
; Margin	displays marginal effects.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec requests computation of the cluster form of corrected covariance estimator. (**; Stratum = specification** for stratified clustered data).
; Robust requests a 'sandwich' estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E49.2.7 Estimated Results for Loglinear Models

As noted earlier, the models in the top half of Table E49.2 are documented extensively in earlier chapters. For those in the bottom half, which are the main subject of this chapter, the following are the estimated results:

Matrices: *b* = estimates of β, or
b = estimates of α followed by β for the beta model
varb = estimated asymptotic covariance matrix for MLE of *b*

Scalars: *logl* = log likelihood
kreg = number of variables in Rhs
nreg = number of observations
pgamma = P for the gamma model
pweibull = P for the Weibull model
pinverse = P for the inverse Gaussian model
s = σ for the normal (exponential regression) model, or
s = σ for the lognormal regression model
exitcode

Last Model: for the **WALD** command using *b_name* for the parts of *b*

Last Function: conditional mean function

E49.3 Discrete Dependent Variable Models

Four types of discrete dependent variables are supported in the GLM group, binary, count, count of successes (binomial) and number of trials until first success (geometric). The fourth of these does not naturally describe a count outcome, but as shown in the application below, purely from the standpoint of a functional form, it might be preferable to the Poisson model.

E49.3.1 Binary Dependent Variables

Six parametric model formulations are provided as internal procedures in *LIMDEP* for binary choice models. The probability models and loglinear forms are shown in Table E49.4.

Model	Probability for $Y = 1$	Loglinear Form*
Probit	$F = \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_i} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt = \Phi(\boldsymbol{\beta}'\mathbf{x}_i)$	$F = \Phi(\log\lambda_i)$
Logit	$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)$	$F = \lambda_i / (1 + \lambda_i)$
Comp. log log	$F = 1 - \exp(-\exp(\boldsymbol{\beta}'\mathbf{x}_i)) = C(\boldsymbol{\beta}'\mathbf{x}_i)$	$F = 1 - \exp(-\lambda_i)$
Gompertz	$F = \exp(-\exp(-\boldsymbol{\beta}'\mathbf{x}_i)) = G(\boldsymbol{\beta}'\mathbf{x}_i)$	$F = \exp(-1/\lambda_i)$
Arctangent	$F = 2/\pi \text{ Arctan}(\exp((\boldsymbol{\beta}'\mathbf{x}_i))$	$F = 2/\pi \tan^{-1}(\lambda)$
Burr	$F = \left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right]^\gamma = [\Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]^\gamma, \gamma > 0$	$F = [\lambda_i / (1 + \lambda_i)]^\gamma$.

* $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$

Table E49.4 Binary Choice Models

These may be invoked with the command

GLIM ; Lhs = dependent variable
; Rhs = independent variables
; Model = Probit, Logit, Comploglog, Gompertz, Arctangent or Burr $

The **GLIM** command is a synonym for the respective commands for each of these, **PROBIT**, **LOGIT**, **COMPLOGLOG**, **GOMPERTZ, ARCTANGENT** and **BURR**. The **GLIM** command does not change the model request; it is merely an equivalent form. The binary choice models are documented in Chapter E27 for cross section and pooled data, and in Chapters E30 and E31 for the various panel data estimators.

E49.3.2 Count Variables

Chapters E41-E44 document a wide variety of models for counts. The basic platform is the Poisson model,

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!},\ y_i = 0,1,...;\ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i).$$

The crucial feature of the Poisson model is its equidispersion property,

$$\text{Var}[y_i|\lambda_i] = E[y_i|\lambda_i] = \lambda_i.$$

Many variations have been developed to relax the equidispersion assumption. The most popular is the negative binomial model (NB2), which has density

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i+1)} u_i^{\theta}(1-u_i)^{y_i}$$

where $\theta = 1/\alpha$

and $u_i = \theta / (\theta + \lambda_i)$.

In the negative binomial (NB2) model,

$$\text{Var}[y_i|\lambda_i] = E[y_i|\lambda_i]\{1 + \alpha\, E[y_i|\lambda_i]\}.$$

There are a variety of other forms for the negative binomial model, based on the relationship

$$\text{Var}[y_i|\lambda_i] = E[y_i|\lambda_i]\{1+ \alpha^{P\text{-}1} E[y_i|\lambda_i]\}.$$

The NB2 model above has $P = 2$. The NB1 model, with $P = 1$, has density

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(\theta\lambda_i)\Gamma(y_i+1)} w^{\theta\lambda_i}(1-w\ )^{y_i}$$

where $w = \theta / (\theta + 1)$.

The general form of the model is the NBP form, which has density

$$\text{Prob}(Y = y_i|\mathbf{x}_i) \quad = \quad \frac{\Gamma(\theta\lambda_i^Q + y_i)}{\Gamma(\theta\lambda_i^Q)\Gamma(y_i+1)}\left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q+\lambda_i}\right)^{\theta\lambda_i^Q}\left(\frac{\lambda}{\theta\lambda_i^Q+\lambda_i}\right)^{y_i}$$

where $Q = P - 2$.

Several other model forms for counts are supported. The gamma (based) probability model is

$$\text{Prob}[y_i = j] \quad = \quad G(\alpha j, \lambda_i) - G(\alpha j + \alpha, \lambda_i)$$

where $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$ (as usual)

and $G(\alpha j, \lambda_i) = 1$ if $j = 0$, or $\dfrac{1}{\Gamma(\alpha j)}\int_0^{\lambda_i} u^{\alpha j-1}e^{-u}du$ if $j > 0, j = 1,...$

The dispersion parameter is α; there is underdispersion if $\alpha > 1$, overdispersion if $\alpha < 1$, and equidispersion if $\alpha = 1$, which reduces the gamma probability to the Poisson model. The gamma distributed count variable may be underdispersed or overdispersed. Underdispersion is usually of lesser interest.

The generalized Poisson model is another that has overdispersion. The density for the generalized Poisson model is

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \left(\frac{\lambda_i}{1+\theta\lambda_i}\right)^{y_i}\frac{(1+\theta y_i)}{y_i!}\exp\left(-\frac{\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\right), y_i = 0,1,2,...;\ \lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}.$$

The mean and variance of this random variable are

$$E[y_i \mid \mathbf{x}_i] = \lambda_i,\ \text{Var}[y_i \mid \mathbf{x}_i] = \lambda_i(1+\theta\lambda_i)^2.$$

Finally, the density for the Polya-Aeppli form of the Poisson model is

$$\text{Prob}[Y = 0 \mid \mathbf{x}_i] = \exp(-\lambda_i),$$

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \exp(-\lambda_i)(1-\theta)^{y_i}\sum_{j=1}^{y_i}\binom{y_i-1}{j-1}\frac{[\lambda_i(1-\theta)/\theta]^j}{j!}, y_i = 1,2,...$$

$$\lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}$$

$$0 < \theta < 1.$$

The mean and variance of this random variable are

$$E[y_i \mid \mathbf{x}_i] = \frac{\lambda_i}{(1-\theta)} \text{ and } \text{Var}[y_i \mid \mathbf{x}_i] = \frac{\lambda_i(1+\theta)}{(1-\theta)^2} = E[y_i \mid \mathbf{x}_i]\frac{(1+\theta)}{(1-\theta)}.$$

All of these models are requested with the command

POISSON ; Lhs = … ; Rhs = … ; Model = NegBin, NB1, NB2, Polya, etc. $

E49.3.3 Number of Successes in *K* Trials – The Binomial Regression Model

The binomial regression model describes the number of success in *K* trials. The model is supported for both cross section and panel data applications. The discrete probability model is

$$\text{Prob}(Y = y_i \mid K_i, \mathbf{x}_i) = \binom{K_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{K_i - y_i},\ y_i = 0,1,\ldots,K_i$$

$$\theta_i = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \frac{\lambda_i}{1+\lambda_i}, 0 < \theta_i < 1.$$

The success probability on any single trial is θ_i. The number of trials may differ at every observation, or may be constant. The conditional mean function is

$$E[y_i|K_i,\mathbf{x}_i] = K_i\theta_i,$$

so the vector of marginal effects is

$$\boldsymbol{\delta}_i = \partial E[y_i|K_i,\mathbf{x}_i]/\partial \mathbf{x}_i = K_i\theta_i (1-\theta_i)\boldsymbol{\beta}.$$

These can be averaged over observations or computed at the sample means, as usual. The command for this model is

```
LOGLINEAR ; Lhs = y ; Rhs = …
          ; Model = Binomial
          ; Trials = specification
          ; ... other options $
```

The **; Trials** definition is the same as a panel data declaration. If the number is constant, then that number is given. If the number is variable, then the name of the variable is provided instead.

A Zero Inflated Binomial Model

An extension to the binomial regression model that allows the zero probability to be inflated would be

$$\text{Prob}(Y = 0 \mid K_i, \mathbf{x}_i) = (1-\Lambda_i) + \Lambda_i \binom{K_i}{0} \theta_i^{0} (1-\theta_i)^{K_i - 0}$$

$$= (1-\Lambda_i) + \Lambda_i (1-\theta_i)^{K_i}$$

$$\text{Prob}(Y = y_i \mid K_i, \mathbf{x}_i) = \Lambda_i \binom{K_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{K_i - y_i},\ y_i = 1,\ldots,K_i$$

$$\theta_i = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \frac{\lambda_i}{1+\lambda_i}, 0 < \theta_i < 1$$

$$\Lambda_i = \frac{\exp(\boldsymbol{\gamma}'\mathbf{z}_i)}{1+\exp(\boldsymbol{\gamma}'\mathbf{z}_i)} = \frac{\phi_i}{1+\phi_i}, 0 < \Lambda_i < 1$$

This formulation shifts some of the mass to the zero outcome. This is a counterpart to the 'zero inflated Poisson model' presented in Chapter E43. The 'regime' probability, Λ_i, is taken to be a logit model. It may involve covariates $\mathbf{z}_i$ or it may be a constant. This model is requested with

LOGLINEAR ; Lhs = y
; Rhs = …
; Model = Binomial
; Trials = specification
; ZIB = list of variables in z
; ... other options $

If the zero inflation probability is to be a constant, then use **; ZIB = one**.

Marginal effects in this model are exerted by the variables in both parts of the probability. The conditional mean is

$$E[y_i|K_i,\mathbf{x}_i] = \Lambda_i K_i \theta_i,$$

so the marginal effects, assuming that variables in $\mathbf{x}_i$ might also appear in $\mathbf{z}_i$, are

$$\boldsymbol{\delta}_i = \partial E[y_i|K_i,\mathbf{x}_i]/\partial \mathbf{x}_i = \Lambda_i K_i \theta_i (1-\theta_i)\boldsymbol{\beta} + \Lambda_i(1-\Lambda_i)K_i\theta_i\boldsymbol{\gamma}.$$

Variables which appear in both $\mathbf{x}_i$ and $\mathbf{z}_i$ exert both terms; those only in $\mathbf{x}_i$, the first, and those only in $\mathbf{z}_i$, only the second.

Technical Details

The log likelihood for the binomial regression model is

$$\log L = \sum_{i=1}^{N} \log \frac{\Gamma(K_i+1)}{\Gamma(y_i+1)\Gamma(K_i-y_i+1)} + y_i \log\theta_i + (K_i - y_i)\log(1-\theta_i)$$

The derivatives of the log likelihood are

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left(\frac{y_i}{\theta_i} - \frac{K_i - y_i}{1-\theta_i}\right)\frac{\partial\theta_i}{\partial\boldsymbol{\beta}}$$

The necessary derivative is $\partial\theta_i/\partial\boldsymbol{\beta} = \theta_i(1-\theta_i)\mathbf{x}_i$. Collecting terms, then

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left(\frac{y_i}{\theta_i} - \frac{K_i - y_i}{1-\theta_i}\right)\theta_i(1-\theta_i)\mathbf{x}_i$$
$$= \sum_{i=1}^{N} \left(y_i(1-\theta_i) - (K_i - y_i)\theta_i\right)\mathbf{x}_i$$

The second derivatives are obtained by using the earlier result and collecting terms;

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{N} \; -K_i \theta_i (1-\theta_i) \mathbf{x}_i \mathbf{x}_i'$$

The same result can be used to obtain the derivatives for the computation of the covariance matrix of the marginal effects. The necessary term is

$$\mathbf{G}_i = \partial \boldsymbol{\delta}_i / \partial \boldsymbol{\beta}' = K_i \theta_i (1-\theta_i)(1 - 2\theta_i) \boldsymbol{\beta} \mathbf{x}_i'.$$

For the zero inflated model,

$$\log L = \sum_{y_i=0} \log\left[(1-\Lambda_i) + \Lambda_i (1-\theta_i)^{K_i}\right]$$
$$+\sum_{y_i>0} \; \log \Lambda_i + \log \frac{\Gamma(K_i+1)}{\Gamma(y_i+1)\Gamma(K_i - y_i + 1)} + y_i \log \theta_i + (K_i - y_i) \log(1-\theta_i)$$

For convenience, call the bracketed term in the first part P_0. Then,

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{y_i=0} \; \left[\frac{-\Lambda_i (1-\theta_i)^{K_i} \theta_i K_i}{P_0}\right] \mathbf{x}_i + \sum_{y_i>0} \; \left[y_i(1-\theta_i) - (K_i - y_i)\theta_i\right] \mathbf{x}_i$$

$$\frac{\partial \log L}{\partial \boldsymbol{\gamma}} = \sum_{y_i=0} \; \left[\frac{\Lambda_i (1-\Lambda_i)\left[(1-\theta_i)^{K_i} - 1\right]}{P_0}\right] \mathbf{z}_i + \sum_{y_i>0} \; (1-\Lambda_i) \mathbf{z}_i$$

The Jacobian for the marginal effects is tedious, but derivable based on earlier results.

Application

To illustrate the binomial regression model estimator, we will simulate a data set that satisfies the assumptions of the model. We begin with the regressors, a continuously (normally) distributed variable and a binary (dummy) variable. The 'regression' model is

$$\lambda_i = \exp(-.5 + 1x_{1i} + 1x_{2i}).$$

The success probabilities are generated as logistic probabilities, $\theta_i = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)$. We then generate the number of trials, K_i, for each individual, using a random draw from the integers 3, 4, 5 and 6. With the number of trials and the success probabilities in hand, we use the built in random number generators to obtain a sample from the observation specific binomial distribution. The last three commands estimate the model, then use a Lagrange multiplier test to test for the joint significance of the two regressors. (The model with only a constant term converges without iterating, because the starting values for the iterations are the MLEs for a model with only a constant term. These results are omitted below.)

The commands are:

```
CALC    ; Ran(12345) $
CREATE  ; x1 = Rnn(0,1) ; x2 = Rnu(0,1)>.5 $
CREATE  ; bx = -.5 + x1 + x2 $
CREATE  ; thetai = Lgp(bx) $
CREATE  ; ki = Rnd(4)+2 $
CREATE  ; y = Rnb(ki,thetai) $
GLIM    ; Lhs = y ; Rhs = one,x1,x2 ; Model = Binomial ; Trials = ki $
GLIM    ; Lhs = y ; Rhs = one ; Model = Binomial ; Trials = ki $
GLIM    ; Lhs = y ; Rhs = one,x1,x2 ; Model = Binomial ; Trials = ki
        ; Start = b,0,0  ; Maxit = 0 $
```

Based on the LM statistic, the hypothesis that the two coefficients are zero is rejected. This might have been expected, given the 't statistics' shown with the first set of results.

```
-----------------------------------------------------------------------------
Binomial (Loglinear) Regression Model
Dependent variable                   Y
Log likelihood function    -43054.94415
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    -.48972***       .00824   -59.41  .0000     -.50587   -.47356
      X1|     .99027***       .00690   143.44  .0000      .97674   1.00381
      X2|     .99523***       .01172    84.89  .0000      .97225   1.01821
--------+--------------------------------------------------------------------
Binomial (Loglinear) Regression Model
Dependent variable                   Y
Log likelihood function    -59413.01283
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     .00787          .00516     1.52  .1274     -.00225    .01798
--------+--------------------------------------------------------------------
Binomial (Loglinear) Regression Model
Dependent variable                   Y
LM Stat. at start values    29393.43182
Log likelihood function    -59413.01283
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     .00787          .00732     1.07  .2829     -.00649    .02222
      X1|        0.0          .00518      .00 1.0000 -.10157D-01  .10157D-01
      X2|        0.0          .01032      .00 1.0000 -.20225D-01  .20225D-01
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E49.3.4 Number of Trials Until Success – The Geometric Regression Model

This model is suitable for a discrete random variable whose values decay geometrically. For example, using the health care data analyzed earlier, the following histogram shows the pattern of the dependent variable, number of doctor visits. Though these data are appropriately modeled using a count model such as the Poisson, the pattern in the histogram is that of a variable that is generated by one with a geometric distribution.

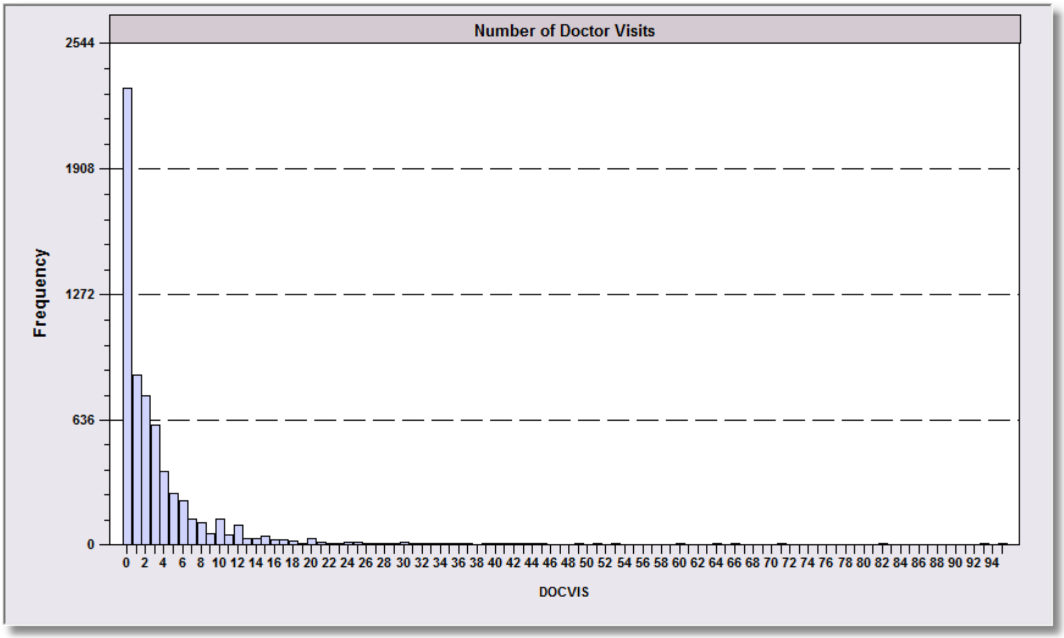


Figure E49.3 Histogram for Doctor Visits

The geometric regression model characterizes a sequences of Bernoulli trials in which the random variable y is the number of failures that occur until the first success occurs. The density is

$$f(y) = \theta(1-\theta)^y$$

where θ is the assumed constant probability of success on each trial. We parameterize the model for regression analysis with

$$\theta_i = \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \frac{1}{1+\lambda_i}.$$

Then $\text{Prob}(Y = y_i) = \lambda_i^{y_i}(1+\lambda_i)^{-(1+y_i)},\ y_i = 0,1,..$

This variable has $E[y_i|\mathbf{x}_i] = \lambda_i$ and $Var[y_i|\mathbf{x}_i] = \lambda_i(1 + \lambda_i)$.

In this model, the conditional mean function is λ_i, so the marginal effects are $\boldsymbol{\delta}_i = \lambda_i\boldsymbol{\beta}$.

The model is requested with

LOGLINEAR ; Lhs = y
; Rhs = variables in x
; Model = Geometric $

Marginal effects, fitted values, restrictions, the cluster estimator, robust covariance matrices, residuals, etc. for all other program features operate as usual.

In the application below, the geometric and Poisson regression models are found to give similar results for the panel of data on hospital visits. However, the Vuong test based on the likelihood functions seems strongly to favor the geometric model.

SAMPLE ; All $
REJECT ; _groupti < 7 $
NAMELIST ; x = one,age,hhninc,hhkids $
LOGLINEAR ; Lhs = docvis ; Rhs = x
; Model = Geometric ; Partial Effects $
CREATE ; lg = logl_obs $
POISSON ; Lhs = docvis ; Rhs = x ; Partial Effects $
CREATE ; lp = logl_obs $
CREATE ; d = lp - lg $
CALC ; List ; v = Sqr(n)*Xbr(d) / Sdv(d) $

```
-----------------------------------------------------------------------------
Geometric (Loglinear) Regression Model
Dependent variable               DOCVIS
Log likelihood function    -14037.19620
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     .16510*        .09139     1.81  .0708      -.01403    .34423
     AGE|     .02590***      .00179    14.43  .0000       .02238    .02941
  HHNINC|    -.49160***      .09013    -5.45  .0000      -.66825   -.31496
  HHKIDS|    -.07532**       .03345    -2.25  .0244      -.14088   -.00975
--------+--------------------------------------------------------------------
Poisson Regression
Dependent variable               DOCVIS
Log likelihood function    -23461.40921
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .12493***      .04815     2.59  .0095       .03055    .21931
     AGE|     .02717***      .00091    29.75  .0000       .02538    .02896
  HHNINC|    -.53730***      .04684   -11.47  .0000      -.62910   -.44550
  HHKIDS|    -.07968***      .01686    -4.73  .0000      -.11272   -.04663
--------+--------------------------------------------------------------------
```

```
(Geometric)
--------------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point   3.0262
Scale Factor for Marginal Effects  3.0262
--------+-----------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AGE|     .07837***      .00549    14.27  .0000       .06761    .08913
  HHNINC|   -1.48769***      .27316    -5.45  .0000     -2.02308   -.95231
  HHKIDS|    -.22792**       .10126    -2.25  .0244      -.42639   -.02945
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
(Poisson)
--------+-----------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  DOCVIS|      Effect        Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AGE|     .08515***      .00293    29.10  .0000       .07941    .09088
  HHNINC|   -1.68390***      .14729   -11.43  .0000     -1.97257  -1.39522
  HHKIDS|    -.24761***      .05197    -4.76  .0000      -.34947   -.14574   #
--------+-----------------------------------------------------------------------
#  Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
[CALC] V       =    -17.2471538
```

Technical Details

The log likelihood function is

$$\log L = \sum_{i=1}^{N} y_i \log \lambda_i - (y_i + 1)\log(1+\lambda_i)$$

The derivatives are

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left[y_i - \frac{\lambda_i (y_i+1)}{(1+\lambda_i)} \right] \mathbf{x}_i = \sum_{i=1}^{N} \left[y_i - (y_i+1)(1-\theta_i) \right] \mathbf{x}_i$$

and

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{N} -\left[(y_i+1)\theta_i (1-\theta_i) \right] \mathbf{x}_i \mathbf{x}_i' .$$

Since $E[y_i|\mathbf{x}_i] = \lambda_i$, the partial effects are $\lambda_i \boldsymbol{\beta}$ and the Jacobian is $\mathbf{G}_i = \lambda_i[\mathbf{I} + \boldsymbol{\beta}\mathbf{x}_i']$.

E50: Generalized Linear Models – 2: Continuous

E50.1 Introduction

This chapter presents a group of ‘generalized linear models’ (GLMs) that can be used for dependent variables whose range is generally restricted, either because they are discrete (such as a binary variable) or because they naturally vary over only a restricted range (such as variables that are only nonnegative). The class of generalized linear models was defined in the pioneering works of Nelder and Wedderburn (1972) and McCullagh and Nelder (1983). As shown below, many of these are models that *LIMDEP* fits under a different heading, but it is convenient to group them here. Formally, as defined in the recent literature, the class of models is a group in which the conditional mean function is of the form $E[y|\mathbf{x}] = h(\boldsymbol{\beta}'\mathbf{x})$ for some continuous function $h(.)$. (McCullagh and Nelder and others since have focused on ‘exponential families,’ but we take some license here, and broaden their class.) This class includes all of the single index function models already considered, such as the binary choice models, censored regression, truncated regression, and all of the count models considered in Chapters E24-E26. This chapter will present a group of models not already considered and also organize several from earlier chapters for the convenience of the user interested in this class of models.

The basic command for estimation of the models described in this chapter is

GLIM ; Lhs = dependent variable
; Rhs = independent variables
; Model = type of model $

where ‘**type of model**’ is one of the generalized linear models presented here.

E50.2 Generalized Linear Models for Continuous Variables

As noted earlier, many of the single index function models described in the preceding chapters fall under the definition of GLMs in that the conditional mean function is a function of the linear index. However, in some of these cases, such as the censored regression, the inverse transformation to the index will be difficult to obtain. Table E50.1 lists most of *LIMDEP*’s GLMs – some not listed here are documented elsewhere in the manual, though perhaps not specifically identified as GLMs. The table also indicates the subset of the indicated families that are typically analyzed in the formal literature on GLMs. As can be seen, we have extended the class a bit. The third column defines the one ‘link function’ used in each of these cases. As discussed earlier, typical tabulations in the literature provide a menu of link functions. However, in most cases, only a single link function makes sense in any particular context. Moreover, in actual practical terms, except where they impose a strong inappropriate restrictions on the model – such as using a probability as the conditional mean in a count or regression model – different link functions will produce similar empirical results.

As many of the models have already been documented in earlier chapters. All of the models may be requested with the command

GLIM ; Lhs = ... ; Rhs = ... ; Model = the model name given above $

The models may all be invoked with the **GLIM** command. For the models documented in this chapter, the command **GLIM** is also synonymous with **LOGLINEAR**. So, for example, the following two commands are identical:

GLIM ; Lhs = y ; Rhs = x ; Model = Beta $
LOGLINEAR ; Lhs = y ; Rhs = x ; Model = Beta $

Tables E50.1 and E50.2 list the loglinear models for continuous data described in this chapter.

Model	Dependent Variable	Conditional Mean	Command
Beta	Bounded in (0,1)	$\exp(t_a)/[\exp(t_a)+\exp(t_b)]$	
Power	Bounded in (0,1)	$[\exp(t)+1]/[\exp(t)+2]$	
Normal (Loglinear)	Continuous	$\exp(t)$	
Gamma	Nonnegative	$P\exp(-t)$	
Weibull	Nonnegative	$[\exp(-t)]^{1/P}\Gamma[(P+1)/P]$	
Exponential[b]	Nonnegative	$\exp(-t)$	
Rayleigh	Nonnegative	$[\pi\exp(-t)/2)]^{1/2}$	
Inverse Gaussian[b]	Nonnegative	$P\exp(-t)$	

[a] In all models, $t = \boldsymbol{\beta}'\mathbf{x}$.
[b] Exponential families typically included in analysis of 'Generalized Linear Models'

Table E50.1 Generalized Linear Models[a]

Model	Conditional Mean	Conditional Variance
Beta	$\lambda_a/(\lambda_a + \lambda_b)$	$\lambda_a\lambda_b/[(\lambda_a+\lambda_b+1)(\lambda_a+\lambda_b)^2]$
Power	$(\lambda+1)/(\lambda+2)$	$(\lambda+1)/(\lambda+3) - [(\lambda+1)/(\lambda+2)]^2$
Normal (Loglinear)	λ	σ^2
Gamma	P/λ	P/λ^2
Weibull	$(1/\lambda)^{1/P}\Gamma[(P+1)/P)]$	$(1/\lambda)^{2/P}\{\Gamma[(P+2)/P)] - \Gamma^2[(P+1)/P]\}$
Exponential	$1/\lambda$	$1/\lambda^2$
Rayleigh	$[\pi/(2\lambda)]^{1/2}$	$(4 - \pi)/(2\lambda)$
Inverse Gaussian	P/λ	P/λ^3

[a] In all results given, $t = \boldsymbol{\beta}'\mathbf{x}$ and $\lambda = \exp(t) = \exp(\boldsymbol{\beta}'\mathbf{x})$.
[b] Mean and variance for the gamma count model are approximate based on increasing event window. See Winkelmann (2003, p. 55).

Table E50.2 Means and Variances of Variables in Generalized Linear Models[a]

E50.2.1 Standard Model Specifications for the Loglinear Regression Models

This is the full list of general specifications that are applicable to this group of model estimators.

Controlling Output from Model Commands

; Par keeps ancillary parameter *p* in main results vector *b*.
; Margin displays marginal effects.
; OLS displays least squares starting values when (and if) they are computed.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec requests computation of the cluster form of corrected covariance estimator. (**; Stratum = specification** for stratified clustered data).
; Robust requests a ‘sandwich’ estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E50.2.2 Estimated Results for Loglinear Models

The following are the estimated results for the models in Table E50.1:

Matrices:	*b*	= estimates of $\boldsymbol{\beta}$, or
	b	= estimates of $\boldsymbol{\alpha}$ followed by $\boldsymbol{\beta}$ for the beta model
	varb	= estimated asymptotic covariance matrix for MLE of *b*
Scalars:	*logl*	= log likelihood
	kreg	= number of variables in Rhs
	nreg	= number of observations
	pgamma	= *P* for the gamma model
	pweibull	= *P* for the Weibull model
	pinverse	= *P* for the inverse Gaussian model
	s	= σ for the normal (exponential regression) model, or
	s	= σ for the lognormal regression model
	exitcode	

Last Model: For the **WALD** command using *b_name* for the parts of *b*

Last Function: Conditional mean function

E50.3 Variables with Unrestricted Range

The simplest form of generalized linear model is the linear regression,

$$\begin{aligned} y_i &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \\ &= \log(\lambda_i) + \varepsilon_i,\ \varepsilon_i \sim N[0,\sigma^2]. \end{aligned}$$

The linear regression model is discussed in detail in Chapters E7 and E8 and in a variety of specifications. One extension might be to use distributions other than the normal for the distribution family. While this does not represent an extension of the model, there are several ways one might proceed. First, least squares in this model is robust to distributional assumptions, so in some sense, the point of the distribution is moot. The Gauss Markov assumptions are met, so without a specific distributional alternative, least squares with a robust covariance matrix would be the estimator of choice. Alternatively, one might want to use a semiparametric method, such as least absolute deviations. Finally, one might be interested in using a specific alternative distribution. The **MAXIMIZE** command can be used to construct the particular maximum likelihood estimator. Alternatively, by treating y_i as if it were the log of a survival time, one can use one of the parametric survival models described in Chapter E60. A variety of panel data treatments for the linear model are presented in Chapter E51.

An alternative form of the normal regression model that remains in the loglinear class of models considered here is

$$\begin{aligned} y_i &= \exp(\boldsymbol{\beta}'\mathbf{x}_i) + \varepsilon_i \\ &= \lambda_i + \varepsilon_i,\ \varepsilon_i \sim N[0,\sigma^2]. \end{aligned}$$

This is a nonlinear regression model that can be fit by nonlinear least squares, using **NLSQ**. The commands could be

NAMELIST ; x = ... the list of variables $
CALC ; k = Col(x) $
NLSQ ; Lhs = y ; Fcn = Exp(b1'x) ; Labels = k_b ; Start = k_0 $

This nonlinear regression may also be fit as a loglinear model with the command

LOGLINEAR ; Lhs = dependent variable
; Rhs = independent variables
; Model = Normal $

All other options described in this chapter for the loglinear models may be used as well. This model is fit by maximum likelihood, which for the normal distribution is nonlinear least squares. If you use **NLSQ**, the estimator will use the Gauss-Marquardt method. The BFGS algorithm is used here, instead.

Of course, you can use a different functional form in **NLSQ**, and the exponential does not have any particularly attractive features. Moreover, perhaps less noted than it might be, one tends to get similar answers for the different functional forms when marginal effects are compared. For the linear model, $\boldsymbol{\beta}$ gives the partial effects. In the loglinear form, $\lambda_i\boldsymbol{\beta}$ gives the marginal effects. The example below compares these for a model of household income using the German health care data.

NAMELIST ; x = one,age,educ,female,married,hhkids $
REGRESS ; Lhs = hhninc ; Rhs = x $
LOGLINEAR ; Lhs = hhninc ; Rhs = x ; Partial Effects ; Model = Normal $

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                 =         .34930
             Standard deviation   =         .16296
             No. of observations  =           6209  Degrees of freedom
Regression   Sum of Squares       =        19.3836           5
Residual     Sum of Squares       =        145.477        6203
Total        Sum of Squares       =        164.860        6208
             Standard error of e  =         .15314 <--
Fit          R-squared            =         .11758  R-bar squared =   .11686
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.04123**      .01698    -2.43  .0152     -.07451    -.00796
     AGE|     .00237***     .00025     9.67  .0000      .00189     .00285
    EDUC|     .02090***     .00096    21.73  .0000      .01902     .02279
  FEMALE|    -.00209        .00412     -.51  .6109     -.01016     .00597
 MARRIED|     .07871***     .00565    13.93  .0000      .06764     .08978
  HHKIDS|    -.01974***     .00456    -4.33  .0000     -.02868    -.01080
--------+--------------------------------------------------------------------
Normal Regression with Exponential Mean
Log likelihood function      2860.43906
R squared = 1-Var(e)/Var(y) =  .1224360
--------+--------------------------------------------------------------------
```

HHINC	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters in conditional mean function					
Constant	-2.18838***	.04789	-45.70	.0000	-2.28224	-2.09451
AGE	.00740***	.00071	10.48	.0000	.00602	.00879
EDUC	.05486***	.00218	25.17	.0000	.05059	.05914
FEMALE	-.00332	.01175	-.28	.7774	-.02635	.01971
MARRIED	.25403***	.01892	13.42	.0000	.21694	.29112
HHKIDS	-.04071***	.01327	-3.07	.0022	-.06671	-.01471
	Standard deviation of normally distributed effect					
Sigma	.15265***	.00137	111.44	.0000	.14996	.15533

Partial derivatives of expected val. with
respect to the vector of characteristics.

HHINC	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
AGE	.00255***	.00024	10.54	.0000	.00208	.00303
EDUC	.01891***	.00074	25.73	.0000	.01747	.02035
FEMALE	-.00115	.00405	-.28	.7774	-.00908	.00679
MARRIED	.08756***	.00645	13.57	.0000	.07491	.10022
HHKIDS	-.01403***	.00457	-3.07	.0021	-.02299	-.00507

E50.4 Nonnegative Random Variables

This section presents estimators for five models for nonnegative variables, the exponential, gamma, Weibull, Rayleigh and inverse Gaussian. A sixth, the lognormal regression, is considered in the next section. In all of these loglinear models, we parameterize the regression using $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$.

E50.4.1 Exponential Regression Model

The exponential model is a single index model with density

$$f(y_i) \;=\; \lambda_i \exp(-\lambda_i y_i)\,,\; y_i \geq 0,$$

$$\lambda_i \;=\; \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

The regression function has

$$E[y_i \mid \mathbf{x}_i] \;=\; 1/\lambda_i \text{ and } \mathrm{Var}[y_i \mid \mathbf{x}_i] \;=\; 1/\lambda_i^2$$

so the slopes of conditional mean are

$$\boldsymbol{\delta}_i \;=\; -\partial \mathrm{E}[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i \;=\; -\lambda_i\,\beta.$$

(Note that the slopes have the opposite signs from the coefficients.) The exponential model forms the most basic (restrictive) model in the group considered here.

E50.4.2 Gamma Regression Model

The gamma and Weibull models both extend the exponential model by allowing a shape parameter to change the model form. The gamma model is

$$f(y_i) = \frac{\lambda_i^P}{\Gamma(P)}\exp(-\lambda_i y_i) y_i^{P-1},\ y_i \geq 0,$$

with conditional mean and variance functions

$$\mathrm{E}[y_i \mid \mathbf{x}_i] = \mathrm{P}/\lambda_i \text{ and } \mathrm{Var}[y_i \mid \mathbf{x}_i] = \mathrm{P}/\lambda_i^2.$$

The vector of slopes is

$$\boldsymbol{\delta}_i = -(\mathrm{P}/\lambda_i)\,\boldsymbol{\beta} = -\,\mathrm{E}[y_i \mid \mathbf{x}_i]\boldsymbol{\beta}.$$

The exponential model results from the gamma model if $P = 1$.

E50.4.3 Weibull Regression Model

The Weibull model is similar to the gamma. The density is

$$f(y_i) = P\lambda_i y_i^{P-1}\exp(-\lambda_i y_i^P),\ y_i \geq 0.$$

The conditional mean is found by integrating this form of the gamma function to obtain

$$\mathrm{E}[y_i \mid \mathbf{x}_i] = \left(\frac{1}{\lambda_i}\right)\Gamma\left(\frac{P+1}{P}\right)$$

and variance

$$\mathrm{Var}[y_i \mid \mathbf{x}_i] = \left(\frac{1}{\lambda_i}\right)^2\left[\Gamma\left(\frac{P+2}{P}\right)-\Gamma^2\left(\frac{P+1}{P}\right)\right]$$

which produces slope vector

$$\boldsymbol{\delta}_i = -\mathrm{E}[y_i \mid \mathbf{x}_i]\boldsymbol{\beta}.$$

Once again, the exponential model is the special case with $P = 1$.

E50.4.4 Rayleigh Regression Model

The Rayleigh distribution applies to a variable that is the square root of twice an exponential variable (i.e., a simple transformation). The density (parameterized to be consistent with our general formulation) is

$$f(y_i) = \lambda_i y_i \exp\left(-\tfrac{1}{2}\lambda_i y_i^2\right), y_i > 0,\ \lambda_i > 0.$$

Precisely, in this form, the variable $z_i = y_i^2/2$ has an exponential density with parameter λ_i. (That is how the model is estimated – we simply transform your Lhs variable during the iterations and use the simpler exponential density to form the likelihood function to estimate the parameters.) The conditional mean and variance functions are

$$E[y_i \mid \mathbf{x}_i] = \sqrt{\frac{\pi}{2\lambda_i}} \text{ and } \mathrm{Var}[y_i \mid \mathbf{x}_i] = \frac{4-\pi}{2\lambda_i}$$

The vector of partial effects is

$$\boldsymbol{\delta}_i = \left[-\left(\frac{1}{2}\sqrt{\frac{\pi}{2}}\right)\frac{1}{\sqrt{\lambda_i}}\right]\boldsymbol{\beta} = -(1/2)\mathrm{E}[y_i \mid \mathbf{x}_i]\boldsymbol{\beta}.$$

The Rayleigh form is a substantial extension of the model, since it has the shapes of the gamma function depending on the value of λ_i.

E50.4.5 Inverse Gaussian Regression Model

The inverse Gaussian model appears in survival modeling and other types of reliability analysis. The standard functional form, where for the present, we omit the covariates is

$$f(y_i) = \left[\frac{\alpha}{2\pi y_i^3}\right]^{1/2} \exp\left[\frac{-\alpha(y_i-\mu)^2}{2\mu^2 y_i}\right], y_i > 0,\ \mu > 0,\ \lambda > 0.$$

In this formulation, $\mathrm{E}[y_i] = \mu$. In order to estimate the parameters of the model, we reparameterize it using

$$P = \sqrt{\alpha}$$

and

$$\lambda = \frac{\sqrt{\alpha}}{\mu}$$

(This is a slight departure from standard references which often replace α with λ in the original form. This is done here to maintain consistency with the other models presented in this section.) The density is now

$$f(y_i) \quad = \quad \frac{P}{\sqrt{2\pi y_i^3}} \exp\left[-\frac{1}{2}\frac{(\lambda_i y_i - P)^2}{y_i}\right], \; y_i > 0, P > 0, \lambda > 0.$$

and the mean is **P**/λ. Finally, to introduce the individual heterogeneity in the parameters, we write

$$\lambda_i \quad = \quad \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

which produces the conditional density just by substitution, while

$$E[y_i|\mathbf{x}_i] = P/\lambda_i. \text{ and } Var[y_i|\mathbf{x}_i] = P/\lambda_i^3$$

The slopes of the conditional mean are

$$\boldsymbol{\delta}_i \quad = \quad (-P/\lambda_i)\boldsymbol{\beta} = -E[y_i|\mathbf{x}_i]\,\boldsymbol{\beta}.$$

E50.4.6 Comparison of Loglinear Models

The first four functional forms differ partly through the shape parameter, P (which equals 1 for the exponential model, which is this special case of the other functions.) The precise shapes of the gamma and Weibull depends on P and whether P is larger than or smaller than 1. The Rayleigh, however, is strictly a function of the exponential, as it does not have a separate shape parameter. The figure below shows the three densities for $\lambda = 1$ and $P = 1.5$.

```
SAMPLE    ; 1-401 $
CREATE    ; y = Trn(0,.01) $
CALC      ; al = 1 ; p = 1.5 $
CREATE    ; exponent = al*Exp(-al*y) $
CREATE    ; gamma = (al^p)/Gma(p)*Exp(-al*y)*(y^(p-1)) $
CREATE    ; weibull = p*al*y^(p-1)*Exp(-al*(y^p)) $
CREATE    ; Rayleigh = al*y*Exp(-.5*al*y^2) $
PLOT      ; Lhs = y
          ; Rhs = exponent,gamma,weibull,rayleigh
          ; Title = Densities for Loglinear Models
          ; Vaxis = Density ; Fill ; Grid $
```

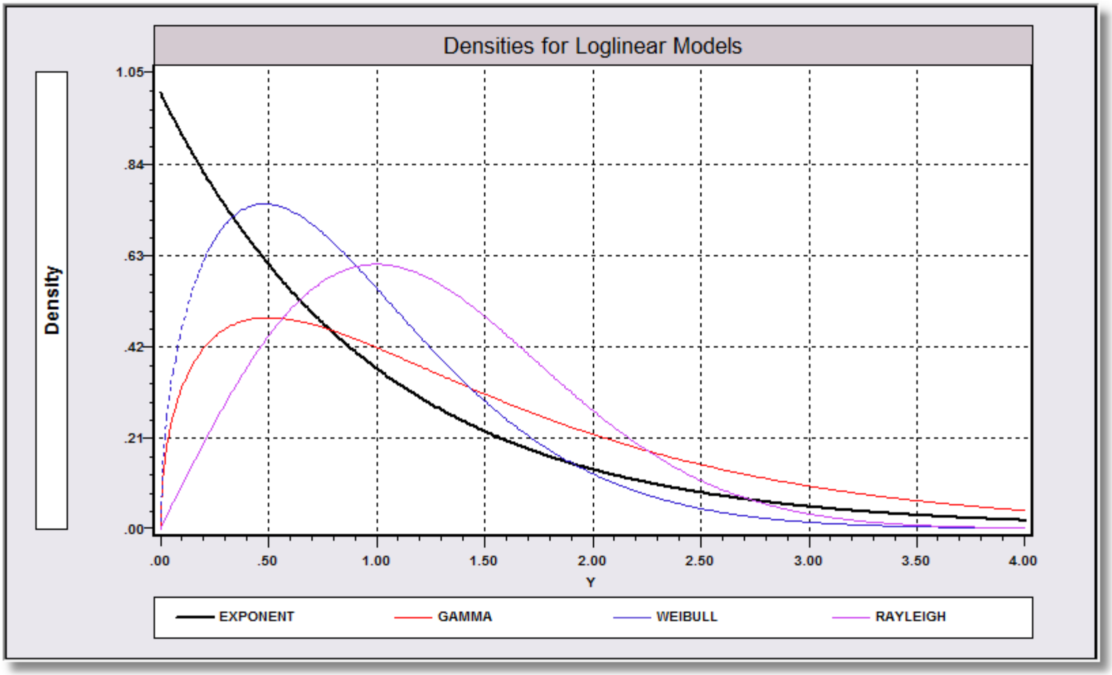


Figure E50.1 Densities for Loglinear Models

E50.4.7 Commands

The estimator is maximum likelihood in all cases. To request these models, use

```
LOGLINEAR ; Lhs = dependent variable (must be nonnegative)
          ; Rhs = one list of independent variables
          ; Model = Rayleigh, Exponential, Gamma, Weibull
                    or Inverse Gaussian $
```

E50.4.8 Applications

To illustrate the different estimators, we have used two of the five distributions to fit a loglinear model to the distribution of incomes. The partial effects at the means and averaged over the sample observations are shown with each model.

```
SAMPLE     ; All $
CREATE     ; agesq = age*age $
SETPANEL   ; Group = id ; Pds = ti $
REJECT     ; ti  <  7 | hhninc = 0 $
NAMELIST   ; x = one,age,age^2,educ,female,married $
LOGLINEAR  ; Lhs = hhninc ; Rhs = x ; Model = Exponential ; Partial Effects $
PARTIALS   ; Effects: age / educ / female / married ; Summary $
LOGLINEAR  ; Lhs = hhninc ; Rhs = x ; Model = Weibull ; Partial Effects $
PARTIALS   ; Effects: age / educ / female / married ; Summary $
```

```
--------------------------------------------------------------------------------
Exponential (Loglinear) Regression Model
Dependent variable                HHNINC
Log likelihood function        398.32536
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    3.14298***      .28090    11.19  .0000     2.59242   3.69354
     AGE|    -.05677***      .01299    -4.37  .0000     -.08223   -.03130
 AGE^2.0|     .00056***      .00015     3.80  .0001      .00027    .00084
    EDUC|    -.05084***      .00632    -8.04  .0000     -.06323   -.03845
  FEMALE|     .02155         .02685      .80  .4222     -.03108    .07418
 MARRIED|    -.18470***      .03593    -5.14  .0000     -.25512   -.11427
--------+-----------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
--------+-----------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  HHNINC|      Effect        Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AGE|     .16454***      .03772     4.36  .0000      .09062    .23846
 AGE^2.0|    -.00161***      .00042    -3.80  .0001     -.00245   -.00078
    EDUC|     .14735***      .01842     8.00  .0000      .11125    .18345
  FEMALE|    -.06246         .07783     -.80  .4222     -.21502    .09009
 MARRIED|     .53533***      .10437     5.13  .0000      .33076    .73989
--------+-----------------------------------------------------------------------
Partial Effects for Exponential Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect       Error      |t|  95% Confidence Interval
----------------------------------------------------------------------
      AGE            .02358      .00460    5.13      .01457       .03258
      EDUC           .14905      .01880    7.93      .11221       .18590
   *  FEMALE       -.06328      .07897     .80     -.21806       .09150
   *  MARRIED        .57631      .11970    4.81      .34171       .81090
----------------------------------------------------------------------

--------------------------------------------------------------------------------
Weibull (Loglinear) Regression Model
Dependent variable                HHNINC
Log likelihood function       3121.57173
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    2.84688***      .08879    32.06  .0000     2.67286   3.02090
     AGE|    -.06208***      .00427   -14.54  .0000     -.07045   -.05371
 AGE^2.0|     .00061***   .4871D-04    12.60  .0000      .00052    .00071
    EDUC|    -.03813***      .00183   -20.85  .0000     -.04171   -.03454
  FEMALE|     .03741***      .00794     4.71  .0000      .02185    .05296
 MARRIED|    -.00793         .00989     -.80  .4229     -.02731    .01146
        |Scale parameter for Weibull model
 P_scale|    2.26749***      .01360   166.76  .0000     2.24084   2.29414
```

```
--------+-------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
--------+-------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  HHNINC|      Effect       Error        z    |z|>Z*          Interval
--------+-------------------------------------------------------------
     AGE|     .01599***      .00110    14.48  .0000      .01382    .01815
 AGE^2.0|    -.00016***   .1260D-04   -12.55  .0000     -.00018   -.00013
    EDUC|     .00982***      .00046    21.15  .0000      .00891    .01073
  FEMALE|    -.00963***      .00205    -4.71  .0000     -.01365   -.00562
 MARRIED|     .00204         .00255      .80  .4229     -.00295    .00703
--------+-------------------------------------------------------------
Partial Effects for Weibull Loglinear Regression Model
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .00243     .00017   14.45      .00210       .00276
      EDUC           .01323     .00062   21.28      .01201       .01444
   *  FEMALE        -.01293     .00275    4.71     -.01832      -.00755
   *  MARRIED        .00274     .00341     .80     -.00395       .00943
---------------------------------------------------------------------
```

E50.5 Technical Details for the Loglinear Models

All five models are parameterized in terms of

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i).$$

The five densities, gradients of the log densities, and Hessians of the log densities are as follows:

E50.5.1 Exponential

For the exponential model,

$$f(y_i) = \lambda_i \exp(-\lambda_i y_i),\ \ y_i \geq 0,$$

$$\partial \log f(y_i)/\partial\lambda_i = (1 - y_i\lambda_i)\,/\,\lambda_i$$

$$\partial \log f(y_i)/\partial\boldsymbol{\beta} = (1 - y_i\lambda_i)\mathbf{x}_i$$

$$\partial^2 \log f(y_i)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -(y_i\,\lambda_i)\mathbf{x}_i\mathbf{x}_i'$$

These terms then define the log likelihood function. The actual Hessian is used for the asymptotic covariance matrix. In this model, the conditional mean function is just $1/\lambda_i$, so the partial effects are

$$\partial \mathrm{E}[y_i|\mathbf{x}_i] = -\lambda_i\boldsymbol{\beta} = -\mathrm{E}[y|\mathbf{x}]\boldsymbol{\beta} = \boldsymbol{\delta}_i$$

which is computed at the means of the data. Standard errors are computed using the delta method. (Note the sign reversal in the marginal effects.) The derivatives matrix for this computation is

$$\partial\boldsymbol{\delta}_i/\partial\boldsymbol{\beta}' \;=\; \mathbf{G}_i \;=\; -\lambda_i[\mathbf{I} \;+\; \boldsymbol{\beta}\mathbf{x}_i']$$

once again, computed at the means of the data.

E50.5.2 Gamma

For the gamma model,

$$f(y_i) \;=\; \frac{\lambda_i^P}{\Gamma(P)}\exp(-\lambda_i y_i)y_i^{P-1},\;\; y_i \geq 0,$$

$$\log f(y_i) \;=\; P\log\lambda_i \;-\; \log\Gamma(P) \;-\; \lambda_i y_i \;+\; (P\text{-}1)\log y_i$$

$$\partial\log f(y_i)/\partial\lambda_i \;=\; P/\lambda_i \;-\; \mathrm{y}_i$$

$$\partial\log f(y_i)/\partial\boldsymbol{\beta} \;=\; (P - \lambda_i y_i)\mathbf{x}_i$$

$$\partial\log f(y_i)/\partial P \;=\; \log\lambda_i \;-\; \Psi(P) \;+\; \log y_i$$

The terms in the Hessian are

$$\partial^2\log f(y_i)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' \;=\; -(\mathrm{y}_i\,\lambda_\mathrm{i})\mathbf{x}_i\mathbf{x}_i'$$

$$\partial^2\log f(y_i)/\partial P^2 \;=\; -\,\Psi'(P) \;+\; \log y_i$$

$$\partial^2\log f(y_i)/\partial\boldsymbol{\beta}\partial P \;=\; \mathbf{x}_i$$

The conditional mean function in the gamma model is

$$\mathrm{E}[y_i|\mathbf{x}_i] \;=\; P\,/\,\lambda_i$$

so, the partial effects are

$$\boldsymbol{\delta}_i \;=\; -P/\lambda_i\,\boldsymbol{\beta} \;=\; -\mathrm{E}[y|\mathbf{x}]\boldsymbol{\beta}$$

For computing standard errors of the partial effects,

$$\partial\boldsymbol{\delta}_i/\partial\boldsymbol{\beta}' \;=\; \mathbf{G}_{i\beta} \;=\; -P/\lambda_i[\mathbf{I} \;-\; \boldsymbol{\beta}\mathbf{x}_i']$$

$$\partial\boldsymbol{\delta}_i/\partial P \;=\; \mathbf{G}_{iP} \;=\; -(1/\lambda_i)\boldsymbol{\beta}$$

E50.5.3 Weibull

For the Weibull model,

$$f(y_i) = P\lambda_i y_i^{P-1}\exp(-\lambda_i y_i^P),\ y_i \geq 0,$$

$$\log f(y_i) = \log P + \log\lambda_i + (P\text{-}1)\log y_i - \lambda_i y_i^P$$

$$\partial\log f(y_i)/\partial\lambda_i = 1/\lambda_i - y_i^P$$

$$\partial \log f(y_i)/\partial\boldsymbol{\beta} = (1 - \lambda_i y_i^P)\mathbf{x}_i$$

$$\partial^2\log f(y_i)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -(y_i\ \lambda_i^P)\mathbf{x}_i\mathbf{x}_i'$$

$$\partial \log f(y_i)/\partial P = 1/P + \log y_i - \lambda_i y_i^P \log y_i$$

$$\partial^2\log f(y_i)/\partial P^2 = -1/P^2 - \lambda_i y_i^P(\log y_i)^2$$

$$\partial^2\log f(y_i)/\partial\boldsymbol{\beta}\partial P = - \lambda_i y_i^P(\log y_i)\,\mathbf{x}_i$$

The conditional mean function in the Weibull model is

$$\text{E}[y_i \mid \mathbf{x}_i] = \left(\frac{1}{\lambda_i}\right)\Gamma\left(\frac{P+1}{P}\right).$$

The partial effects are

$$\boldsymbol{\delta}_i = -\left(\frac{1}{\lambda_i}\right)\boldsymbol{\Gamma}\left(\frac{P+1}{P}\right)\boldsymbol{\beta} = -\text{E}[y_i|\mathbf{x}_i]\ \boldsymbol{\beta}$$

For computing marginal effects,

$$\partial\boldsymbol{\delta}_i/\partial\boldsymbol{\beta}' = \mathbf{G}_{i\beta} = -\text{E}[y_i|\mathbf{x}_i]\ [\mathbf{I} - (1/P)\boldsymbol{\beta}\mathbf{x}_i']$$

$$\partial\boldsymbol{\delta}_i/\partial P = \mathbf{G}_{iP} = -\boldsymbol{\delta}_i\ [\Psi((\text{P}+1)/\text{P}) \times 1/\text{P}^2]$$

(We make use of the fact that $\Gamma'(t) = \Gamma(t)\Psi(t)$.)

E50.5.4 Rayleigh Distribution

The actual density for the Rayleigh distribution is

$$f(y_i) = \lambda_i y_i \exp\left(-\tfrac{1}{2}\lambda_i y_i^2\right), y_i > 0,\ \lambda_i > 0.$$

However, rather than manipulate this distribution for estimation purposes, internally, we create $z_i = y_i^2/2$, which has an exponential distribution based on the same parameter λ_i. Thus, we use the exponential model to estimate the parameters.

The marginal effects are

$$\boldsymbol{\delta}_{\mathrm{i}} = \left(\frac{-1}{2}\sqrt{\frac{\pi}{2\lambda_i}}\right)\boldsymbol{\beta}\,.$$

The derivative matrix for computing the asymptotic covariance for the marginal effects is

$$\mathbf{G}_i = \left(\frac{-1}{2}\sqrt{\frac{\pi}{2\lambda_i}}\right)[\mathrm{I} - (1/2)\boldsymbol{\beta}\mathbf{x}'].$$

E50.5.5 Inverse Gaussian

For the inverse Gaussian model,

$$f(y_i) = \frac{P}{\sqrt{2\pi y_i^{\,3}}}\exp\left[-\frac{1}{2}\frac{(\lambda_i y_i - P)^2}{y_i}\right],\ y_i > 0,\ P > 0,\ \lambda > 0.$$

$$\log f(y_i) = \log P \ - ½ \log(2\pi\, y_i^{3}) - ½\, (\lambda_i y_i - P)^2/y_i$$

$$\partial \log f(y_i)/\partial\lambda_i = -e_i \ \text{ where } e_i = \lambda_i y_i - P$$

$$\partial \log f(y_i)/\partial\boldsymbol{\beta} = -e_i\lambda_i\mathbf{x}_i$$

$$\partial^2 \log f(y_i)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\,\lambda_i(e_i + \lambda_i y_i)\mathbf{x}i\mathbf{x}i'$$

$$\partial \log f(y_i)/\partial P = 1/P \ + \ e_i/y_i$$

$$\partial^2 \log f(y_i)/\partial P^2 = -1/P^2 - 1/y_i$$

$$\partial^2 \log f(y_i)/\partial\boldsymbol{\beta}\partial P = \lambda_i\mathbf{x}_i$$

The mean in the inverse Gaussian model in the original form is μ. Therefore, as reparameterized, the conditional mean in this model is

$$\mathrm{E}[y_i|\mathbf{x}_i] = P/\lambda_i$$

(note the similarity to the other models, which is what motivated this reparameterization). The partial effects are

$$\boldsymbol{\delta}_i = -P/\lambda_i\, \boldsymbol{\beta}$$

(note, again, the sign reversal). For computing standard errors of the marginal effects,

$$\partial\boldsymbol{\delta}_i/\partial\boldsymbol{\beta}' = \mathbf{G}_{i\beta} = -P/\lambda_i[\mathbf{I} \ - \ \boldsymbol{\beta}\mathbf{x}_i']$$

$$\partial\boldsymbol{\delta}_i/\partial P = \mathbf{G}_{iP} = -(1/\lambda_i)\boldsymbol{\beta}.$$

E50.6 The Lognormal Regression Model

The lognormal regression model is specified to include a particular type of heteroscedasticity as well as to deal explicitly with the nonnegative values of certain variables. (See Amemiya (1973).) If y has a lognormal distribution, then its variance is proportional to the square of its mean. The general form of the underlying regression is

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

where y is positive, $$E[y] = \boldsymbol{\beta}'\mathbf{x}$$

and, $$\text{Var}[y] = \sigma^2[\boldsymbol{\beta}'\mathbf{x}]^2.$$

In this model, $$E[\log y] = \log(\boldsymbol{\beta}'\mathbf{x}) - 2\sigma^2$$

and $$\text{Var}[\log y] = \sigma^2.$$

The lognormal regression also allows censoring. (In the literature on this model, this variant is erroneously called truncation.) In this case, censoring may only be on the right. This model has been applied to the length of program participation, in which y must be positive and does not exceed the length of the program. Another natural application is the distribution of incomes, as in the application below.

The command is

LOGNORMAL ; Lhs = dependent variable
(or GLIM) ; Rhs = regressors $

The censored form of the model can be specified by adding the specification

; Limit = limit value

where limit is a fixed value or a variable. The limit must always be positive, as it is an upper limit.

The model parameters are ($\boldsymbol{\beta}$,σ^2). Estimation parameters are $\boldsymbol{\beta}$ and $\theta^2 = \log(1 + \sigma^2)$. Finally, for the lognormal regression model, the predicted value is just the mean, $\boldsymbol{\beta}'\mathbf{x}$. The other values displayed by **; List** are the residual, $\boldsymbol{\beta}'\mathbf{x}$ again, and the probability that y_i would exceed the limit value. The latter is zero if the data are not censored. This model, save for these considerations, is the same as the tobit model discussed in Chapter E45. All other options and specifications are identical. Note, however, that since the conditional mean is linear, **; Partial Effects** does not produce additional results.

There are no panel data forms of the lognormal model. For modeling in this context with panel data, the truncated regression or the loglinear models, Weibull, gamma, inverse Gaussian or Rayleigh should provide satisfactory alternatives. (The exponential model is likely to be too restrictive.)

E50.6.1 Application

In the results below, we fit a lognormal distribution to the income variable *hhninc*, in the German health care data analyzed earlier. The first model illustrates the truncated lognormal estimator. The second and third estimates compare the uncensored lognormal distribution to an ordinary truncated regression model with the same data. These two models are roughly comparable. The commands are:

SAMPLE ; All $
REJECT ; _groupti < 7 $
NAMELIST ; x = one,educ,hhkids,female,married,age $
LOGNORMAL ; Lhs = hhnins ; Rhs = x ; Limit = 2 $
LOGNORMAL ; Lhs = hhninc ; Rhs = x $
TRUNCATION ; Lhs = hhninc ; Rhs = x $

Note that in comparing the models, the parameter ‘σ’ is completely different; in the lognormal model it is a scale factor in the scedastic function while in the truncation model, it carries the scale of the dependent variable. Also, for comparing the estimated effects, the comparison would be between the coefficients in the lognormal model and the marginal effects in the truncation model, which, as can be seen below, are fairly similar. The differences arise because of the intrinsic differences in the functional forms.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - LOGNORMA
Dependent variable                HHNINC
Log likelihood function     -3336.11491
Estimation based on N =   6208, K =   7
Lower =      .0000    Upper =     2.0000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    -.01587         .01561    -1.02  .3093     -.04646    .01472
    EDUC|     .02123***      .00105    20.20  .0000      .01917    .02329
  HHKIDS|    -.03004***      .00375    -8.01  .0000     -.03738   -.02269
  FEMALE|    -.00505         .00354    -1.42  .1542     -.01199    .00190
 MARRIED|     .09845***      .00360    27.32  .0000      .09139    .10551
     AGE|     .00148***      .00019     7.62  .0000      .00110    .00186
        |Variance for lognormal distribution
   Sigma|     .43267***      .00409   105.75  .0000      .42465    .44069
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - LOGNORMA
Dependent variable               HHNINC
Log likelihood function     -3382.89616
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    -.00895          .01473     -.61  .5436     -.03783     .01993
    EDUC|     .02087***       .00102    20.56  .0000      .01888     .02286
  HHKIDS|    -.02972***       .00372    -7.98  .0000     -.03702    -.02242
  FEMALE|    -.00520          .00354    -1.47  .1417     -.01214     .00174
 MARRIED|     .09688***       .00363    26.68  .0000      .08976     .10399
     AGE|     .00146***       .00020     7.46  .0000      .00107     .00184
        |Variance for lognormal distribution
   Sigma|     .43611***       .00394   110.74  .0000      .42840     .44383
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Limited Dependent Variable Model - TRUNCATE
Dependent variable               HHNINC
Log likelihood function      2965.37636
Lower =       .0000    Upper = +infinity
Observations after truncation      6208
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|    -.09010***       .01931    -4.67  .0000     -.12795    -.05225
    EDUC|     .02260***       .00105    21.45  .0000      .02054     .02467
  HHKIDS|    -.02198***       .00513    -4.28  .0000     -.03204    -.01192
  FEMALE|    -.00262          .00462     -.57  .5701     -.01167     .00643
 MARRIED|     .09149***       .00662    13.83  .0000      .07852     .10446
     AGE|     .00263***       .00028     9.50  .0000      .00209     .00318
        |Disturbance standard deviation
   Sigma|     .16204***       .00174    93.12  .0000      .15863     .16545
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E50.6.2 Technical Details for the Lognormal Regression Model

The log likelihood function is

$$\log L = -\tfrac{1}{2}\, \Sigma_{nonlimit\ observations} \{\log\theta^2 + \log 2\pi + (1/\theta^2)[\log y_i - \log(\boldsymbol{\beta}'\mathbf{x}_i) + \theta^2/2]^2\}$$

$$+ \Sigma_{limit\ observations} \log\Phi[-(1/\theta)(\log(\boldsymbol{\beta}'\mathbf{x}_i) - \log U_i - \theta^2/2)],$$

where $\theta^2 = \log(1 + \sigma^2)$

and U_i = upper censoring point or $+\infty$, in which case, there are no limit values.

The function and derivatives are the same as those for the tobit model with upper censoring at $\log U_i$, where we would use the analogy

$$\log y_i = \log(\boldsymbol{\beta}'\mathbf{x}_i) - \theta^2/2 + \varepsilon_i.$$

Let $$\varepsilon_i = \log y_i - \log(\boldsymbol{\beta}'\mathbf{x}_i) + \theta^2/2.$$

Thus, $$\partial\log L/\partial\boldsymbol{\beta} = (1/\theta^2)\Sigma_{nonlimit}[\varepsilon_i/(\boldsymbol{\beta}'\mathbf{x}_i)]\mathbf{x}_i - \Sigma_{limit}[(\phi/\Phi)\, 1/(\theta\boldsymbol{\beta}'\mathbf{x}_i)]\mathbf{x}_i$$

and $$\partial\log L/\partial\theta^2 = 1/(2\theta^2)\Sigma_{nonlimit}[(\varepsilon_i/\theta)^2 - \varepsilon_i - 1] + \Sigma_{limit}[(\phi/\Phi)/(2\theta)](1 + \varepsilon_i^2/\theta^2).$$

The BHHH estimator is used to estimate the asymptotic covariance matrix of the coefficient estimates.

An aspect of the model should be noted. The model implies the log of a linear function enters the log likelihood. Since a linear function cannot be directly constrained, there is the possibility that the function can become noncomputable. If the data and the model are well matched, this should be unusual. Users are warned of this possibility, however. The program cannot restrict the estimates in any way to prevent this. A value of $\boldsymbol{\beta}'\mathbf{x}$ that is nonpositive is fixed at a small positive value so that estimation can continue. However, if the problem occurs many times, the estimation is likely to break down at some point, claiming to be unable to maximize the function.

E50.7 Variable Limited to the (0,1) Interval

These models are defined for a random variable with range of variation restricted to an interval (L,U) which is usually (0,1) but may be any fixed interval. To use this model, you will need to rescale your variable if (0,1) is not the range of variation by dividing it by U-L by using a **CREATE** command. As such, we assume from this point on that y ranges in the unit interval. The distributions for y are the beta distribution and the power distribution.

The beta distribution is defined by two parameters, a and b, such that

$$f(y|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}, 0 \le y \le 1.$$

We parameterize this distribution by assuming

$$a = \exp(\boldsymbol{\alpha}'\mathbf{x}) = \lambda_a$$

and $$b = \exp(\boldsymbol{\beta}'\mathbf{x}) = \lambda_b.$$

This defines the conditional density, $f(y_i|\mathbf{x}_i,\boldsymbol{\alpha},\boldsymbol{\beta})$ as well as the log likelihood function needed for parameter estimation. The mean and variance of the beta distributed random variable are

$$E[y|\mathbf{x}] = \frac{\lambda_a}{\lambda_a + \lambda_b} \text{ and } \mathrm{Var}[y|\mathbf{x}] = \frac{\lambda_a\lambda_b}{(\lambda_a + \lambda_b + 1)(\lambda_a + \lambda_b)^2}.$$

Therefore, by differentiation, the marginal effects in this model are

$$\boldsymbol{\delta} = \frac{\partial E[y\,|\,\mathbf{x}]}{\partial \mathbf{x}} = \{E[y|\mathbf{x}] \times (1 - E[y|\mathbf{x}])\}(\boldsymbol{\alpha} - \boldsymbol{\beta}) = d(\boldsymbol{\alpha},\boldsymbol{\beta}) \times (\boldsymbol{\alpha} - \boldsymbol{\beta}).$$

Note that the signs of the individual coefficients are not necessarily indicative of the signs of the marginal effects.

The power model is also used to analyze a variable whose range is (0,1). The density for the random variable (as formulated in *LIMDEP*) is

$$f(y\,|\,\lambda) = (\lambda + 1)\, y^{\lambda},\ 0 < y < 1.$$

We parameterize the model by setting

$$\lambda = \exp(\boldsymbol{\beta}'\mathbf{x}).$$

For this model,

$$E[y|\mathbf{x}] = \frac{\lambda+1}{\lambda+2} \text{ and } \mathrm{Var}[y|\mathbf{x}] = \frac{\lambda+1}{\lambda+3} - \left(\frac{\lambda+1}{\lambda+2}\right)^2.$$

The marginal effects are

$$\boldsymbol{\delta} = \frac{\partial E[y\,|\,\mathbf{x}]}{\partial \mathbf{x}} = \left(\frac{\lambda}{(\lambda+2)^2}\right)\boldsymbol{\beta}.$$

We note, this model is a bit volatile, but it generally works. For analyzing a variable which is a proportion or is simply bounded by zero and one, one could use this formulation, or any of the binary choice models (probit, logit, complementary log log, Gompertz). These estimators automatically detect and adjust the estimation procedure for a proportions variable.

The model request for these models is

LOGLINEAR ; Lhs = dependent variable (must be in the range (0,1))
; Rhs = list of independent variables
; Model = Beta or Power $

All other options available for loglinear models are extended to this one. This model allows cross section analysis as above and all three panel data treatments, random parameters (**; RPM**), latent class (**; LCM**) and fixed effects (**; FEM**) as discussed below. Standard errors for the marginal effects are computed using the delta method. The models are fit by maximum likelihood. For the beta model, two vectors of parameters are produced. There is no obvious connection between them, nor any priority in their entry into the conditional mean function. The differences between the corresponding parameters feeds into the shape of the distribution. Note, for example, if $\boldsymbol{\alpha} = \boldsymbol{\beta}$, regardless of the values, the distribution is standard uniform between zero and one. Other configurations produce different shapes of the distribution.

E50.7.1 Application

A constructed example appears below:

CALC ; Ran(12345) $
SAMPLE ; 1-1000 $
CREATE ; x1 = Rnn(0,1) ; x2 = Rnn(0,1) $
CREATE ; y = Rnu(0,1) $
LOGLINEAR ; Lhs = y ; Rhs = one,x1,x2
; Model = Beta ; Partial Effects ; List $

```
-----------------------------------------------------------------------------
Beta (Loglinear) Regression Model
Dependent variable                    Y
Log likelihood function         3.07985
Restricted log likelihood       -.13875
Chi squared [   6 d.f.]         6.43719
Significance level               .37604
McFadden Pseudo R-squared    23.1973633
Estimation based on N =   1000, K =   6
Inf.Cr.AIC  =      5.8 AIC/N =     .006
Model estimated: Aug 04, 2011, 22:15:45
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in ALPHA
Constant|    -.00923         .04376     -.21  .8329     -.09499    .07653
      X1|    -.06026         .04816    -1.25  .2108     -.15465    .03413
      X2|    -.03412         .04647     -.73  .4627     -.12520    .05695
        |Parameters in BETA
Constant|     .01694         .04510      .38  .7072     -.07145    .10532
      X1|    -.09959**       .04963    -2.01  .0448     -.19687   -.00231
      X2|    -.01744         .04616     -.38  .7055     -.10791    .07302
--------+--------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point    .4939
Scale Factor for Marginal Effects   .2500
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
       Y|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      X1|    -.01506         .08991     -.17  .8669     -.19127    .16115
      X2|    -.00853         .08528     -.10  .9203     -.17567    .15861
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Predicted Values         (* => observation was not in estimating sample.)
Observation         Observed Y   Predicted Y   Residual       Lambda-A   Lambda-B
        1            .9596596     .4954268     .4642329     -.0683902   -.0500967
        2            .2399725     .4856611    -.2456886      .0565713    .1139426
        3            .7801775     .4896772     .2905003     -.0459020   -.0046050
        4            .6930874     .4941819     .1989055     -.0756854   -.0524118
        5            .7516940     .4903664     .2613276      .0361852    .0747242
        6            .8457808     .4833286     .3624522      .0843637    .1510741
        7            .5540159     .4952044     .0588115     -.0727616   -.0535786
        8            .2979716     .4991834    -.2012118      .0043903    .0076567
        9            .6014696     .4842701     .1171995      .1500031    .2129434
       10            .7536326     .4733595     .2802731      .0463947    .1530577
```

E50.7.2 Technical Details

The log likelihood for the beta model is

$$\log L = \sum_{i=1}^{N} \log\Gamma(\lambda_{a,i}+\lambda_{b,i}) - \log\Gamma(\lambda_{a,i}) - \log\Gamma(\lambda_{b,i}) + (\lambda_{a,i}-1)\log y_i + (\lambda_{b,i}-1)\log(1-y_i)\,.$$

The derivatives are

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^{N} [\Psi(\lambda_{a,i}+\lambda_{b,i}) - \Psi(\lambda_{a,i}) + \log y_i]\lambda_{a,i}\mathbf{x}_i \\ \frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{N} [\Psi(\lambda_{a,i}+\lambda_{b,i}) - \Psi(\lambda_{b,i}) + \log(1-y_i)]\lambda_{b,i}\mathbf{x}_i\end{aligned}.$$

The BHHH estimator is used for the asymptotic covariance matrix. The conditional mean function is

$$E[y|\mathbf{x}] \;=\; \frac{\lambda_a}{\lambda_a+\lambda_b} \;=\; \mu$$

The partial effects are

$$\boldsymbol{\delta} \;=\; \frac{\partial E[y\,|\,\mathbf{x}]}{\partial \mathbf{x}} \;=\; [\mu(1\text{-}\mu)](\boldsymbol{\alpha} - \boldsymbol{\beta}).$$

To find the derivatives to use the delta method for the asymptotic covariance matrix for the estimated marginal effects, we require

$$\mathbf{G} \;=\; [\partial\boldsymbol{\delta}/\partial\boldsymbol{\alpha}',\, \partial\boldsymbol{\delta}/\partial\boldsymbol{\beta}'] \;=\; [\mathbf{G}_\alpha,\mathbf{G}_\beta].$$

By a tedious application of the chain rule, we find

$$\partial[\mu(1\text{-}\mu)]/\partial\lambda_a = (1/\lambda_a)\mu(1\text{-}\mu)(1 - 2\mu)$$

and by symmetry $\quad \partial[\mu(1\text{-}\mu)]/\partial\lambda_b = (1/\lambda_b)\mu(1\text{-}\mu)(1 - 2\mu).$

Therefore,

$$\begin{aligned}\mathbf{G}_\alpha &= \mu(1\text{-}\mu)\mathbf{I} \;+\; [(1/\lambda_b)\mu(1\text{-}\mu)(1 - 2\mu)]\boldsymbol{\alpha}\,[\lambda_a\mathbf{x}'] \\ &= \mu(1\text{-}\mu)[\mathbf{I} + (1 - 2\mu)\boldsymbol{\alpha}\mathbf{x}'].\end{aligned}$$

By the symmetry of the function, we can deduce

$$\mathbf{G}_\beta = \mu(1\text{-}\mu)[\mathbf{I} + (1 - 2\mu)\boldsymbol{\beta}\mathbf{x}'].$$

For the power model, the density is

$$f(\text{y} \mid \lambda) = (\lambda + 1)\, y^{\lambda},\ 0 < y < 1$$

$$\lambda = \exp(\boldsymbol{\beta}'\mathbf{x}).$$

so the log likelihood and its derivatives are

$$\log L = \sum_{i=1}^{N} \log(1+\lambda_i) + \lambda_i \log y_i$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left[\frac{\lambda_i}{1+\lambda_i} + \lambda_i \log y_i\right] \mathbf{x}_i$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum_{i=1}^{N} \left[\frac{\lambda_i}{(1+\lambda_i)^2} + \lambda_i \log y_i\right] \mathbf{x}_i\mathbf{x}_i'$$

The marginal effects are

$$\boldsymbol{\delta} = \left(\frac{\lambda}{(\lambda+2)^2}\right)\boldsymbol{\beta}$$

so,

$$\mathbf{G} = \frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\beta}'} = \frac{\lambda}{(\lambda+2)^2}\mathbf{I} + \frac{\lambda(2-\lambda)}{(\lambda+2)^3}\boldsymbol{\beta}\mathbf{x}'$$

$$= \frac{\lambda}{(\lambda+2)^2}\left[\mathbf{I} + \frac{2-\lambda}{2+\lambda}\boldsymbol{\beta}\mathbf{x}'\right]$$

E51: Generalized Linear and Fractional Response Models for Panel Data

E51.1 Introduction

This chapter presents the panel data estimators for the generalized linear models (GLMs) described in Chapters E49 and E50.

The basic command for estimation of the models described in this chapter is

GLIM ; Lhs = dependent variable
; Rhs = independent variables
; Model = type of model
; Panel
; Panel model specification $

where ‘**type of model**’ is one of the generalized linear models presented here. The panel model specification indicates the form of the stochastic specification, fixed or random effects, random parameters, or latent class.

The models listed in Table E51.1 provide specific extensions for panel data methods.

Model	Type of Random Variable
Lognormal	Nonnegative
Binomial	Count of successes
Geometric	Count until success
Power	Bounded in (0,1)
Normal (Loglinear)	Continuous
Gamma	Nonnegative
Weibull	Nonnegative
Exponential	Nonnegative
Rayleigh	Nonnegative
Inverse Gaussian	Nonnegative

Table E51.1 Loglinear Models with Supported Panel Data Treatments

There are no panel data estimators supported for the lognormal or beta models. The full set of panel data treatments, fixed effects, random effects, random parameters, and latent class, are supported for these nine models.

An additional modeling framework that is similar to the ones listed above is Papke and Wooldridge’s (2008) *fractional response model* for panel data, which is presented in Section E51.8.

E51.2 GEE Modeling

The four panel data treatments noted above provide the most common applications of longitudinal, or repeated measures methods to microeconomic data of the sort of interest here. A widely cited development in the statistics literature, generalized equation estimation (GEE) modeling appears to be yet another form of estimator. (See Liang and Zeger (1986) and Diggle, Liang and Zeger, (1994).) This is an extension of the GLM framework to panel data applications. The GEE estimator is not explicitly supported in *LIMDEP* directly as a preprogrammed routine. However, most of the internally consistent forms of GLM/GEE models are actually contained in the random parameters model package in *LIMDEP*. As this is a frequently asked question, we consider it in detail.

The GEE approach adds what is essentially a random effects form to a panel data treatment in the preceding GLM models. We redefine the link function as

$$f(\mathrm{E}[y_{it}|\,\mathbf{x}_{it}]) \;=\; \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \varepsilon_{it},\ t = 1,...,T_i.$$

Now, consider some different approaches to formulating the $T_i \times T_i$ covariance matrix for the heterogeneity: (We borrow the nomenclature from the GEE literature):

Independent:	$\mathrm{Corr}[\varepsilon_{it}, \varepsilon_{is}] \;=\; 0,\ t \neq s$
Exchangeable:	$\mathrm{Corr}[\varepsilon_{it}, \varepsilon_{is}] \;=\; \rho,\ t \neq s$
AR(1):	$\mathrm{Corr}[\varepsilon_{it}, \varepsilon_{is}] \;=\; \rho^{\lvert t-s \rvert},\ t \neq s$
Nonstationary:	$\mathrm{Corr}[\varepsilon_{it}, \varepsilon_{is}] \;=\; \rho_{ts},\ t \neq s,\ \lvert t-s \rvert \leq g$
Unstructured:	$\mathrm{Corr}[\varepsilon_{it}, \varepsilon_{is}] \;=\; \rho_{ts},\ t \neq s.$

The GEE approach to estimation is a form of generalized method of moments. Most of these models are already available in other forms in *LIMDEP*. The first one is obvious – this is just the pooled estimator ignoring any group effects – we considered this model in Chapters E49 and E50. The second is the random effects model. We have noted a large number of models, including most of those in the valid set of GLIMs that *LIMDEP* can fit in the random effects form. In addition, all models that are available in the random parameters form can be fit with just a random constant term to provide this random effects model. This includes most of the GLM models and some others, such as the tobit model. This model is produced simply by writing the random constants model as

$$f(\mathrm{E}[y_{it}|\,\mathbf{x}_{it}]) \;=\; \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; u_{it},\ t = 1,...,T_i$$

$$\alpha_i \;=\; \alpha \;+\; w_i,\ i = 1,...,N.$$

Thus, the random constants model is functionally equivalent to the GEE model/estimator with the ‘exchangeable’ form of the covariance matrix. There is, however, an important difference in the treatment. The GEE estimator is a type of method of moments estimator. (See Diggle et al. (1994) for documentation.) The estimator in *LIMDEP* is maximum simulated likelihood. In addition, the random parameters model allows an AR(1) format for the random constant term, so all the models that fit in the exchangeable case can also be fit as in the AR(1) case. (See Chapter R24 for details on random parameters estimation.)

The nonstationary covariance matrix is a restricted form of the of the unstructured covariance matrix, in which covariances are restricted to be zero after a certain lag. It is possible to obtain both of these forms by using freely correlated random period specific constant terms (i.e., time dummy variables) in the model. It might also be desired to force the means of the variables to be equal, so as to match exactly the structure above. We do note, however, these sorts of models are very weakly identified in any estimation setting, owing to the large number of parameters that must be estimated to characterize the distribution of an unobserved random vector. A fully unstructured correlation matrix, for example, is nearly inestimable as an ancillary parameter in a model fit by maximum likelihood, because the log likelihood becomes quite flat in the space of the correlations. If the panel is at all large, users should not be optimistic about fitting models such as the unstructured one above using GEE, MSL or any other technique. (For example, *LIMDEP*'s multinomial probit and multivariate probit models face this difficulty.)

LIMDEP can estimate most GEE models. The estimation technique however, is simulated maximum likelihood, not the method of moments. By construction, *LIMDEP*'s estimator will be more efficient asymptotically, though in the sizes typical of panel data sets, this will probably be a minor consideration. We note, finally, ability to structure the random parameters model with random coefficients on all variables, rather than just the constant term, makes this estimator, in fact, far more general than the GEE estimator. The end result would be, in answer to the frequently asked question, yes, *LIMDEP* does do GLIM and GEE estimation, and considerably more with the random parameters model.

E51.3 Panel Data Models

There are several general formulations for extensions of the regression models to a panel data setting. These include, where $f(.)$ denotes the density for the observed random variable (i.e., the model),

- **Fixed effects:** $f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$; α_i may be correlated with $\mathbf{x}_{it}$,
- **Random effects:** $f(y_{it}) = f(\boldsymbol{\beta}'\mathbf{x}_{it} + u_i)$; u_i is uncorrelated with $\mathbf{x}_{it}$,
- **Random parameters:** $f(y_{it}) = f(\boldsymbol{\beta}_i'\mathbf{x}_{it})$, $\boldsymbol{\beta}|i \sim h(\boldsymbol{\beta}|i)$ with mean vector $\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i$ and covariance matrix $\boldsymbol{\Sigma}$
- **Latent class:** $f(y_{it}|\text{class } j) = f(\boldsymbol{\beta}_j'\mathbf{x}_{it})$, $\text{Prob}[\text{class} = j] = F_j(\boldsymbol{\theta})$

NOTE: The inverse Gaussian regression model is extremely volatile, particularly with fixed effects. It is essential to have a good set of starting values, and even with these, the model is still often difficult to fit. To fit this model in any of the panel data forms, you must precede your command with estimation of the same model with no heterogeneity. That is, the immediately preceding command (each time you use it) must be

LOGLINEAR ; Lhs = ... ; Rhs = ... ; Model = Inverse Gaussian $

There is no natural starting value for the exponential regression model, other than the one fit to the pooled sample, as developed above. You must provide the values in the same way for this generalized regression model, with

LOGLINEAR ; Lhs = ... ; Rhs = ... ; Model = Normal $

Once again, the model specification must be otherwise identical in the pooled and panel data estimators.

The same features, options, and panel data treatments are provided for all ten models listed in Table E51.1. For convenience, the relevant specifications and restrictions for all of them are listed below. In all cases, the basic random effects model can be estimated by using the random parameters model with only a random constant term. The binomial regression model (only) also supports a quadrature based (Butler and Moffitt estimator) for the random effects model. This is noted again below.

Standard Model Specifications for the Panel Data Loglinear Models

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Margin displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm (not available for FEM).
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator (not for FEM).
; Rst = list specifies equality and fixed value restrictions (not for FEM).

E51.4 Fixed Effects Models

The fixed effects model assumes a group specific effect:

$$f(y_{it}) = f(\lambda_{it})$$

where $\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)$

where α_i is the parameter to be estimated. You may also fit a two way fixed effects model

$$\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + \gamma_t)$$

where γ_t is an additional, time (period) specific effect. This model is fit (in principle) as a dummy variable with separate dummy variables in the model for each individual (and group for the two way model).

The command for estimation of the fixed effects models is

LOGLINEAR ; Lhs = dependent variable
; Rhs = independent variables
; Model = Binomial, Geometric, Power, Exponential, Gamma, Weibull,
Inverse Gaussian, Rayleigh, Power, Geometric, Binomial
; Pds = panel specification
; FEM (for fixed effects model) $

(See the earlier note about the command for the inverse Gaussian regression model. The fixed effects form is not supported for the normal model.) You may request residuals, fitted values, marginal effects, and all other optional features with this model. Restrictions, with **; Rst**, however, must be built into the model at the outset. The algorithm does not accommodate restrictions. Full details on estimating fixed effects models appear in Section R23.2.

NOTE: Your Rhs list should not include a constant term, as the fixed effects model fits a complete set of constants for the set of groups. If you do include *one* in your Rhs list, it is removed prior to beginning estimation.

The fixed effects models are estimated by maximum likelihood. The time specific effect is requested by adding

; Time

to the command if the panel is balanced, and

; Time = variable name

if the panel is unbalanced.

For the unbalanced panel, we assume that overall, the sample observation period is

$$t = 1,2,..., T_{max}$$

and that the time variable gives for the specific group, the particular values of t that apply to the observations. Thus, suppose your overall sample is five periods. The first group is three observations, periods 1, 2, 4, while the second group is four observations, 2, 3, 4, 5. Then, your panel specification would be

; Pds = Ti, for example, where *Ti* = 3, 3, 3, 4, 4, 4, 4
and **; Time = Pd**, for example, where *Pd* = 1, 2, 4, 2, 3, 4, 5.

Results that are kept for this model are

Matrices: *b* = estimate of β
varb = asymptotic covariance matrix for estimate of β
alphafe = estimated fixed effects

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

The upper limit on the number of groups is 100,000.

To illustrate the fixed effects estimator, we fit a Weibull model to the distribution of income in the health care data. The model is somewhat limited, since it cannot accommodate time invariant regressors. We use health satisfaction, marital status, presence of children and working status. (It would be natural to include age and education. This produces perfect collinearity in the two way fixed effects model – age can be expressed as a linear combination of the time dummy variables and any other nonzero variable.) The results below compare a pooled model, the fixed effects model, and a two way model with time effects. One household with zero income is also removed from the sample.

```
SAMPLE      ; All $
REJECT      ; hhninc = 0 $
SETPANEL    ; Group = id ; Pds = ti $
REJECT      ; ti < 7 $
LOGLINEAR   ; Lhs = hhninc ; Rhs = one,hsat,married,hhkids,working
            ; Model = Weibull ; Partial Effects $
LOGLINEAR   ; Lhs = hhninc ; Rhs = one,hsat,married,hhkids,working
            ; Model = Weibull ; Partial Effects
            ; FEM ; Panel $
LOGLINEAR   ; Lhs = hhninc ; Rhs = one,hsat,married,hhkids,working
            ; Model = Weibull ; Partial Effects
            ; FEM ; Panel ; Time $
```

```
----------------------------------------------------------------------------
Weibull (Loglinear) Regression Model
Dependent variable               HHNINC
Skipped    0 groups with inestimable ai
Log likelihood function      3047.13922
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.24391***      .01755    70.88  .0000     1.20952   1.27831
    HSAT|    -.01057***      .00213    -4.97  .0000     -.01474   -.00640
 MARRIED|    -.12722***      .00881   -14.44  .0000     -.14449   -.10995
  HHKIDS|     .08500***      .00767    11.08  .0000      .06996    .10004
 WORKING|    -.21951***      .00767   -28.61  .0000     -.23455   -.20448
        |Scale parameter for Weibull model
 P_scale|    2.24607***      .01379   162.87  .0000     2.21904   2.27310
--------+-------------------------------------------------------------------
FIXED EFFECTS Weibull Model
Log likelihood function      6350.69266
Estimation based on N =   6202, K = 891
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Index function for probability
    HSAT|     .04355***      .00974     4.47  .0000      .02446    .06265
 MARRIED|   -1.04619***      .08997   -11.63  .0000    -1.22254   -.86985
  HHKIDS|    1.03632***      .05126    20.22  .0000      .93585   1.13678
 WORKING|    -.71050***      .05750   -12.36  .0000     -.82320   -.59779
        |Scale parameter for Weibull distribution
 P_scale|    4.20649***      .04216    99.77  .0000     4.12386   4.28912
--------+-------------------------------------------------------------------
FIXED EFFECTS Weibull Model
Log likelihood function      7378.57387
No. of period specific effects= 6
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Index function for probability
    HSAT|    -.01554         .00962    -1.62  .1063     -.03440    .00332
 MARRIED|   -1.02930***      .09712   -10.60  .0000    -1.21965   -.83895
  HHKIDS|     .41377***      .05553     7.45  .0000      .30493    .52261
 WORKING|    -.98886***      .06115   -16.17  .0000    -1.10871   -.86900
 Period1|    2.08793***      .06153    33.94  .0000     1.96734   2.20852
 Period2|    2.06704***      .06072    34.04  .0000     1.94804   2.18605
 Period3|    1.94827***      .05857    33.26  .0000     1.83347   2.06307
 Period4|    1.86938***      .05743    32.55  .0000     1.75682   1.98193
 Period5|    1.67965***      .05629    29.84  .0000     1.56932   1.78999
 Period6|     .74510***      .05269    14.14  .0000      .64184    .84836
        |Scale parameter for Weibull distribution
 P_scale|    5.13238***      .05313    96.60  .0000     5.02825   5.23651
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the Xs.
--------+-----------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
  HHNINC|      Effect       Error        z     |z|>Z*        Interval
--------+-----------------------------------------------------------------------
(Pooled)
    HSAT|     .00274***     .00055     4.96  .0000      .00166     .00382
 MARRIED|     .03298***     .00227    14.53  .0000      .02853     .03743
  HHKIDS|    -.02204***     .00199   -11.09  .0000     -.02593    -.01814
 WORKING|     .05691***     .00199    28.62  .0000      .05301     .06081
--------+-----------------------------------------------------------------------
(One way fixed effects)
    HSAT|    -.00220***    -.06933    -4.75  .0000     -.00311    -.00129
 MARRIED|     .05296***     .21021     9.54  .0000      .04208     .06383
  HHKIDS|    -.05246***    -.11190   -18.64  .0000     -.05797    -.04694
 WORKING|     .03596***     .12487    10.65  .0000      .02934     .04258
--------+-----------------------------------------------------------------------
(Two way fixed effects)
    HSAT|     .00054        .02027     1.58  .1133     -.00013     .00122
 MARRIED|     .03610***     .16951     9.03  .0000      .02826     .04393
  HHKIDS|    -.01451***    -.03662    -7.53  .0000     -.01829    -.01073
 WORKING|     .03468***     .14244    13.56  .0000      .02967     .03969
--------+-----------------------------------------------------------------------
(Random effects)
  HSAT|    -.00085*       -.01636    -1.66  .0965     -.00186     .00015
 MARRIED|     .05185***     .12573    21.99  .0000      .04723     .05647
  HHKIDS|    -.06594***    -.08592   -28.44  .0000     -.07048    -.06139
 WORKING|     .05527***     .11723    25.98  .0000      .05110     .05944
--------+-----------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E51.5 Random Effects Models

The random effects model also assumes a group specific effect:

$$f(y_{it}) = f(\lambda_{it})$$

where $$\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i),\ \mathrm{u_i} \sim N[0,1],$$

where σ_u is the one additional parameter to be estimated. This differs from the fixed effects model in the assumption that u_i is uncorrelated with $\mathbf{x}_{it}$. For example, the binomial regression with this formulation would have

$$\pi_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i)/[1+ \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_u u_i)] \text{ where } \sigma_u u_i \sim N[0,\sigma_u^2].$$

The Butler and Moffitt procedure for estimating this model has been incorporated in many random effects estimators, including many models in *LIMDEP*. A full listing of the frameworks appears in Section R23.3. The approach uses Hermite quadrature to evaluate the one dimensional normal integral in the conditional log likelihood. An alternative method of estimating one factor random effects models is via maximum simulated likelihood in a random parameters model with only a random constant term. This is described in Chapter R24, and in the next section below.

Random effects estimators for the loglinear models identified in Table E51.1 are obtained by using the random parameters approach. The generic command would be

```
LOGLINEAR ; Lhs = the dependent variable
          ; Rhs = one,... the remaining independent variables
          ; ... any other model specifications
          ; Pds = the panel data specification
          ; RPM ; Fcn = one(n) [; Halton ; Pts = the desired number]
          ; Model = one of Normal, Exponential, Gamma, Weibull, Rayleigh,
                    Inverse Gaussian, Power, Geometric, Binomial $
```

An example appears below. In addition, (only) one of these models, the binomial model, may be fit with the Butler and Moffitt estimator. The command would be

```
LOGLINEAR ; Lhs = the dependent variable
          ; Rhs = one,... the remaining independent variables
          ; Trials = the specification
          ; Pds = the panel data specification ; Normal
          ; Model = Binomial $
```

To illustrate, the following estimates a Weibull model for the distribution of incomes in the health care data. Note that a pooled exponential regression (P = 1) (not shown) is used to obtain the starting values. The likelihood ratio statistic of over 8,000 firmly rejects this hypothesis. Whether this is from the random parameters part of the model or the shape of the distribution remains to be determined. Strictly within the Weibull results, we find the null hypothesis of P = 1, can be firmly rejected based on a Wald (t) statistic of (2.24607 – 1)/0.01378 = 90.36.

```
SAMPLE    ; All $
REJECT    ; hhninc = 0 $
SETPANEL  ; Group = id ; Pds = ti $
REJECT    ; ti < 7 $
LOGLINEAR ; Lhs = hhninc
          ; Rhs = one,hsat,married,hhkids,working
          ; Model = Weibull ; Partial Effects
          ; RPM ; Panel ; Fcn = one(n) ; Pts = 25 ; Halton $
```

```
-----------------------------------------------------------------------------
Random Coefficients  WeiblReg Model
Dependent variable               HHNINC
Log likelihood function      4558.85190
Unbalanced panel has    886 individuals
Weibull loglinear regression model
Simulation based on     25 Halton draws
-----------------------------------------------------------------------------
```

```
--------+-------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
        |Nonrandom parameters
    HSAT|     .00888*         .00525     1.69  .0909     -.00141    .01917
 MARRIED|    -.54048***       .02534   -21.33  .0000     -.59014   -.49081
  HHKIDS|     .68732***       .02077    33.09  .0000      .64661    .72803
 WORKING|    -.57613***       .02166   -26.60  .0000     -.61858   -.53368
        |Means for random parameters
Constant|    3.96401***       .04425    89.59  .0000     3.87729   4.05073
        |Scale parameters for dists. of random parameters
Constant|    1.71943***       .02040    84.29  .0000     1.67944   1.75941
        |Scale parameter for Weibull distribution
 P_scale|    3.63332***       .02142   169.60  .0000     3.59133   3.67530
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

E51.6 Random Parameters Models

The random parameters model may be specified for all nine loglinear models. The structure of the random parameters model is based on the conditional density

$$f(y_{it}|\, \mathbf{x}_{it}, \boldsymbol{\beta}_i) \;=\; f(\boldsymbol{\beta}_i{}'\mathbf{x}_{it}),\; i = 1,...,N,\; t = 1,...,T_i.$$

where $f(.)$ is the density for the particular model. The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) means

$$\text{E}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] \;=\; \boldsymbol{\beta} \;+\; \boldsymbol{\Delta}\mathbf{z}_i.$$

(The second term is optional – the mean may be constant.)

$$\text{Var}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] \;=\; \boldsymbol{\Sigma}.$$

The model is operationalized by writing

$$\boldsymbol{\beta}_i \;=\; \boldsymbol{\beta} \;+\; \boldsymbol{\Delta}\mathbf{z}_i \;+\; \boldsymbol{\Gamma}\mathbf{v}_i.$$

As noted earlier, the heterogeneity term is optional. In addition, it may be assumed that some of the parameters are nonrandom. It is convenient to analyze the model in this fully general form here. One could easily accommodate nonrandom parameters just by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$.

NOTE: If there is no heterogeneity in the mean, and only the constant term is considered random – the model may specify that some parameters are nonrandom – then this model is equivalent to the random effects model.

E51.6.1 Command for the Random Parameters Model

The basic model command for this form of the model is

LOGLINEAR ; Lhs = dependent variable
; Rhs = independent variables
; Model = Exponential, Gamma, Weibull, Inverse Gaussian,
Rayleigh, Power, Normal, Geometric, Binomial
; Pds = panel specification
; RPM (or ; RPM = list of variables in z)
; Fcn = specifications of the random parameters
[; Pts = number of replications and ; Halton are optional] $

(See the earlier note about the command for the inverse Gaussian regression and normal exponential models.)

NOTE: For this model, your Rhs list should include a constant term.

NOTE: The **; Pds** specification is optional. You may fit these models with cross section data. There is nothing inherent in the model that limits it to a panel data application.

Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one, x1, x2, x3, x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be constant. For those that you wish to specify as random, use

; Fcn = variable name (distribution), variable name (distribution), ...

Three distributions may be specified. All random variables have mean zero.

n = standard normal distribution, variance = 1,
t = triangular (tent shaped) distribution in [-1,+1], variance = 1/6,
u = standard uniform distribution [-1,1], variance = 1/3.

(Several other available distributions are listed in Section R24.3.) Note that each of these is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. The latter two distributions are provided as one may wish to reduce the amount of variation in the tails of the distribution of the parameters across individuals and to limit the range of variation. (See Train (2010) for discussion.). To specify that the constant term and the coefficient on *x*1 are normally distributed with fixed mean and variance, use

; Fcn = one(n), x1(n)

This specifies that the first and second coefficients are not random while the remainder are. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

NOTE: The model with only a random constant term (**; Fcn = one(n)**) is precisely equivalent to a 'random effects' model.

Correlated Random Parameters

The preceding defines an estimator for a model in which the covariance matrix of the random parameters is diagonal. To extend it to a model in which the parameters are freely correlated, add

; Correlation (or just ; Cor)

to the command.

Heterogeneity in the Means

The preceding examples have specified that the mean of the random variable is fixed over individuals. If there is measured heterogeneity in the means, in the form of

$$E[\beta_{ki}] = \beta_k + \Sigma_m \delta_{km} z_{mi}$$

where z_m is a variable that is measured for each individual, then the command may be modified to

; RPM = list of variables in z

In the data set, these variables must be repeated for each observation in the group.

The Parameter Vector and Starting Values

The parameter vector is laid out as follows, in this order:

$\alpha_1, ..., \alpha_K$	are the K nonrandom parameters,
$\beta_1,...,\beta_M$	are the M means of the distributions of the random parameters,
$\sigma_1,\sigma_2,...,\sigma_M$	are the M scale parameters for the distributions of the random parameters,
P	is the shape parameter in the gamma, inverse Gauss or Weibull model (last parameter).

These are the essential parameters. If you have specified that parameters are to be correlated, then the σs are followed by the below diagonal elements of **Γ**. (The σs are the diagonal elements.) If you have specified heterogeneity variables, **z**, then the preceding are followed by the rows of **Δ**. Consider an example: The model specifies:

; Model = Weibull
; RPM = z1,z2
; Rhs = one,x1,x2,x3,x4 ? The base parameters are β_1, β_2, β_3, β_4, β_5.
; Fcn = one(n),x2(n),x4(n)
; Correlated

Then, after rearranging, the model becomes

Variable	**Parameter**
$x1$	α_1
$x3$	α_2
one	$\beta_1 + \sigma_1 v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
$x2$	$\beta_2 + \sigma_2 v_{i2} + \gamma_{21} v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
$x4$	$\beta_3 + \sigma_3 v_{i3} + \gamma_{31} v_{i1} + \gamma_{32} v_{i2} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$

and the parameter vector would be

$$\theta = \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2, \sigma_3, \gamma_{21}, \gamma_{31}, \gamma_{32}, \delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}, \delta_{31}, \delta_{32}, P$$

You may use **; Rst** and **; CML** to impose restrictions on the parameters. Use the preceding as a guide to the arrangement of the parameter vector.

Results saved by this estimator are:

Matrices: *b* = estimate of $\boldsymbol{\theta}$
varb = asymptotic covariance matrix for estimate of $\boldsymbol{\theta}$
beta_i = individual specific parameters, if **; Par** is requested
sdbeta_i = estimated standard deviations of conditional distributions

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

Last Function: None

E51.6.2 Application

The example below continues the application shown in the preceding sections. Here, we fit the model with two random parameters, both heterogeneous and heteroscedastic. The specification for the random coefficient on *working*, for example, is

$$\beta_{working,i} = \beta_{working} + \delta Female_i + \sigma_{working,i,} w_{working,i}$$

$$\sigma_{working,i} = \sigma_{working} + \theta_{working} age_i$$

$$w_{working,i} \sim \text{N}[0,1].$$

The value used for age_i is the observation in the first period of the observation.

The command set is:

SAMPLE ; **All $**
REJECT ; **hhninc = 0 $** (There are four bad observations in the data set)
SETPANEL ; **Group = id ; Pds = ti $**
REJECT ; **ti # 7 $**
LOGLINEAR ; **Lhs = hhninc**
; **Rhs = one,hsat,married,hhkids,working**
; **Model = Weibull ; Partial Effects**
; **RPM = female ; Panel ; Fcn = one(n),working(n); Het ; Hfr = age**
; **Pts = 25 ; Halton $**

```
-----------------------------------------------------------------------------
Expnontl Regression Start Values for HHNINC
Dependent variable               HHNINC
Log likelihood function       388.47572
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    HSAT|    -.00628         .00577    -1.09  .2761     -.01758    .00502
 MARRIED|    -.26601***      .03585    -7.42  .0000     -.33627   -.19575
  HHKIDS|     .11582***      .02629     4.41  .0000      .06430    .16734
Constant|    1.46378***      .05179    28.26  .0000     1.36227   1.56530
 WORKING|    -.25276***      .02937    -8.60  .0000     -.31034   -.19519
--------+--------------------------------------------------------------------
Random Coefficients  WeiblReg Model
Dependent variable               HHNINC
Log likelihood function      3854.38119
Restricted log likelihood     388.47572
Unbalanced panel has     886 individuals
Weibull loglinear regression model
Simulation based on     25 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    HSAT|    -.02500***      .00581    -4.30  .0000     -.03640   -.01361
 MARRIED|    -.71013***      .02641   -26.89  .0000     -.76189   -.65837
  HHKIDS|     .43228***      .02414    17.91  .0000      .38498    .47959
        |Means for random parameters
Constant|    3.19326***      .06649    48.03  .0000     3.06294   3.32357
 WORKING|    -.42880***      .05322    -8.06  .0000     -.53310   -.32449
        |Scale parameters for dists. of random parameters
Constant|     .22335***      .01970    11.34  .0000      .18474    .26195
 WORKING|     .17416***      .02536     6.87  .0000      .12447    .22386
        |Heterogeneity in the means of random parameters
cONE_FEM|     .13341**       .05595     2.38  .0171      .02374    .24308
cWOR_FEM|    -.44091***      .06662    -6.62  .0000     -.57149   -.31034
        |Heterogeneity in the variances of random parameters
hONE_AGE|     .02637***      .00204    12.91  .0000      .02237    .03037
hWOR_AGE|    -.15711***      .00348   -45.19  .0000     -.16392   -.15029
        |Scale parameter for Weibull distribution
 P_scale|    2.35600***      .01635   144.12  .0000     2.32396   2.38804
--------+--------------------------------------------------------------------
```

```
Partial derivatives of expected val. with
respect to the vector of characteristics.
--------+-------------------------------------------------------------------
        |     Partial                           Prob.      95% Confidence
  HHNINC|      Effect    Elasticity      z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
    HSAT|     .00349***       .07105     4.33  .0000       .00191    .00507
 MARRIED|     .09915***       .25476    25.78  .0000       .09162    .10669
  HHKIDS|    -.06036***      -.08334   -17.34  .0000      -.06718   -.05354
 WORKING|     .05987***       .13455     7.11  .0000       .04336    .07638
--------+-------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

E51.7 Latent Class Loglinear Regression Models

The model for a panel of data, $i = 1,...,N$, $t = 1,...,T_i$ is

$$f(y_{it}|\,\mathbf{x}_{it}) = f(y_{it},\boldsymbol{\beta}'\mathbf{x}_{it}) = f(i,t).$$

Henceforth, we use the term 'group' to indicate the T_i observations on respondent i in periods $t = 1,...,T_i$. Unobserved heterogeneity in the distribution of y_{it} is assumed to impact the density in the form of a random effect. The continuous distribution of the heterogeneity is approximated by using a finite number of 'points of support.' The distribution is approximated by estimating the location of the support points and the mass (probability) in each interval. In implementation, it is convenient and useful to interpret this discrete approximation as producing a sorting of individuals (by heterogeneity) into J classes, $j = 1,...,J$. (Since this is an approximation, J is chosen by the analyst.)

Thus, we modify the model for a latent sorting of y_{it} into J 'classes' with a model which allows for heterogeneity as follows: The density of the observed y_{it} given that regime j applies is

$$f(y_{i,t}|j) = f(y_{it}|\,\mathbf{x}_{it}, j)$$

where the density is now specific to the group. The analyst does not observe directly which class, $j = 1,...,J$ generated observation $y_{it}|j$, and class membership must be estimated. Heckman and Singer (1984) suggest a simple form of the class variation in which only the constant term varies across the classes. This would produce the model

$$f(y_{i,t}\,|j) = f(y_{it},\,\boldsymbol{\beta}'\mathbf{x}_{it} + \delta_j),\ \text{Prob}(\text{class} = j) = F_j.$$

We formulate this approximation more generally as,

$$f(y_{i,t}\,|j) = f[y_{it}\,|\,\boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}_j'\mathbf{x}_{it},\,P_j)$$

$$F_j = \exp(\theta_j) / \Sigma_j \exp(\theta_j),\ \text{with } \theta_J = 0.$$

If the prior class probabilities are functions of observed variables, then they may be extended in the form of a multinomial logit model, with

$$\theta_{ij} = \boldsymbol{\theta}_j'\mathbf{z}_i.$$

This is done by adding the specification of **z** to the command as shown below. In this formulation, each group has its own parameter vector, $(\boldsymbol{\beta}_j',\sigma_j) = (\boldsymbol{\beta} + \boldsymbol{\delta}_j,P_j)$ though the variables that enter the mean are assumed to be the same. (This can be changed by imposing restrictions on the full parameter vector, as described below.) This allows the Heckman and Singer formulation as a special case by imposing restrictions on the parameters.

The estimation command for this model is

```
LOGLINEAR ; Lhs = dependent variable
          ; Rhs = independent variables
          ; Model = Exponential, Gamma, Weibull, Inverse Gaussian,
                    Rayleigh, Power, Normal, Geometric, Binomial
          ; Pds = panel specification
          ; LCM (or ; LCM = list of variables in z)
          ; Pts = number of classes $
```

The default number of support points is five. But, this is fairly high. You may set J to 2, 3, ...,9 with

; Pts = the value you wish

Some particular values computed for the latent class model are

; Group = the index of the most likely latent class
; Cprob = estimated probability for the most likely latent class

You can obtain a listing of these two results by using

; List

An example appears below. Computation of these values is described in the technical details in Chapter R25.

You can use the **; Rst = list** option to structure the latent class model so that different variables appear in different classes. For example, the following would restrict a model so that $x1$ appears in one class and $x2$ in the other:

; Rst = a1,bx1,0, a2,0,bx2,theta1,theta2.

Alternatively, you can use this device to construct the Heckman and Singer form of the model as follows, where we use a three class model as an example:

```
NAMELIST  ; x = ... one, list of variables $
CALC      ; kx1 =  Col(x) - 1 $
LOGLINEAR ; Lhs = ... ; Rhs = x ; LCM ; Pts = 3
          ; Model = Exponential, Gamma, Weibull, etc.
          ; Rst =  d1,kx1_b, pshape1,
                   d2,kx1_b, pshape2,
                   d3,kx1_b, pshape3, t1,t2,t3 $
```

Estimates retained by this model include

Matrices:	*b*	= full parameter vector, $[\boldsymbol{\beta}_1', P_1, \boldsymbol{\beta}_2', P_2, \ldots F_1, \ldots, F_J]$
	varb	= full covariance matrix
	beta_i	= individual specific parameters, if **; Par** requested
	classp_i	= individual specific posterior class probabilities if **; Par** requested

Note that *b* and *varb* involve $J\times(K+2)$ estimates. Two additional matrices are created,

	b_class	= a $J\times K$ matrix with each row equal to the corresponding $\boldsymbol{\beta}_j$
	class_pr	= a $J\times 1$ vector containing the estimated class probabilities

Scalars:	*kreg*	= number of variables in Rhs list
	nreg	= total number of observations used for estimation
	logl	= maximized value of the log likelihood function
	exitcode	= exit status of the estimation procedure

Last Function: None

Application

The following repeats the earlier example in the latent class framework. We fit the model with three latent classes, and allow the prior class probabilities to depend on *female*. The general formulation is as follows:

```
SAMPLE      ; All $
REJECT      ; hhninc = 0 $
SETPANEL    ; Group = id ; Pds = ti $
REJECT      ; ti # 7 $
LOGLINEAR   ; Lhs = hhninc
            ; Rhs = one,hsat,married,hhkids,working
            ; Model = Weibull ; Partial Effects
            ; LCM = female ; Panel ; Pts = 3 ; Par $
```

The model requests that the posterior class probabilities be retained in the matrix *classp_i*. This is an 886×3 matrix, shown below. The average class probabilities are shown below the model results. The matrix command,

```
MATRIX      ; List ; 1/886 * 1'classp_i $
```

verifies the computation.

A restriction that the coefficients on *hhkids* and *working* and the shape parameter, *P*, be equal in all three classes is imposed by adding

```
; Rst = a1,a2,a3,b1,b2,pw,c1,c2,c3,b1,b2,pw,d1,d2,d3,b1,b2,pw,t1,t2,t3
```

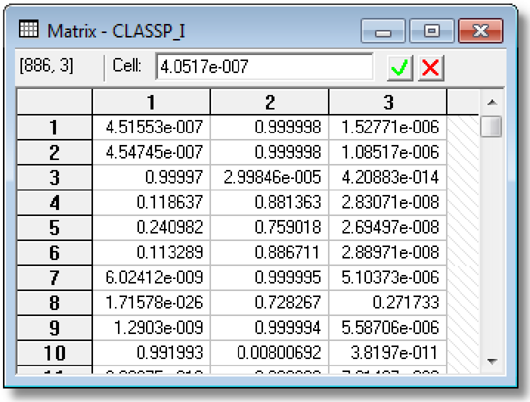
to the command.

The results for this model are shown below. This and the general model are nested, and we can use a likelihood ratio to test the restriction as a hypothesis. The log likelihoods for the unrestricted and restricted models are 4615.52967 and 4476.74896, respectively. The chi squared likelihood ratio statistic would be twice the difference, or 277.56 with 6 degrees of freedom (3 restrictions times (3-1) classes). The hypothesis would be rejected. The restriction for producing the Heckman and Singer model would be

; Rst = a1,4_b,pw,a2,4_b,pw,a3,4_b,pw,t1,t2,t3

```
-----------------------------------------------------------------------------
Latent Class / Panel WeiblReg Model
Dependent variable              HHNINC
Log likelihood function     4615.52967
Restricted log likelihood    388.47572
Weibull loglinear regression model
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    1.63419***      .16181    10.10  .0000     1.31704   1.95133
    HSAT|    -.05480***      .01628    -3.37  .0008     -.08671   -.02290
 MARRIED|     .30384**       .11854     2.56  .0104      .07150    .53617
  HHKIDS|     .16158**       .07431     2.17  .0297      .01594    .30723
 WORKING|    -.51480***      .08730    -5.90  .0000     -.68591   -.34370
 P_scale|    2.35389***      .06473    36.37  .0000     2.22703   2.48076
        |Model parameters for latent class 2
Constant|    4.01033***      .11676    34.35  .0000     3.78148   4.23919
    HSAT|     .02925***      .01096     2.67  .0076      .00777    .05073
 MARRIED|    -.77124***      .07256   -10.63  .0000     -.91346   -.62902
  HHKIDS|     .71216***      .05378    13.24  .0000      .60676    .81756
 WORKING|    -.66530***      .06554   -10.15  .0000     -.79376   -.53685
 P_scale|    3.73992***      .06670    56.07  .0000     3.60919   3.87064
        |Model parameters for latent class 3
Constant|    6.22776***      .16645    37.41  .0000     5.90152   6.55400
    HSAT|     .00548         .01243      .44  .6595     -.01889    .02984
 MARRIED|   -1.33406***      .08064   -16.54  .0000    -1.49210  -1.17601
  HHKIDS|     .53559***      .06418     8.35  .0000      .40981    .66137
 WORKING|    -.93793***      .07017   -13.37  .0000    -1.07546   -.80041
 P_scale|    3.59117***      .06778    52.98  .0000     3.45831   3.72402
        |Estimated prior probabilities for class membership
   ONE_1|    -.91044***      .17327    -5.25  .0000    -1.25004   -.57084
FEMALE_1|    -.01473         .23463     -.06  .9500     -.47459    .44514
   ONE_2|     .32248**       .13658     2.36  .0182      .05479    .59017
FEMALE_2|    -.02702         .17348     -.16  .8762     -.36704    .31300
   ONE_3|        0.0    .....(Fixed Parameter).....
FEMALE_3|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

```
+-------------------------------------------------------------+
|  Prior class probabilities at data means for LCM variables |
|   Class 1     Class 2     Class 3     Class 4     Class 5  |
|    .14463      .49367      .36170      .00000      .00000  |
+-------------------------------------------------------------+
   |             1              2              3
--------+--------------------------------------------
       1|     .144623        .493670        .361707
```


Matrix - CLASSP_I

[886, 3] Cell: 4.0517e-007

	1	2	3
1	4.51553e-007	0.999998	1.52771e-006
2	4.54745e-007	0.999998	1.08517e-006
3	0.99997	2.99846e-005	4.20883e-014
4	0.118637	0.881363	2.83071e-008
5	0.240982	0.759018	2.69497e-008
6	0.113289	0.886711	2.88971e-008
7	6.02412e-009	0.999995	5.10373e-006
8	1.71578e-026	0.728267	0.271733
9	1.2903e-009	0.999994	5.58706e-006
10	0.991993	0.00800692	3.8197e-011

Figure E51.1 Matrix Results

```
-----------------------------------------------------------------------------
Latent Class / Panel WeiblReg Model
Dependent variable               HHNINC
Log likelihood function      4476.74896
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    1.25414***     .25515     4.92  .0000      .75405   1.75423
    HSAT|    -.10689***     .02517    -4.25  .0000    -.15623   -.05755
 MARRIED|    1.11362***     .16610     6.70  .0000      .78807   1.43916
  HHKIDS|     .46848***     .04315    10.86  .0000      .38391    .55305
 WORKING|    -.74635***     .04931   -15.14  .0000    -.84299   -.64971
 P_scale|    3.30483***     .03145   105.08  .0000     3.24319   3.36648
        |Model parameters for latent class 2
Constant|    3.55050***     .11101    31.98  .0000     3.33293   3.76807
    HSAT|     .03419***     .01179     2.90  .0038      .01107    .05730
 MARRIED|    -.63192***     .07773    -8.13  .0000    -.78427   -.47957
        |Model parameters for latent class 3
Constant|    5.67164***     .11968    47.39  .0000     5.43707   5.90622
    HSAT|     .00033        .01181      .03  .9779    -.02283    .02348
 MARRIED|   -1.27033***     .07774   -16.34  .0000   -1.42270  -1.11796
        |Estimated prior probabilities for class membership
   ONE_1|   -1.45501***     .21827    -6.67  .0000   -1.88281  -1.02721
FEMALE_1|     .08828        .27242      .32  .7459    -.44566    .62222
   ONE_2|     .19068        .16191     1.18  .2389    -.12667    .50802
FEMALE_2|     .07895        .17808      .44  .6575    -.27009    .42799
   ONE_3|        0.0    .....(Fixed Parameter).....
FEMALE_3|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

E51.8 Fractional Responses

Papke and Wooldridge (2998) panel data model for a fractional response is

$$E[y_{it}|\ \mathbf{x}_{it},a_i]\ =\ \Phi(\boldsymbol{\beta}'\mathbf{x}_{it} + a_i),\ \ 0 \leq y_{it} \leq 1$$

where a_i is unobserved heterogeneity. The model is completed with the assumption about the heterogeneity,

$$a_i|\ \mathbf{X}_i \ \sim\ \ N[\alpha + \delta'\overline{\mathbf{x}}, \sigma^2]$$

where $\mathbf{X}_i$ is the $T_i{\times}K$ matrix of data on $\mathbf{x}_{it}$ for the T_i periods. After integrating out the heterogeneity, the conditional mean function in terms of the observables is

$$E[y_{it}\mid \mathbf{X}_i] = \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \alpha}{\sqrt{1+\sigma^2}}\right) = \Phi\left(\boldsymbol{\beta}_a'\mathbf{x}_{it} + \boldsymbol{\delta}_a'\overline{\mathbf{x}}_i + \alpha_a\right).$$

The second result provides the useable model in terms of the scaled coefficients. Since the parameter σ^2 is not identified, estimation and inference is based on the scaled coefficients. The model should contain a constant term. Note that if $\boldsymbol{\beta}$ contains a constant to begin with, β_0, the parameter estimated is $(\beta_0 + \alpha)/(1 + \sigma^2)^{1/2}$. We extend the model specification slightly to allow time invariant variables, $\mathbf{z}_i$, so

$$E[y_{it}\mid \mathbf{X}_i, \mathbf{z}_i] = \Phi\left(\boldsymbol{\beta}_a'\mathbf{x}_{it} + \boldsymbol{\delta}_a'\overline{\mathbf{x}}_i + \alpha_a + \gamma_a'\mathbf{z}_i\right).$$

The estimator for this model is obtained with command

FRACTIONAL ; Lhs = dependent variable
; Rhs = independent variables
; Panel (or ; Pds = name) $

The group means of the Rhs variables are added to the model during estimation – the Rhs list should not contain the means. The constant term, *one*, is automatically created, as it is part of the structure of the latent common effect. All of the standard options are available. Average partial effects based on the scaled coefficients are requested with

; Partial effects.

Estimation and computation of the partial effects are developed further in the technical details in Section E51.8.4. Predictions and residuals are retained with

; Keep = name
and **; Res = name**.

E51.8.1 Standard Model Specifications for the Fractional Response Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Partials displays marginal effects.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm (not available for FEM).
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator (not for FEM).
; Rst = list specifies equality and fixed value restrictions (not for FEM).

E51.8.2 Application

To illustrate the estimator, we have constructed the fractional variable *frac* which, for each household in the sample equals the proportion of the total income for the T_i years that is reported in each period, t. Thus, $frac_{it} = hhninc_{it}/(\Sigma_t hhninc_{it})$. For a few households, this equals zero in a few periods. The estimated fractional response model appears below.

```
SAMPLE       ; All $
SETPANEL     ; Group = id ; Pds = ti $
REJECT       ; ti < 7 $
CREATE       ; sums = Group Sums(hhninc, Pds = ti) $
CREATE       ; frac = hhninc/sums $
FRACTIONAL   ; Lhs = frac ; Rhs = one,age,educ,female,hhkids,married
             ; Panel ; Partial Effects $
```

```
Normal exit:  11 iterations. Status=0, F=    64.57008
-----------------------------------------------------------------------------
Fractional Response Model - Panel Data
Dependent variable                 FRAC
Log likelihood function        64.57008
Estimation based on N =   6209, K =  10
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    FRAC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Time Invariant Variables in Conditional Mean
Constant|   -1.07117***      .00133  -805.70  .0000    -1.07378  -1.06856
  FEMALE|-.60292D-04         .00021     -.28  .7756 -.47494D-03  .35435D-03
        |Time Varying Variables in Conditional Mean
     AGE|     .02349***      .00102    23.02  .0000      .02149    .02549
    EDUC|     .04549**       .01884     2.42  .0157      .00858    .08241
  HHKIDS|    -.05524***      .01133    -4.87  .0000     -.07745   -.03303
 MARRIED|     .15575***      .02286     6.81  .0000      .11094    .20056
        |Group Means of Time Varying Variables
     AGE|    -.02345***      .00102   -23.03  .0000     -.02545   -.02145
    EDUC|    -.04565**       .01885    -2.42  .0154     -.08260   -.00871
  HHKIDS|     .05419***      .01111     4.88  .0000      .03241    .07597
 MARRIED|    -.15558***      .02268    -6.86  .0000     -.20003   -.11112
--------+--------------------------------------------------------------------
Partial derivatives of expected value.
--------+--------------------------------------------------------------------
    FRAC|     Partial      Standard            Prob.      95% Confidence
  E[y|x]|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Partial effect for dummy variable is P|1 - P|0.
  FEMALE|-.13552D-04      .4755D-04     -.28  .7756 -.10675D-03  .79649D-04
     AGE|     .00528***      .00023    23.16  .0000      .00483    .00573
    EDUC|     .01023**       .00423     2.42  .0157      .00193    .01852
        |Partial effect for dummy variable is P|1 - P|0.
  HHKIDS|    -.01238***      .00253    -4.89  .0000     -.01735   -.00742
        |Partial effect for dummy variable is P|1 - P|0.
 MARRIED|     .03303***      .00456     7.24  .0000      .02409    .04197
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E51.8.3 Endogenous Explanatory Variables

The possibility of accommodating endogenous variables on the right hand side of the equation is raised in Papke and Wooldridge (2008). Their application involves an endogenous continuous variable. Since the estimator is not based on a likelihood function, the sort of full information maximum likelihood estimator proposed, for example, for the count model with sample selection, or other two part models, will not be appropriate. The authors suggest, instead, the control function approach developed by Rivers and Vuong (1988) and investigated further by Terza, Basu and Rathouz (2008).

Consistent estimation is suggested by adding the residual from a reduced form equation for the endogenous variable to the fractional response model. For example, assume that one of the variables in $\mathbf{x}_{it}$ is $x_{it,e}$, a continuous endogenous variable such as 'spending,' that depends on a $\mathbf{z}_{it}$ that contains at least one variable that is not contained in $\mathbf{x}_{it}$. Then, consistent estimation of the parameters of the fractional response model is achieved by adding v_{it}, the residual in the linear regression of $x_{it,e}$ on $\mathbf{z}$. (Note, it is not suggested that predictions of $x_{it,e}$ be placed in the equation in place of the original data – rather, the residual is *added* to the equation.) Three issues remain:

1. If there is more than one endogenous variable, then a residual is added to the fractional response for each of them. Identification requires that each equation be uniquely identified by its own exogenous variable(s).

2. If the endogenous variables are not continuous, there may be an ambiguity as to how the residual is to be computed. We suggest Chesher and Irish's (1987) generalized residuals. For the single index models that are very likely to be behind the endogenous variables, the generalized residual is the derivative of the log density with respect to the constant term in the model. For a linear model, this would be e_{it}/s^2. For a binary choice (probit) model, this would be the signed inverse Mills ratio. For a count data model, this would be $y_{it} - \lambda_{it}$ where λ_{it} is the conditional mean. And so on.

3. There is no obvious approach suggested for obtaining the appropriate asymptotic covariance matrix for this estimator. The Murphy and Topel (2002) two step approach seems like the natural candidate, but is likely to be cumbersome in the extreme. Papke and Wooldridge mention bootstrapping in passing – this seems like an attractive alternative to the tedious development of a Murphy and Topel estimator.

The following general template could be used to incorporate these steps:

```
SETPANEL      ; Group = the variable ; Pds = ti $
PROCEDURE = FracResp(y1,x,y2,z,model2) $
MODEL2        ; Lhs = y2 ; Rhs = z $
CREATE        ; ey2 = _genres $
FRACTIONAL    ; Lhs = y1 ; Rhs = x, ey2 ; Panel $
ENDPROC $
EXEC          ; Proc = FracResp(y1,x,y2,z,model2) ; Bootstrap = b
              ; Panel ; N = number of bootstrap reps $
```

The following contrived example supposes that work status is endogenous in the fractional response model, and is determined by a probit model for the data generating process. We use a panel bootstrap method to do the estimation. In order to speed up the estimation, we have restricted the sample to groups with seven periods. The program illustrates several features. The procedure uses adjustable parameters, so it can be used with different model specifications. Note one of the parameters in the parameter list is the model command name, '*model2*.' The **EXECUTE** command that calls this procedure below requests *model2* to be a **PROBIT** command. The program also illustrates use of the panel bootstrap – bootstrap replications sample groups of observations defined by the **SETPANEL** command. Finally, the estimator that is constructed is a two step MLE/NLSQ estimator with an endogenous right hand side variable, working, in the second equation. The last command below computes the covariance matrix for the estimator without using the bootstrapping procedure. The **; Maxit = 0** specification just reuses the previous estimates.

```
? Initial preparation of the variables in the model
SAMPLE       ; All $
SETPANEL     ; Group = id ; Pds = ti $
CREATE       ; sums = Group Sums(hhninc, Pds = ti) $
CREATE       ; frac = hhninc/sums $
NAMELIST     ; x = one,age,educ,hhkids $
NAMELIST     ; z = one,age,hsat,public $
? Two step bootstrap estimator
PROC = FracResp(y1,x,y2,z,model2) $
MODEL2       ; Lhs = y2 ; Rhs = z $
CREATE       ; genres = score_fn $
FRACTIONAL   ; Lhs = y1 ; Rhs = x,y2,genres ; Panel $
ENDPROC $
? Set the sample for estimation then estimate model
REJECT       ; ti < 7 $
SETPANEL     ; Group = id ; Pds = ti $
CALC         ; Ran(123457) $
EXEC         ; N = 25 ; Bootstrap = b
             ; Labels = constant,age,educ,kids,working,residual,
                        mean_age,mean_edc,mean_kds,mean_work,mean_res
             ; Proc = FracResp(frac,x,working,z,probit)
             ; Pds = ti  $
? Estimate the model without the bootstrap iterations
FRACTIONAL   ; Lhs = frac ; Rhs = x,working,genres
             ; Panel ; Maxit = 0 $
```

```
----------------------------------------------------------------------------
Results of bootstrap estimation of model.
Model has been reestimated      25 times.
Coefficients shown below are the original
model estimates based on the full sample.
Bootstrap samples have  887 observations.
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
BootStrp|  Coefficient       Error       z     |z|>Z*        Interval
--------+-------------------------------------------------------------------
CONSTANT|   -1.07102***      .00279  -384.01  .0000    -1.07649  -1.06556
     AGE|     .02636***      .00134    19.64  .0000      .02373    .02899
    EDUC|     .02370         .01586     1.49  .1351     -.00739    .05479
    KIDS|    -.03033***      .00920    -3.30  .0010     -.04836   -.01230
 WORKING|     .33992***      .07040     4.83  .0000      .20193    .47790
RESIDUAL|    -.12308***      .04276    -2.88  .0040     -.20690   -.03927
MEAN_AGE|    -.02632***      .00134   -19.59  .0000     -.02896   -.02369
MEAN_EDC|    -.02378         .01586    -1.50  .1339     -.05487    .00732
MEAN_KDS|     .02868***      .00910     3.15  .0016      .01084    .04651
MEAN_WOR|    -.34129***      .06996    -4.88  .0000     -.47841   -.20417
MEAN_RES|     .12460***      .04248     2.93  .0034      .04133    .20787
--------+-------------------------------------------------------------------
(Estimated without bootstrapped standard errors)
--------+-------------------------------------------------------------------
        |Time Invariant Variables in Conditional Mean
Constant|   -1.07126***      .00239  -448.80  .0000    -1.07594  -1.06658
        |Time Varying Variables in Conditional Mean
     AGE|     .02634***      .00122    21.57  .0000      .02394    .02873
    EDUC|     .02335         .01698     1.38  .1691     -.00993    .05662
  HHKIDS|    -.03010***      .01124    -2.68  .0074     -.05213   -.00806
 WORKING|     .33236***      .07141     4.65  .0000      .19239    .47233
  GENRES|    -.11875***      .04112    -2.89  .0039     -.19935   -.03815
        |Group Means of Time Varying Variables
     AGE|    -.02629***      .00121   -21.65  .0000     -.02868   -.02391
    EDUC|    -.02342         .01699    -1.38  .1680     -.05672    .00988
  HHKIDS|     .02844**       .01104     2.58  .0100      .00680    .05009
 WORKING|    -.33352***      .07079    -4.71  .0000     -.47227   -.19477
  GENRES|     .12015***      .04072     2.95  .0032      .04034    .19996
--------+-------------------------------------------------------------------
```

E51.8.4 Technical Details

Full technical details of this model are given in Papke and Wooldridge (2008). We will provide a sketch of the main results here. The structure of the model is

$$E[y_{it}|\,\mathbf{x}_{it},a_i] \;=\; \Phi(\boldsymbol{\beta}'\mathbf{x}_{it} + a_i),\;\; 0 \leq y_{it} \leq 1$$

where a_i is unobserved heterogeneity with projection onto the group means of the data approximated with,

$$a_i|\,\mathbf{X}_i \sim \;\; N[\alpha+\delta'\overline{\mathbf{x}},\sigma^2]\,.$$

The reduced form is

$$E[y_{it}\,|\,\mathbf{X}_i] = \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_{it}+\boldsymbol{\delta}'\overline{\mathbf{x}}_i+\alpha}{\sqrt{1+\sigma^2}}\right) = \Phi\left(\boldsymbol{\beta}_a'\mathbf{x}_{it}+\boldsymbol{\delta}_a'\overline{\mathbf{x}}_i+\alpha_a\right).$$

Objects of estimation are β_a, δ_a and α_a. We estimate average partial effects as the partial derivatives of $E[y_{it}|X_i]$ and use the delta method to obtain the asymptotic standard errors for these. The average is taken over all the sample observations.

The estimator is based on nonlinear multivariate least squares. First step estimates of $(\beta_a,\delta_a,\alpha_a)$ are obtained by a pooled grouped probit estimator, which maximizes the log likelihood

$$\log L = \sum_{i=1}^{n}\sum_{t=1}^{T_i} \begin{matrix} y_{it}\log\Phi(\boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \alpha) + \\ (1-y_{it})\log[1-\Phi(\boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \alpha)] \end{matrix}$$

This initial estimator of the parameters, $\boldsymbol{\gamma}^0$, is used in construction of the weighting matrix for generalized least squares. The criterion function for estimation is, then

$$\text{Min}_{\gamma}\, S(\boldsymbol{\gamma}) = \Sigma_i\, (\mathbf{y}_i - E[\mathbf{y}_i|\mathbf{X}_i])'\, [\mathbf{V}(\mathbf{X}_i,\boldsymbol{\gamma}^0]^{-1}\, (\mathbf{y}_i - E[\mathbf{y}_i|\mathbf{X}_i])$$

where $\mathbf{y}_i = (y_{i1},y_{i2},\ldots,y_{i,Ti})'$ and the T_i conditional means are stacked in $E[\mathbf{y}_i|\mathbf{X}_i]$. The asymptotic covariance matrix for the estimator is computed using (3.8) in Papke and Wooldridge (2008). Partial effects are based on their (3.11) and (3.12), for the entire sample.

E52: Linear Sample Selection Models

E52.1 Introduction

Many variants of the 'sample selection' model can be estimated with *LIMDEP*. (See Heckman (1979), Maddala (1983) and Greene (2011) for further discussion.) Most of them share the following structure: A specified model, denoted **A**, applies to the underlying data. However, the observed data are not sampled randomly from this population. Rather, a related variable z^* is such that an observation is drawn from **A** only when z^* crosses some threshold. If the observed data are treated as having been randomly sampled from **A** instead of from the subpopulation of **A** associated with the 'selected' values of z^*, potentially serious biases result. The general solution to the selectivity problem relies upon an auxiliary model of the process generating z^*. Information about this process is incorporated in the estimation of **A**.

Several of the forms of this model which can be estimated with *LIMDEP* depart from Heckman's now canonical form, a linear regression with a binary probit selection criterion model:

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

$$z^* = \boldsymbol{\alpha}'\mathbf{w} + u,$$

$$\varepsilon,u \sim N[0,0,\sigma_\varepsilon^2, \sigma_u^2, \rho].$$

A bivariate classical (seemingly unrelated) regressions model applies to the structural equations. The standard deviations are σ_ε and σ_u, and the covariance is $\rho\sigma_\varepsilon\sigma_u$. If the data were randomly sampled from this bivariate population, the parameters could be estimated by least squares, or GLS combining the two equations. However, z^* is not observed. Its observed counterpart is z, which is determined by

$$z = 1 \text{ if } z^* > 0$$

and

$$z = 0 \text{ if } z^* \le 0.$$

Values of y and **x** are only observed when z equals one. The essential feature of the model is that under the sampling rule, $E[y|\mathbf{x},z=1]$ is not a linear regression in **x**, or **x** and z. The development below presents estimators for the class of essentially nonlinear models that emerge from this specification.

The basic command structure for the models described in this chapter is

```
PROBIT     ; Lhs = variable z ; Rhs = variables in w ; Hold $
SELECT     ; Lhs = variable y ; Rhs = variables in x $
```

Note that two commands are required for estimation of the sample selection model, one for each structural equation.

This is the simplest form of this model. In this chapter, we consider estimation by two step least squares and maximum likelihood. In addition, we provide a large number of different forms and associated estimation techniques. Because there are so many different models considered here, this chapter will depart from our usual format. Rather than gather material by function (theoretical background, command, application, mathematical details), we will gather the material on sample selection by model. This chapter will develop the two step and ML estimators for sample selection models using cross section data. Chapter E53 presents panel data formulations of the sample selection models. These two chapters are concerned with variations on the binary selection mechanism with linear primary equation. Chapters E54 and E55 develop a variety of models that involve different types of equations and different types of selection mechanisms, respectively.

E52.2 Regression Model with Sample Selection

The models described in this section are based on a dichotomous selection mechanism. Heckman's approach to estimation is based on the following observations: In the selected sample,

$$\begin{aligned} \mathrm{E}[y_i|\mathbf{x}_i, \text{in sample}] &= \mathrm{E}[y_i \mid \mathbf{x}_i, z_i = 1] \\ &= \mathrm{E}[y_i \mid \mathbf{x}_i, \boldsymbol{\alpha}'\mathbf{w}_i + u_i > 0] \\ &= \boldsymbol{\beta}'\mathbf{x}_i + \mathrm{E}[\varepsilon_i \mid u_i > -\boldsymbol{\alpha}'\mathbf{w}_i] \\ &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho\sigma_\varepsilon\sigma_u)\{\phi(-\boldsymbol{\alpha}'\mathbf{w}_i)/[1 - \Phi(-\boldsymbol{\alpha}'\mathbf{w}_i)]\} \\ &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho\sigma_\varepsilon\sigma_u)[\phi(\boldsymbol{\alpha}'\mathbf{w}_i)/\Phi(\boldsymbol{\alpha}'\mathbf{w}_i)]. \end{aligned}$$

Given the structure of the model and the nature of the observed data, σ_u cannot be estimated, so it is normalized to 1.0. (We observe the same values of z_i regardless of the value of σ_u.) Then,

$$\begin{aligned} \mathrm{E}[y_i \mid \mathbf{x}_i, \text{in sample}] &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho\sigma_\varepsilon)\lambda_i \\ &= \boldsymbol{\beta}'\mathbf{x}_i + \theta\lambda_i\,. \end{aligned}$$

There are some subtle ambiguities in the received applications of this model. First, it is unclear whether the index function, $\boldsymbol{\beta}'\mathbf{x}_i$, or the conditional mean is really the function of interest. If the model is to be used to analyze the behavior of the selected group, then it is the latter. If not, it is unclear. The index function would be of interest if attention were to be applied to the entire population, rather than those expected to be selected. This is application specific. Second, the partial effects in this model are complicated as well. For the moment, assume that $\mathbf{x}_i$ and $\mathbf{w}_i$ are the same variables. Then,

$$\frac{\partial E[y_i \mid \mathbf{x}_i, z_i = 1]}{\partial \mathbf{x}_i} = \boldsymbol{\beta} + \theta(-\lambda_i\boldsymbol{\alpha}'\mathbf{x}_i - \lambda_i^2)\boldsymbol{\alpha}$$

For any variable x_k which appears in both the selection equation (for z_i) and the regression equation, the partial effect consists of both the direct part (β_k) and the indirect part, which is of opposite sign – the term in parentheses is always negative; $\theta(-\lambda_i\boldsymbol{\alpha}'\mathbf{x}_i - \lambda_i^2)\alpha_k$. It is not obvious which part will dominate. Most applications have at least some variables that appear in both equations, so this is an important consideration. Note also that variables which do not appear in the index function still affect the conditional mean function through their affect on the inverse Mills ratio (the 'selection variable'). (We note the risk of conflict in the notation used here for the selection term, λ_i, and the loglinear term in the conditional mean functions of the generalized linear models in the preceding chapters. There is no relationship between the two. The two uses of 'lambda' are so common in the received literature as to have become part of the common parlance and as such, the risk of ambiguity is worse if we try to change the notation used here for clarity.)

LIMDEP contains three estimators for this model, Heckman's two step (or 'Heckit') estimator, full information maximum likelihood, and two step maximum likelihood (which is, more or less, a limited information maximum likelihood estimator). The first is presented in Section E52.2. The MLEs are presented in the Sections E52.2.3 and E52.3.3. The latter develops a model with heteroscedasticity, and uses the limited information maximum likelihood estimator.

E52.2.1 Defining Limit Observations and Control Observations

Limit Observations

The default specification for the selection model is to select on the value one for z. If you wish, instead, to select on zero use

; Limits

to define observations with $z = 1$ as 'limit' observations in the commands. In this case, λ is computed using the appropriate formula,

$$\lambda = -\phi(\boldsymbol{\alpha}'\mathbf{w})/[1 - \Phi(\boldsymbol{\alpha}'\mathbf{w})]$$

instead of $\phi(\boldsymbol{\alpha}'\mathbf{w})/\Phi(\boldsymbol{\alpha}'\mathbf{w})$. An application is suggested by the mover stayer model discussed in Section E56.2. This brings no other changes in the model or results. However, if you select on zero, *LIMDEP* saves *sigma*0 and *bsr*0 instead *sigma*1 and *bsr*1 when it retains the results. This is noted again below. Once again, these have relevance to the mover stayer model presented later.

Control Observations

In some experimental situations, some observations might actually be randomly selected from the full population, not the selected one. These might be controls, included by the experimental design. For any such observation, the correct value of λ to include in the equation is zero. You can indicate that there are control observations in your sample with a binary indicator included as a second Lhs variable in the command. Observations for which the variable is zero are given a value of 0.0 for λ; observations for which it is one will get the appropriate value computed by the formulas given earlier.

NOTE: This option is not used with the maximum likelihood estimators.

This variable, d_i must be coded 0/1. When it is present, the value of λ_i that is inserted in the equation is $d_i\lambda_i^*$, where λ_i^* is computed as prescribed earlier. As such, if you do not use a binary variable, the results may be seriously distorted.

E52.2.2 Two Step Estimation of the Standard Model

Heckman's two step, or 'Heckit' estimation method, is based on the method of moments. It is a consistent, but not efficient two step estimator.

Step 1. Use a probit model for z_i to estimate $\boldsymbol{\alpha}$.
For each observation, compute $\lambda_i = \phi(\boldsymbol{\alpha}'\mathbf{w}_i)/\Phi(\boldsymbol{\alpha}'\mathbf{w}_i)$ using the probit coefficients.

Step 2. Linearly regress y_i on $\mathbf{x}_i$ and λ_i to estimate $\boldsymbol{\beta}$ and $\theta = \rho\sigma_\varepsilon$.
Adjust the standard errors and the estimate of σ_ε^2, which is inconsistent.

The corrected asymptotic covariance matrix for the two step estimator, $(\mathbf{b},c)$, is

$$\text{Asy.Var}[\mathbf{b},c] = \sigma_\varepsilon^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}[\mathbf{X}^{*\prime}(\mathbf{I} - \rho^2\boldsymbol{\Delta})\mathbf{X}^* + \rho^2(\mathbf{X}^{*\prime}\boldsymbol{\Delta}\mathbf{W})\boldsymbol{\Sigma}(\mathbf{W}'\boldsymbol{\Delta}\mathbf{X}^*)](\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$$

where $\mathbf{X}^* = [\mathbf{X} : \boldsymbol{\lambda}]$,

$\boldsymbol{\Delta} = \text{diag}[\boldsymbol{\delta}]$,

$\boldsymbol{\delta}_i = -\lambda_i(\boldsymbol{\alpha}'\mathbf{w}_i + \lambda_i)$ $(-1 \le \delta_i \le 0)$,

and $\boldsymbol{\Sigma}$ = asymptotic covariance matrix for the estimator of $\boldsymbol{\alpha}$.

A consistent estimator of σ_ε^2 is $\hat{\sigma}_\varepsilon^2 = \mathbf{e}'\mathbf{e}/n - \hat{\theta}^2\bar{\hat{\delta}}$. The remaining parameters are estimated using the least squares coefficients. The computations used in the estimation procedure are those discussed in Heckman (1979) and in Greene (1981).

NOTE: (This is one of our frequently asked questions.) *LIMDEP* always computes the corrected asymptotic covariance matrix, for all variants of selection models in all model frameworks.

The estimator of the correlation coefficient, ρ, is $\text{sign}(\hat{\theta})\sqrt{\hat{\theta}^2/\hat{\sigma}_\varepsilon^2}$. This is the ratio of a regression coefficient (the coefficient on λ_i) and the variance estimator above. Note that it is not a sample moment estimator of the correlation of two variables. This ratio is not guaranteed to be between -1 and +1. (See Greene (1981), which is about this result.) But, note also that an estimate of ρ is needed to compute the asymptotic covariance matrix above, so this is a potential complication. When this occurs, *LIMDEP* uses either +1 or -1, and continues. We emphasize, this is not an error, nor is it a program failure. It is a characteristic of the data. (It may signal some problems with the model.) When this condition occurs, the model results will contain the diagnostic

```
Estimated correlation is outside the range -1 < r < 1. Using 1.0
```

If the estimate of ρ is invalid, so that the polar value must be used, the corrected standard errors can be negative. If this happens, a warning is given and the OLS standard errors for the estimates are used instead. This condition is specific to the two step regression estimators. The maximum likelihood estimators discussed below force the coefficient to lie in the unit interval – ρ is estimated directly, not by the method of moments.

To estimate this model with *LIMDEP*, it is necessary first to estimate the probit model, then request the selection model. The pair of commands is

PROBIT ; Lhs = name of z ; Rhs = list for w ; Hold results $
SELECT ; Lhs = name of y ; Rhs = list for x $

For this simplest case, **; Hold ...** may be abbreviated to **; Hold**. All of the earlier discussion for the probit model applies. This application differs only in the fact the **; Hold** requests that the model specification and results be saved to be used later. Otherwise, they disappear with the next model command. The **PROBIT** command is exactly as described in Chapter E26. The selection model is completely self contained. You do not need to compute or save λ_i.

Standard Model Specifications for the Sample Selection Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

None

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.

Predictions for this model are computed according to the formula given above for the conditional mean, including λ_i. The other variables listed are the residual computed as usual, $\boldsymbol{\beta}'\mathbf{x}$, and λ_i. The optional specifications

; Keep = name
; Res = name
; List

are the same as for the linear regression model, **REGRESS**.

You may also specify weights for the regression with

; Wts = weighting variable

LIMDEP recomputes the scale factor for the weights so that the weights used in estimation sum to the number of observations in the selected sample, not the number in the full data set.

In most cases, it is not necessary for you to compute the selection variable, λ; this is taken care of internally. Nonetheless, you may have occasion to use this variable for some other purpose. If so, change the **; Hold** specification in the **PROBIT** command to

; Hold (IMR = name)

(IMR stands for Inverse Mill's Ratio.) This places λ_i in your data array as variable *name* and you can use it for any other purpose or write it to a data file for later use.

NOTE: The *imr* variable is computed as $\lambda = \phi/\Phi$ if $z = 1$ and $\lambda = -\phi/(1-\Phi)$ if $z = 0$. Note that the data saved are determined by the current sample. This means that if you have partitioned the sample before giving the **PROBIT** command, after it is executed, the data array may have some cells which are undefined. The variable is only computed for observations used to fit the probit model.

The retrievable results saved by the estimator are

Matrices:	*b*	contains $(\boldsymbol{\beta},\theta)$
	v*arb*	contains the corrected VC matrix
	*bsr*1	contains $[\boldsymbol{\beta},\sigma_\varepsilon,\rho]$ (Used with the mover stayer model.)

Scalars:	*sy* and *ybar*	for the dependent variable
	x	= $\sigma_{,}$ (this is also saved in *sigma*1)
	ssqrd	= $\mathbf{e}'\mathbf{e}/N$
	sumsqdev	= $\mathbf{e}'\mathbf{e}$
	rsqrd	= R^2 in the linear regression
	rho	= ρ
	degfrdm	= $N - K - 1$
	kreg	= $K + 1$
	nreg	= number of observations in selected sample

Last Model: *b_variables*, *theta*

Last Function: Conditional mean function, $\beta'x + \theta\lambda_i$

The conditional mean function given above is used in the **SIMULATE** and **PARTIAL EFFECTS** commands when computing predictions and partial effects.

The estimate of ρ needed to compute the appropriate standard errors for the estimates is the square root of

$$\hat{\rho}^2 = \hat{\theta}^2 / \left(\mathbf{e}'\mathbf{e}/N \ - \ \hat{\theta}^2\bar{\delta}\right).$$

The output for the sample selection model consists of a short summary similar to the results for a least squares regression, with some additional information about the selection model.

To illustrate the estimator, we use one of the most familiar data sets used in the pedagogical segment of this literature, Mroz's (1987) female labor supply data. The data are described in detail in Section E45.7. The variables in the data set are

lfp	=	a dummy variable = 1 if woman worked in 1975, else 0
whrs	=	wife's hours of work in 1975
kl6	=	number of children less than 6 years old in household
k618	=	number of children between ages 6 and 18 in household
wa	=	wife's age
we	=	wife's educational attainment, in years
ww	=	wife's average hourly earnings, in 1975 dollars
rpwg	=	wife's wage reported at the time of the 1976 interview
hhrs	=	husband's hours worked in 1975
ha	=	husband's age
he	=	husband's educational attainment, in years
hw	=	husband's wage, in 1975 dollars
faminc	=	family income, in 1975 dollars
mtr	=	marginal tax rate facing the wife
wmed	=	wife's mother's educational attainment, in years
wfed	=	wife's father's educational attainment, in years
un	=	unemployment rate in county of residence, in percentage points
cit	=	dummy variable = 1 if live in large city (SMSA), else 0
ax	=	actual years of wife's previous labor market experience
prin	=	*faminc* - (*whrs***ww*) = wife's property income

We will use the two step estimator to build the selection model:

$$lfp \ = \ f(kl6,\ k618,\ prin,\ un,\ hinc)$$

$$winc \ = \ g(wa,\ wa^2,\ we,\ cit,\ ax)$$

where *hinc* = husband's income = $hhrs \times hw$, and *winc* = wife's wage income = $ww \times whrs$. The estimates also compare the corrected results to uncorrected ordinary least squares.

The commands are:

CREATE ; prin = faminc - ww*whrs $
CREATE ; hinc = hw*hhrs $
CREATE ; winc = ww*whrs $
CREATE ; wasq = wa*wa $
NAMELIST ; w = one,kl6,k618,prin,un,hinc $
NAMELIST ; x = one,wa,wasq,we,cit,ax $
PROBIT ; Lhs = lfp ; Rhs = w ; Hold $
SELECTION ; Lhs = winc ; Rhs = x ; Partial Effects $
REGRESS ; Lhs = winc ; Rhs = x $

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    LFP
Log likelihood function      -491.93905
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LFP|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .60344***      .16379     3.68  .0002      .28242    .92446
     KL6|    -.54906***      .09590    -5.73  .0000     -.73702   -.36111
    K618|     .01988         .03555      .56  .5760     -.04980    .08956
    PRIN|-.15882D-04**    .7770D-05    -2.04  .0410 -.31111D-04  -.65219D-06
      UN|    -.00979         .01509     -.65  .5165     -.03936    .01978
    HINC| .46551D-05      .9747D-05      .48  .6329 -.14449D-04  .23759D-04
--------+--------------------------------------------------------------------
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on LFP                     |
| Selection rule is: Observations with LFP      =  1         |
| Results of selection:                                      |
|                  Data points     Sum of weights            |
| Data set              753              753.0               |
| Selected sample       428              428.0               |
+------------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model...........................
Two step     least squares regression ............
LHS=WINC     Mean                 =     5192.94004
             Standard deviation   =     4301.55079
             Number of observs.   =            428
Model size   Parameters           =              7
             Degrees of freedom   =            421
Residuals    Sum of squares       =    .614172E+10
             Standard error of e  =     3819.47630
Fit          R-squared            =         .20973
             Adjusted R-squared   =         .19847
Model test   F[  6,   421] (prob) =    18.6(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Standard error corrected for selection..4309.55979
Correlation of disturbance in regression
and Selection Criterion (Rho)     =        -.60967
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    WINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -7379.04       5831.829    -1.27  .2058   -18809.21   4051.14
      WA|    270.821       266.8043     1.02  .3101    -252.106   793.748
    WASQ|   -3.92282        3.10679    -1.26  .2067   -10.01201   2.16638
      WE|    560.264***    84.95503     6.59  .0000     393.755   726.773
     CIT|    671.261*      397.7575     1.69  .0915    -108.329  1450.852
      AX|    196.088***    26.77697     7.32  .0000     143.606   248.570
  LAMBDA|   -2627.39**     1272.297    -2.07  .0389    -5121.05   -133.73
--------+--------------------------------------------------------------------
Ordinary     least squares regression ............
--------+--------------------------------------------------------------------
Constant|   -11587.8***    3857.324    -3.00  .0027    -19148.0   -4027.6
      WA|    394.398**     178.6805     2.21  .0273      44.191   744.606
    WASQ|   -5.33362***     2.06127    -2.59  .0097    -9.37363  -1.29361
      WE|    429.404***    59.27814     7.24  .0000     313.221   545.587
     CIT|    202.857       279.6910      .73  .4683    -345.327   751.041
      AX|    221.461***    17.38617    12.74  .0000     187.385   255.537
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Partial effects of E[y] = Xb + c*L   with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Means for direct effects are for selected
observations.  Means for indirect effects
are the  full sample used for the probit.
If a variable appears in both Xb and in L
the second effect shown in the table is
b + c*dL/dx = direct+indirect.
--------+--------------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
    WINC|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Direct effects in the regression
      WA|    270.821       266.8043     1.02  .3101    -252.106   793.748
    WASQ|   -3.92282        3.10679    -1.26  .2067   -10.01201   2.16638
      WE|    560.264***    84.95503     6.59  .0000     393.755   726.773
     CIT|    671.261*      397.7575     1.69  .0915    -108.329  1450.852
      AX|    196.088***    26.77697     7.32  .0000     143.606   248.570
        |Indirect effects in LAMBDA (means are for all obs.)
     KL6|   -861.881*      444.3537    -1.94  .0524   -1732.798     9.036
    K618|    31.2102       64.78761      .48  .6300    -95.7712  158.1916
    PRIN|    -.02493       29.21305      .00  .9993   -57.28145  57.23159
      UN|   -15.3645       38.50192     -.40  .6899    -90.8269   60.0979
    HINC|     .00731       29.21252      .00  .9998   -57.24818  57.26279
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E52.2.3 Maximum Likelihood Estimation

The full log likelihood for the sample selection model is built up from

Prob(Selection) × density|selection for selected observations

and Prob(Nonselection) for nonselected observations.

Combining the various parts, this produces the following log likelihood function for the sample selection model:

$$\log L \;=\; \sum\nolimits_{z=1} \log\left[\frac{\exp\left(-(1/2)\varepsilon_i^2/\sigma_\varepsilon^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(\frac{\rho\varepsilon_i/\sigma_e+\boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho^2}}\right)\right] \;+\; \sum\nolimits_{z=0} \log\Phi(-\boldsymbol{\alpha}'\mathbf{w}_i)$$

where $\varepsilon_i \;=\; y_i - \boldsymbol{\beta}'\mathbf{x}_i$.

NOTE: There are two (only apparent) inconsistencies between this and the statement of the counterpart in Maddala’s (1983) widely used reference on his page 266. First, he appears to multiply the first term by $1/\Phi(\boldsymbol{\alpha}'\mathbf{w}_i)$. The reason for this is that his result gives the conditional density for the selected observations whereas the likelihood function is built up from the unconditional densities for the entire sample. Thus, the log likelihood function above results from the construction,

density = Prob(selected) × Maddala’s result

which gives our result. The second inconsistency appears to be the sign on the residual term in the second function above, which is -ρ... in Maddala. The reason for this is the inexplicable negative sign on ε_1 in his statement (9.17) as opposed to our positive sign on u in our statement of the model. Since ε is normally (symmetrically) distributed, the formulations are equivalent. There is no obvious reason for the sign reversal in Maddala’s treatment – at this juncture, the literature has settled on the slightly simpler formulation adopted herein.

Maximum likelihood estimates of the model parameters can be obtained by adding the specification

; MLE

to the **SELECTION** command described earlier. This activates a number of the standard optional features, as shown in the revised listing below. It is still necessary to precede this estimator with the probit model in order to provide starting values for the MLE. The full set of output for the earlier methodology is produced as well. The final values from the Heckman procedure are used as the starting values for the maximum likelihood procedure. The method is that of BFGS.

NOTE: Although this model computes an estimate of α, it does *not* replace the estimates that have been retained with the **; Hold** instruction in the preceding **PROBIT** command.

Standard Model Specifications for the Sample Selection MLE

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps ancillary parameter σ with main parameter **β** vector in *b*.
; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec requests computation of the cluster form of corrected covariance estimator. (includes **; Stratum** as well for stratified and clustered data sets).
; Robust requests a sandwich estimator or robust VC for TSCS and some discrete choice models (uses **; Cluster = 1**).

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg [= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb [= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.
; Fill fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Once the model has been estimated by maximum likelihood, all remaining computations are the same as for the earlier treatment with estimation by two step least squares. But, in this case, fewer scalars are saved. In particular, only *ybar*, *sy*, *nreg*, as before, plus *logl*, *sigma*1, *rho*, and *varrho* for the MLE are added. The matrix *bsr*1 is as described above and *b* and *varb* contain $\boldsymbol{\beta}, \sigma_\varepsilon$, and ρ.

NOTE: When the parameters are estimated by maximum likelihood, there is no 'λ_i' variable in the equation. (See the technical details below.) Therefore, there is one fewer parameter in *b* for the regression.

TECHNICAL NOTE: During the optimization process, the parameter ρ is replaced by the transformation of $\tau = \log[(1+\rho)/(1-\rho)]$, so that $\rho = [\exp(\tau)-1]/[\exp(\tau)+1]$. By this reparameterization, τ may be estimated as an unrestricted parameter – its range is unbounded. This circumvents problems of ρ straying outside the allowable range of (-1,+1).

The model estimated earlier is shown below using the maximum likelihood approach, instead.

```
CREATE      ; prin = faminc - ww*whrs $
CREATE      ; hinc = hw*hhrs $
CREATE      ; winc = ww*whrs $
CREATE      ; wasq  = wa*wa $
NAMELIST    ; w = one,kl6,k618,prin,un,hinc $
NAMELIST    ; x  = one,wa,wasq,we,cit,ax $
PROBIT      ; Lhs = lfp ; Rhs = w ; Hold $
SELECTION   ; Lhs = winc ; Rhs = x ; Partial Effects ; MLE $
```

Two sets of partial effects are reported. The first set are the partial effects based on the maximum likelihood estimator. The second set are those estimated earlier based on the two step estimator.

```
-----------------------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                  WINC
Log likelihood function       -4630.35920
FIRST  6 estimates are probit equation.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    WINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Selection (probit) equation for LFP
Constant|      .64244***      .16261     3.95  .0001       .32374     .96115
     KL6|     -.57716***      .09854    -5.86  .0000     -.77030    -.38402
    K618|      .00659         .03532      .19  .8520     -.06263     .07581
    PRIN| -.14511D-04**     .6667D-05   -2.18  .0295 -.27578D-04  -.14427D-05
      UN|     -.01380         .01487     -.93  .3535     -.04295     .01535
    HINC|  .40862D-05       .9065D-05     .45  .6522 -.13681D-04  .21853D-04
        | Corrected regression, Regime 1
Constant|    -8801.21       5833.866    -1.51  .1314   -20235.37    2632.96
      WA|     304.939       266.2113     1.15  .2520    -216.826    826.703
    WASQ|    -4.29431        3.10729    -1.38  .1670   -10.38449    1.79587
      WE|     538.907***    81.09275     6.65  .0000     379.968    697.846
     CIT|     614.533       468.8614     1.31  .1900    -304.419   1533.484
      AX|     200.737***    28.25734     7.10  .0000     145.353    256.120
SIGMA(1)|     3950.06***    222.7821    17.73  .0000     3513.41    4386.70
RHO(1,2)|     -.31471         .22337    -1.41  .1589     -.75250     .12309
```

```
--------+-------------------------------------------------------------
 (Partial effects, ML)
--------+-------------------------------------------------------------
        |     Partial      Standard            Prob.      95% Confidence
    WINC|      Effect        Error       z     |z|>Z*         Interval
--------+-------------------------------------------------------------
        | Direct effects in the regression
      WA|     304.939       266.2113     1.15  .2520    -216.826   826.703
    WASQ|    -4.29431        3.10729    -1.38  .1670   -10.38449   1.79587
      WE|     538.907***    81.09275     6.65  .0000     379.968   697.846
     CIT|     614.533       468.8614     1.31  .1900    -304.419  1533.484
      AX|     200.737***    28.25734     7.10  .0000     145.353   256.120
        | Indirect effects in LAMBDA (means are for all obs.)
     KL6|    -428.987       332.6624    -1.29  .1972   -1080.994   223.019
    K618|     4.89735       29.82765      .16  .8696   -53.56376  63.35847
    PRIN|    -.01079        13.88839      .00  .9994   -27.23152  27.20995
      UN|    -10.2558       19.39879     -.53  .5970    -48.2767   27.7651
    HINC|      .00304       13.88739      .00  .9998   -27.21574  27.22182
--------+-------------------------------------------------------------
 (Partial effects, two step)
--------+-------------------------------------------------------------
        | Direct effects in the regression
      WA|     270.821       266.8043     1.02  .3101    -252.106   793.748
    WASQ|    -3.92282        3.10679    -1.26  .2067   -10.01201   2.16638
      WE|     560.264***    84.95503     6.59  .0000     393.755   726.773
     CIT|     671.261*      397.7575     1.69  .0915    -108.329  1450.852
      AX|     196.088***    26.77697     7.32  .0000     143.606   248.570
        | Indirect effects in LAMBDA (means are for all obs.)
     KL6|    -861.881*      444.3537    -1.94  .0524   -1732.798     9.036
    K618|     31.2102       64.78761      .48  .6300    -95.7712  158.1916
    PRIN|    -.02493        29.21305      .00  .9993   -57.28145  57.23159
      UN|    -15.3645       38.50192     -.40  .6899    -90.8269   60.0979
    HINC|      .00731       29.21252      .00  .9998   -57.24818  57.26279
```

E52.2.4 A Selection Model with Heteroscedasticity

We extend the sample selection model to incorporate heteroscedasticity in the regression variance with the usual loglinear formulation,

$$\sigma_{\varepsilon i} = \sigma_\varepsilon \exp(\boldsymbol{\gamma}'\mathbf{v}_i).$$

This adds a considerable complication to the model. The full structure becomes

$$\begin{aligned} y_i &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \\ z_i^* &= \boldsymbol{\alpha}'\mathbf{w}_i + u_i, \\ \varepsilon_i, u_i &\sim N[0,0,\sigma_{\varepsilon i}^2, 1, \rho], \\ z &= 1 \text{ if } z^* > 0 \text{ and } z = 0 \text{ if } z^* \le 0, \end{aligned}$$

Values of y_i and $\mathbf{x}_i$ are only observed when z_i equals one. The major complication arises because

$$\begin{aligned} E[y_i \mid \mathbf{x}_i, \text{in sample}] &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho\sigma_{\varepsilon i})\lambda_i \\ &= \boldsymbol{\beta}'\mathbf{x}_i + \theta_i\lambda_i\,. \end{aligned}$$

Note that the heterogeneity in the variance now shows up in the mean. The interesting effects in this model now come in three parts. As we did earlier, assume for the moment that all three data vectors in the model are the same. Then,

$$\frac{\partial E[y_i \mid \mathbf{x}_i, z_i = 1]}{\partial \mathbf{x}_i} = \boldsymbol{\beta} + [\theta_i(-\lambda_i\boldsymbol{\alpha}'\mathbf{x}_i - \lambda_i^2)[\boldsymbol{\alpha} + [\sigma_{\varepsilon i}\lambda_i\rho]\boldsymbol{\gamma}.$$

It is not unlikely that a variable would appear in all three parts, so the marginal effect in this model is extremely complicated. The example below and the technical details present further details.

The preceding implies that conventional least squares based estimation, such as Heckman's estimator, will no longer be consistent. In order to use Heckman's approach, one would require a consistent estimator of $\boldsymbol{\gamma}$ before the least squares step, and it is unclear where that would come from. *LIMDEP* uses a two step, maximum likelihood estimator for this model. The command for this model is

PROBIT ; Lhs = z ; Rhs = variables in w ; Hold $
SELECT ; Lhs = y ; Rhs = variables in x ; Hfn = variables in v $

Do not include *one* in the Hfn list. The constant term in the variance model is already included (implicitly) as $\log\sigma_\varepsilon$, so if you include one of your own, the model will become inestimable.

Estimates of the heteroscedastic model are obtained in three steps. First, the two step least squares estimator is obtained ignoring the heteroscedasticity. Second, the full maximum likelihood estimator is obtained, again ignoring the heteroscedasticity. This is done to obtain the starting values for the parameters, under the assumption that $\boldsymbol{\gamma} = \mathbf{0}$. Finally, the maximum likelihood estimates for the heteroscedasticity model are obtained, allowing $\boldsymbol{\gamma}$ to be unrestricted. The parameter vector in this final model is $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho]$. The estimator of α is not recomputed – this is a limited information maximum likelihood estimator.

Other optional features for this model are the same as for the maximum likelihood estimator of the model with homoscedastic disturbances. (The list of standard model specifications is identical, so it is not repeated here.) If you provide starting values or impose constraints, use this arrangement of the parameters. Starting values are computed in two steps for this model. First, the sample selection model is computed using Heckman's method and ignoring the heteroscedasticity. Second, the MLE, once again ignoring the heteroscedasticity is computed, to sharpen the starting values of ρ and σ. Then, the starting values for $\boldsymbol{\beta}$, σ, and ρ from the MLE and a vector of zeros for γ are used as the start values for this estimator. Therefore, your model results for this model will contain all three sets of estimates.

The estimation results retained are

Matrices:	*b*	= $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho]$ (all parameters are retained)
	varb	= the full asymptotic covariance matrix
Scalars:	*s*	= estimate of σ
	rho	= estimate of ρ
	varrho	= estimated asymptotic variance for estimated ρ
	ybar, sy	= descriptive statistics for dependent variable
	nreg	= number of observations in selected sample
	logl	= log likelihood
	exitcode	

To illustrate the model, we will layer heteroscedasticity on the earnings equation based on our previous application. The probit equation is omitted from the results below, as it was estimated earlier.) We note, as happens frequently in models with heteroscedasticity, the parameter estimates differ from those in the model with homoscedasticity, but the total marginal effects are very similar.

CREATE ; prin = faminc - ww*whrs $
CREATE ; hinc = hw*hhrs $
CREATE ; winc = ww*whrs $
CREATE ; wasq = wa*wa $
NAMELIST ; w = one,kl6,k618,prin,un,hinc $
NAMELIST ; x = one,wa,wasq,we,cit,ax $
PROBIT ; Lhs = lfp ; Rhs = w ; Hold $
SELECTION ; Lhs = winc ; Rhs = x ; Partial Effects
; Hfn = cit,wa,kl6 $

The intermediate results that already appear above are omitted.

```
-----------------------------------------------------------------------------
Selection with heteroscedasticity
Dependent variable                  WINC
Log likelihood function      -4582.58916
Restricted log likelihood    -4592.48517
Chi squared [   3 d.f.]         19.79201
Significance level                .00019
McFadden Pseudo R-squared       .0021548
Estimation based on N =    753, K =  11
Inf.Cr.AIC  =   9187.2 AIC/N =   12.201
Model estimated: Aug 09, 2011, 19:27:31
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    WINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Slopes in regression function
Constant|   -6666.83       5723.047    -1.16  .2441   -17883.79   4550.14
      WA|    221.284       254.3162      .87  .3842    -277.167   719.734
    WASQ|   -3.40692       488.7922     -.01  .9944  -961.42205  954.60821
      WE|    516.942***    78.61677     6.58  .0000     362.856   671.028
     CIT|    913.228**     443.9549     2.06  .0397      43.093  1783.364
      AX|    194.232***    27.17843     7.15  .0000     140.963   247.500
        |Parameters of heteroscedasticity function
     CIT|      .30289*       .18030     1.68  .0930     -.05048    .65627
      WA|      .00389       4.32370      .00  .9993    -8.47041   8.47819
     KL6|      .12457        .15586      .80  .4242     -.18091    .43004
        |Variance and correlation parameters
SIGMA(1)|    2660.49       1878.462     1.42  .1567    -1021.23   6342.21
RHO(1,2)|    -.32939         .22600    -1.46  .1450     -.77233    .11356
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
|            Marginal Effects in Heteroscedastic Selection Model        |
+----------------------+----------------+----------------+----------------+
|Variable        Direct|       Selection|         Hetero.|           Total|
+----------------------+----------------+----------------+----------------+
|WA          221.28378|          .00000|        -3.48631|       217.79746|
|WASQ         -3.40692|          .00000|          .00000|        -3.40692|
|WE          516.94222|          .00000|          .00000|       516.94222|
|CIT         913.22822|          .00000|      -271.57199|       641.65623|
|AX          194.23157|          .00000|          .00000|       194.23157|
+----------------------+----------------+----------------+----------------+
```

These are the marginal effects estimated in the earlier model without heteroscedasticity. As often happens, though the models are very different, the marginal effects are quite similar.

```
        |Direct effects in the regression
      WA|    270.821        266.8043     1.02  .3101     -252.106    793.748
    WASQ|   -3.92282         3.10679    -1.26  .2067    -10.01201    2.16638
      WE|    560.264***     84.95503     6.59  .0000      393.755    726.773
     CIT|    671.261*       397.7575     1.69  .0915     -108.329   1450.852
      AX|    196.088***     26.77697     7.32  .0000      143.606    248.570
```

Technical Details on Estimation of the Heteroscedasticity Model

The full log likelihood function for the full sample is

$$\log L \;=\; \sum\nolimits_{z=1} \log\left[\frac{\exp\left(-(1/2)\varepsilon_i^2/\sigma_{\varepsilon i}^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(\frac{\rho\varepsilon_i/\sigma_{\varepsilon i}+\boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho^2}}\right)\right] \;+\; \sum\nolimits_{z=0} \log\Phi(-\boldsymbol{\alpha}'\mathbf{w}_i)$$

where the parameter vector is [$\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma},\sigma_\varepsilon,\rho$.] The parameters of the probit selection equation can be estimated consistently in isolation using the probit model. We do so, and insert the estimated $\boldsymbol{\alpha}$ in the model, to obtain

$$\log L\,|\,\hat{\boldsymbol{\alpha}} \;=\; \sum\nolimits_{z=0} \log\Phi\left(-\hat{\boldsymbol{\alpha}}'\mathbf{w}_i\right) \;+\; \sum\nolimits_{z=1} \log\left[\frac{\exp\left(-(1/2)\varepsilon_i^2/\sigma_{\varepsilon i}^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(\frac{\rho\varepsilon_i/\sigma_{\varepsilon i}+\hat{\boldsymbol{\alpha}}'\mathbf{w}_i}{\sqrt{1-\rho^2}}\right)\right]$$

Conditioned on $\hat{\boldsymbol{\alpha}}$, we may now maximize the likelihood with respect to the remaining parameters, [$\boldsymbol{\beta},\boldsymbol{\gamma},\sigma,\rho$]. Any terms involving $\hat{\boldsymbol{\alpha}}$ are irrelevant to the solution. To simplify the process, we make the following substitutions:

$$q_i \;=\; \hat{\boldsymbol{\alpha}}'\mathbf{w}_i$$

$$\eta \;=\; 1/\sigma_\varepsilon$$

$$\tau \;=\; \rho/\sqrt{1-\rho^2}$$

$$\boldsymbol{\delta} \;=\; -\boldsymbol{\gamma}$$

$$\boldsymbol{\mu} \;=\; \boldsymbol{\beta}/\sigma_\varepsilon$$

(Note, we are using the Olsen transformation.) Then, the log likelihood becomes

$$\log L \mid \hat{\boldsymbol{\alpha}} = \sum_{z=0} \log\Phi(-q_i) + \sum_{z=1} \log\eta + \boldsymbol{\delta}'\mathbf{v}_i - \frac{\log 2\pi}{2} - \frac{1}{2}\left(\exp(\boldsymbol{\delta}'\mathbf{v}_i)\right)^2 (\eta y_i - \boldsymbol{\mu}'\mathbf{x}_i)^2 + \log\Phi\left[\tau\left(\exp(\boldsymbol{\delta}'\mathbf{v}_i)\right)(\eta y_i - \boldsymbol{\mu}'\mathbf{x}_i) + q_i\sqrt{1+\tau^2}\right]$$

In spite of its length, this is not a particularly difficult log likelihood to maximize, and estimation of this model is fairly routine.

The derivatives of the log likelihood are as follows, where we give the result for a single observation: Let

$$\varepsilon_i = \eta y_i - \boldsymbol{\mu}'\mathbf{x}_i,$$

$$\kappa_i = \exp(\boldsymbol{\delta}'\mathbf{v}_i),$$

$A_i = \phi(\,.\,) / \Phi(\,.)$ based on the $\log\Phi(.)$ term in $\log L$.

Then,

$$\frac{\partial \log L_i}{\partial \boldsymbol{\mu}} = [\kappa_i^2\, \varepsilon_i - A_i\, \tau\, \kappa_i\,]\mathbf{x}_i,$$

$$\frac{\partial \log L_i}{\partial \boldsymbol{\delta}} = [1 - \kappa_i^2\varepsilon_i^2 + A_i\tau\, \kappa_i\, \varepsilon_i]\mathbf{v}_i,$$

$$\frac{\partial \log L_i}{\partial \eta} = \frac{1}{\eta} - [\kappa_i^2\varepsilon_i - A_i\, \tau\, \kappa_i]y_{,i,}$$

$$\frac{\partial \log L_i}{\partial \tau} = A_i\, [\kappa_i\, \varepsilon_i + q_i\, \tau \,/\, \sqrt{1+\tau^2}\,].$$

The BHHH estimator is used for the asymptotic covariance matrix.

Since this is a two step estimator, we now make the Murphy and Topel correction. Let the vector of derivatives given above, evaluated at the maximum likelihood estimators, be denoted $\mathbf{g}_i$. The transpose of this vector forms the ith row of the matrix **G**, and the BHHH estimator noted above is

$$\mathbf{V}_2 = (\mathbf{G}'\mathbf{G})^{-1}.$$

Now, let

$$\mathbf{m}_i = \frac{\partial \log L_i}{\partial \boldsymbol{\alpha}}$$

$$= -\mathbf{1}(z_i = 0)[\phi(q_i) / \Phi(-q_i)]\mathbf{w}_i + \mathbf{1}(z_i = 1)[\mathrm{A}_i\tau \,/\, \sqrt{1+\tau^2}\,]\,\mathbf{w}_i$$

$$\mathbf{d}_i = \{(2z_i - 1)\phi(q_i) / \Phi[(2z_i - 1)q_i]\}\mathbf{w}_i$$

and define matrices **M** and **D** in the same manner as **G**. Finally, let $\mathbf{V}_1$ denote the estimated asymptotic covariance matrix for the first round estimator of $\boldsymbol{\alpha}$. (This could be the BHHH estimator, $(\mathbf{D}'\mathbf{D})^{-1}$, but *LIMDEP* uses the Hessian for this purpose, instead.) The corrected covariance matrix is

$$\mathbf{V}_2^* = \mathbf{V}_2 + \mathbf{V}_2\, [(\mathbf{G}'\mathbf{M})\mathbf{V}_1(\mathbf{M}'\mathbf{G}) - (\mathbf{G}'\mathbf{D})\mathbf{V}_1(\mathbf{M}'\mathbf{G}) - (\mathbf{G}'\mathbf{M})\mathbf{V}_1(\mathbf{D}'\mathbf{G})\,]\,\mathbf{V}_2.$$

E52.3 Treatment Effects – Using All Observations

If you wish to use the entire sample, that is, not select out any observations, use the specification

; All

in the **SELECTION** command, and otherwise, set it up in the usual manner. In this instance, all computations are exactly as described earlier, save that in the calculations,

$$\lambda_i = (2z_i - 1)\,\phi(\boldsymbol{\alpha}'\mathbf{w}_i) / \Phi[(2z_i - 1)\boldsymbol{\alpha}'\mathbf{w}_i].$$

The model of an endogenous binary variable is an example that would use this formulation. A specification of the selection, known as a 'treatment effects model,' has been used, for example, in the returns to education literature (see Barnow, Cain, and Goldberger (1981));

$$y = \boldsymbol{\beta}'\mathbf{x} + \delta z + \varepsilon,$$

$$z^* = \boldsymbol{\alpha}'\mathbf{w} + u,$$

$$z = 1 \text{ if } z^* > 0 \text{ and } z = 0 \text{ if } z^* \leq 0.$$

The indicator, z is assumed to indicate the presence or absence of some treatment, for example, participation in an experiment or going to college. This is the same as the selectivity model discussed earlier except that z itself appears in the primary equation. Thus, there is an endogenous variable in the regression equation. On the other hand, note that conditioned on z (i.e., the 'selection') this is the same model we have been examining so far. There are three approaches to estimation.

E52.3.1 Two Step Estimation

Barnow, et. al. suggest two methods of estimating this model. The simplest method is to use the selection model exactly as before. It is still necessary to estimate the probit equation for z and pass the results to **SELECT**. If z is now simply included among the Rhs variables in the **SELECT** command, consistent estimates of $\boldsymbol{\beta}$ and δ are obtained. It is necessary, however, in this case, to use the entire sample of data, so the additional specification **; All** is necessary. All other output, saved results, options, etc. for the **SELECT** command are the same. The initial results for the model will indicate that the entire sample is in use, as in the following:

```
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on LFP                     |
| Sample is all observations.                                |
| Results of selection:                                      |
|                       Data points     Sum of weights       |
| Data set                   753             753.0           |
| Selected sample            753             753.0           |
+------------------------------------------------------------+
```

E52.3.2 Two Stage Least Squares – Instrumental Variable Estimation

A second means of estimating the model is with two stage least squares. The problem with ordinary least squares estimates of the model based on the observed data is the correlation between z and ε. A solution to the inconsistency of OLS is to use 2SLS, using as the instrumental variable for z the predicted probabilities from the probit equation. It is not necessary to **; Hold** the results of the probit in this case. The set of commands would be

```
NAMELIST   ; w = ...  ; x = ... $
PROBIT     ; Lhs = z ; Rhs = w ; Prob = zfit $
2SLS       ; Lhs = y ; Rhs = x,z
           ; Inst = x,zfit $
```

We note, there is a tendency in the literature to equate the simple replacement of z_i in the regression with the fitted probability as an instrumental variable estimator. Ordinary least squares is then used to estimate the parameters. We emphasize, this is not 2SLS for this model and the replacement variable is not an instrument, it is a proxy. Whether the estimator so constructed is even consistent is debatable. The following, developed below in the application, illustrates use of 2SLS to fit the treatment model. In the main equation, we fit an hours equation for the husband, where the 'treatment' is whether the wife is in the labor force.

```
NAMELIST   ; x  = one,ha,he,hw,faminc $
NAMELIST   ; w = one,we,age,agesq,kl6,k618 $
PROBIT     ; Lhs = lfp ; Rhs = w
           ; Prob  = pfit $
2SLS       ; Lhs = hhrs ; Rhs = x,lfp
           ; Inst = x,pfit $
```

E52.3.3 Maximum Likelihood Estimation

Finally, a third approach is full information maximum likelihood. The log likelihood for the treatment effects model is

$$\log L = \sum_{i=1}^{N} \log\left[\frac{\exp\left(-(1/2)\varepsilon_i^2/\sigma_\varepsilon^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(\frac{(2z_i-1)(\rho\varepsilon_i/\sigma_e + \boldsymbol{\alpha}'\mathbf{w}_i)}{\sqrt{1-\rho^2}}\right)\right]$$

where $\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i - \delta z_i$.

This is a straightforward modification of the estimator developed earlier for the selection model. To fit the treatment effects model, just add **; MLE** to the two step estimator. The commands are

```
PROBIT     ; Lhs = variable z ; Rhs = variables in w ; Hold $
SELECT     ; Lhs = variable y ; Rhs = variables in x, variable z ; All ; MLE $
```

E52.3.4 Application

In the following, we fit a 'treatment model' for the husband's hours, where the endogenous dummy variable is the wife's labor force participation. The following uses all three estimators.

```
NAMELIST    ; x  = one,ha,he,hw,faminc $
NAMELIST    ; w = one,we,age,agesq,kl6,k618 $
PROBIT      ; Lhs = lfp    ; Rhs = w ; Hold ; Prob = pfit $
SELECT      ; Lhs = hhrs ; Rhs = x,lfp ; All $
2SLS        ; Lhs = hhrs ; Rhs = x,lfp ; Inst = x,pfit $
SELECT      ; Lhs = hhrs ; Rhs = x,lfp ; All ; MLE $
```

These are the two step estimators using Heckman's method.

```
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on LFP                     |
| Sample is all observations.                                |
| Results of selection:                                      |
|                  Data points     Sum of weights            |
| Data set              753              753.0               |
| Selected sample       753              753.0               |
+------------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step     least squares regression ............
LHS=HHRS     Mean                     =     2267.27092
             Standard deviation       =      595.56665
             Number of observs.       =            753
Model size   Parameters               =              7
             Degrees of freedom       =            746
Residuals    Sum of squares           =    .181436E+09
             Standard error of e      =      493.16533
Fit          R-squared                =         .31340
             Adjusted R-squared       =         .30788
Model test   F[  6,   746] (prob) =    56.8(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Standard error corrected for selection.. 495.44230
Correlation of disturbance in regression
and Selection Criterion (Rho)...........   -.12587
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    HHRS|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2350.53***    150.0000    15.67  .0000     2056.53    2644.52
      HA|   -6.42709***     2.32011    -2.77  .0056   -10.97442   -1.87976
      HE|    30.4775***     6.94286     4.39  .0000     16.8697    44.0852
      HW|   -112.030***     6.43336   -17.41  .0000    -124.639    -99.421
  FAMINC|     .03248***      .00223    14.58  .0000      .02812     .03685
     LFP|   -150.410       109.3608    -1.38  .1690    -364.753     63.934
  LAMBDA|   -62.3601       70.67071     -.88  .3776   -200.8722    76.1519
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Two stage    least squares regression ...........
LHS=HHRS     Mean                 =    2267.27092
             Standard deviation   =     595.56665
             Number of observs.   =           753
Model size   Parameters           =             6
             Degrees of freedom   =           747
Residuals    Sum of squares       =   .183027E+09
             Standard error of e  =     494.99127
Fit          R-squared            =        .30831
             Adjusted R-squared   =        .30368
Instrumental Variables:
ONE       HA        HE        HW        FAMINC    PFIT
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2334.61***    159.8080    14.61  .0000     2021.39   2647.83
      HA|   -6.26441***     2.35599    -2.66  .0078   -10.88206  -1.64675
      HE|    31.4818***     6.78263     4.64  .0000     18.1881   44.7755
      HW|   -109.163***     7.48834   -14.58  .0000    -123.840   -94.486
  FAMINC|     .03159***      .00260    12.17  .0000      .02651    .03668
     LFP|   -159.015       111.4911    -1.43  .1538    -377.533    59.504
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                 HHRS
Log likelihood function     -6202.52230
FIRST  6 estimates are probit equation.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selection (probit) equation for LFP
Constant|    -.23352        1.54515     -.15  .8799    -3.26195   2.79491
      WE|     .11944***      .02223     5.37  .0000      .07588    .16300
      WA|     .00276         .07099      .04  .9690     -.13638    .14190
    WASQ|    -.00047         .00081     -.58  .5625     -.00207    .00112
     KL6|    -.87593***      .11397    -7.69  .0000    -1.09932   -.65255
    K618|    -.05539         .04028    -1.38  .1691     -.13434    .02355
        |Corrected regression, Regime 1
Constant|    2351.32***    140.8639    16.69  .0000     2075.23   2627.41
      HA|   -6.43033***     2.19962    -2.92  .0035   -10.74150  -2.11916
      HE|    30.5281***     6.68052     4.57  .0000     17.4345   43.6217
      HW|   -112.027***     4.13167   -27.11  .0000    -120.125  -103.929
  FAMINC|     .03250***      .00153    21.25  .0000      .02950    .03549
     LFP|   -153.227       115.7198    -1.32  .1855    -380.034    73.580
   SIGMA|    495.319***    11.30461    43.82  .0000     473.162   517.475
     RHO|    -.12200         .14665     -.83  .4055     -.40944    .16543
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E52.4 Simultaneous Equations Models with Selectivity

A simultaneous equations model which is 'selected' is estimated in exactly the same fashion as a single equation, using a form of two stage least squares. (See Lee, Maddala, and Trost (1980).) We consider a simple three equation case; the extensions to other cases would be analogous. The assumed model is:

$$y_1 = \alpha_1 + \beta_{1,1}x_1 + \beta_{1,2}x_2 + \beta_{1,3}x_3 + \gamma_{1,2}y_2 + \varepsilon_1,$$

$$y_2 = \alpha_2 + \beta_{2,1}x_1 + \beta_{2,3}x_3 + \beta_{2,4}x_4 + \gamma_{2,1}y_1 + \varepsilon_2,$$

$$y_3 = \alpha_3 + \beta_{3,5}x_5 + \gamma_{3,1}y_1 + \gamma_{3,2}y_2 + \varepsilon_1 .$$

This structure is observed if $z = 1$; some other if $z = 0$. The general procedure would be:

Step 1. Estimate the probit selection equation.

Step 2. Estimate each equation of the reduced form and keep the fitted values.

Step 3. Estimate the structural equations using **SELECT**, using fitted instead of actual values on the right hand side of each equation.

Step 4. As in a conventional simultaneous equations model, it is necessary to use the original data, not the predicted values, when computing the estimate of the disturbance variance. (Step 4 is done automatically, internally.)

To accommodate the last of these, in order to estimate a simultaneous equations model with selectivity, after obtaining the predicted values at Step 3, the **SELECT** commands should be the same as the usual **2SLS** commands. That is,

```
SELECT      ; Lhs  = left hand side variable
            ; Rhs  = original variables including endogenous
            ; Inst = instruments, with fitted values in place of actual $
```

For the model shown above, the commands would be

```
NAMELIST    ; x = one,x1,x2,x3,x4,x5 $
PROBIT      ; Lhs = z   ; Rhs = list of variables in w ; Hold $
SELECT      ; Lhs = y1 ; Rhs = x ; Keep = y1fit $
SELECT      ; Lhs = y2 ; Rhs = x ; Keep = y2fit $
SELECT      ; Lhs = y3 ; Rhs = x ; Keep = y3fit $
SELECT      ; Lhs = y1 ; Rhs = one,x1,x2,x3,y2
            ; Inst = one,x1,x2,x3,y2fit $
SELECT      ; Lhs = y2 ; Rhs = one,x1,x3,x4,y1
            ; Inst = one,x1,x3,x4,y1fit $
SELECT      ; Lhs = y3 ; Rhs = one,x5,y1,y2
            ; Inst = one,x5,y1fit,y2fit $
```

No mention is made in the output of the simultaneous nature of the model, but the necessary adjustments are made internally. Since it is assumed that you are generating the predicted values yourself, it follows that the number of instruments is always identical to the number of Rhs variables. If not, an error is assumed, and estimation is terminated.

A small application based on our earlier example is shown below. In this model, it is assumed that hours (*whrs*) and wage (*ww*) are simultaneously determined.

```
NAMELIST    ; x  = one,kl6,k618,wa,we,wmed $
NAMELIST    ; w = one,kl6,wa $
PROBIT      ; Lhs = lfp    ; Rhs = w ; Hold $
SELECT      ; Lhs = whrs ; Rhs = x  ; Keep = whrsfit $
SELECT      ; Lhs = ww    ; Rhs = x  ; Keep = wwfit $
SELECT      ; Lhs = whrs ; Rhs = one,kl6,k618,wa,ww
            ; Inst = one,kl6,k618,wa,wwfit $
SELECT      ; Lhs = ww    ; Rhs = one,wa,we,wmed,whrs
            ; Inst = one,wa,we,wmed,whrsfit $
```

E52.5 Incidental Truncation

In the single equation probit/selection model, if there are no observations with $z = 0$, then the probit model cannot be estimated. Bloom and Killingsworth (1985) demonstrated that the model can still be fit, by maximum likelihood. The full specification is:

$$\begin{aligned} y &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \\ z^* &= \boldsymbol{\alpha}'\mathbf{w} + u, \\ z &= 1 \text{ if } z^* > 0,\ z = 0 \text{ if } z^* \le 0, \\ \varepsilon,u &\sim N[0,0,\sigma_\varepsilon^2,1,\rho]. \end{aligned}$$

A probit model applies to z. In the usual circumstance, the familiar selectivity model applies. However, if observations on y, z, $\mathbf{x}$, and $\mathbf{w}$ are obtained only when z equals one, then the model is no longer estimable by the method of Heckman. (The dependent variable in the probit equation is always one.) But, the model is identified as long as ρ is nonzero, and can be estimated by maximum likelihood. The paper by Bloom and Killingsworth provides the methodology.

The data consist of a set of observations on y, $\mathbf{x}$, and $\mathbf{v}$, with which it is known at every observation that z equals one. (Observations on z are unnecessary.) The model command is

```
INCIDENTAL  ; Lhs  = y
            ; Rh1 = x
            ; Rh2 = w $
```

The only internal starting value for this model which seems viable is a pair of least squares estimators. Thus, starting values are obtained by regressing y on $\mathbf{x}$ to estimate $\boldsymbol{\beta}$ and y on $\mathbf{w}$ to estimate $\boldsymbol{\alpha}$. Admittedly, these starting values are inconsistent. The second set are particularly problematic. We use them only to provide some information on the relative scales of the coefficients in the second equation. The first variable in $\mathbf{w}$ must be the constant term, *one*. The starting value for ρ is the ratio of the first coefficient in the second equation to the estimate of σ from the first. (This is computed internally.) We assume that the constant term is first in the second equation.

The model parameters are $\boldsymbol{\theta} = [\boldsymbol{\beta},\boldsymbol{\alpha},\sigma,\rho]$. Use this ordering if you provide your own starting values for the iterations or use **; Rst** to impose restrictions. The retrievable results are

Matrices: *b* and *varb* include $[\boldsymbol{\beta},\boldsymbol{\alpha}]$. **; Par** adds (σ,ρ) to the parameter vector.

Scalars: *ybar, sy, nreg, kreg, logl, rho, varrho, exitcode*

Last Model: *b_variables in x, a_variables in v , sigma, reu*.

Last Function: None

All options for nonlinear models are available. In particular, the **; Test:** specification can be used to test the equality of coefficients across equations, and **; Rst** can be used to impose fixed value and within or cross equation equality restrictions. Predictions are computed using the formula given earlier for the standard selection model – once the parameters are in hand, the observed data can be treated as if the standard model applies.

Standard Model Specifications for the Incidental Truncation Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps ancillary parameters σ and ρ with main parameter $\boldsymbol{\beta}$ vector in *b*.
; Partial Effects	displays marginal effects, same as **; Marginal Effects**.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec	requests computation of the cluster form of corrected covariance estimator.
; Robust	requests a ‘sandwich’ estimator or robust covariance matrix for TSCS and some discrete choice models (uses **; Cluster = 1**).

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg [= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb [= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E53: Sample Selection Models for Panel Data

E53.1 Introduction

Heckman's now canonical form of the sample selection model is a linear regression with a binary probit selection criterion model:

$$
\begin{aligned}
y &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \\
z^* &= \boldsymbol{\alpha}'\mathbf{w} + u, \\
\varepsilon, u &\sim N[0,0,\sigma_\varepsilon^2, \sigma_u^2, \rho].
\end{aligned}
$$

A bivariate classical (seemingly unrelated) regressions model applies to the structural equations. The standard deviations are σ_ε and σ_u, and the covariance is $\rho\sigma_\varepsilon\sigma_u$. If the data were randomly sampled from this bivariate population, the parameters could be estimated by least squares, or GLS combining the two equations. However, z^* is not observed. Its observed counterpart is z, which is determined by

$$z = 1 \text{ if } z^* > 0$$

and

$$z = 0 \text{ if } z^* \leq 0.$$

Values of y and $\mathbf{x}$ are only observed when z equals one. The essential feature of the model is that under the sampling rule, $E[y|\mathbf{x},z=1]$ is not a linear regression in $\mathbf{x}$, or $\mathbf{x}$ and z. The development below presents estimators for the class of essentially nonlinear models that emerge from this specification.

The basic command structure for the models described in this chapter is

PROBIT ; Lhs = variable z ; Rhs = variables in w ; Hold $
SELECT ; Lhs = variable y ; Rhs = variables in x $

Note that two commands are required for estimation of the sample selection model, one for each structural equation. Chapter E52 defined several estimators appropriate for cross sectional treatment of the model. This chapter will develop some panel data approaches.

E53.2 Panel Data Treatments

The literature on panel data models for the sample selection is rather incomplete and ambiguous. Applications are relatively sparse, and few useable general modeling frameworks have been proposed. The earliest contribution appears to be Hausman and Wise's (1979) paper on attrition (see Section E53.7 for the estimator) which is a contemporary of Heckman's seminal paper on cross sections. The Hausman and Wise model is a two period fully parametric model. The literature has come nearly full circle since then, in that some of the later work (Kyriazidou (1997) focuses once again on the two period framework. (In Hausman and Wise's case, two periods was a natural application, as their interest lay in the beginning (baseline) and ending point of a study, whereas in the more recent analyses, two periods is often assumed of necessity to make the analysis tractable.)

Fixed and random effects, and hybrid models have been suggested by Verbeek (1990), Zabel (1992) and Verbeek and Nijman (1992). The estimators suggested here build on the suggestions by these authors, and extend them in several directions. The following modeling frameworks are provided:

Fixed Effects

$$y_{it} = \theta_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ [u_{it}, \varepsilon_{it}] \sim \text{BVN}[(0,0),1,\sigma,\rho]$$

$$z_{it}^* = \alpha_i + \boldsymbol{\delta}'\mathbf{w}_{it} + u_{it}$$

$$z_{it} = \mathbf{1}(z_{it}^* > 0)$$

y_{it}, $\mathbf{x}_{it}$ observed only when $z_{it} = 1$.

Random Effects

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + c_i,\ \varepsilon_{it} \sim \text{N}[0,\sigma^2],\ c_i \sim \text{N}[0,\sigma_c^2]$$

$$z_{it}^* = \boldsymbol{\alpha}'\mathbf{w}_{it} + u_{it} + d_i$$

$$z_{it} = 1(z_{it}^* > 0),\ u_{it} \sim \text{N}[0,1],\ d_i \sim \text{N}[0,\sigma_d^2]$$

y_{it}, $\mathbf{x}_{it}$ observed only when $z_{it} = 1$,

$$\text{Corr}[\varepsilon_{it},u_{it}] = \rho$$

$$\text{Corr}[c_i, d_i] = \theta$$

Random Parameters

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it}\ \ \varepsilon_{it} \sim \text{N}[0,\sigma^2], \qquad \boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}_\beta\mathbf{f}_i + \boldsymbol{\Gamma}_\beta\mathbf{v}_i$$

$$z_{it}^* = \boldsymbol{\alpha}_i'\mathbf{w}_{it} + u_{it},\ u_{it} \sim \text{N}[0,1], \qquad \boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \boldsymbol{\Delta}_\alpha\mathbf{g}_i + \boldsymbol{\Gamma}_\alpha\mathbf{h}_i,$$

$$z_{it} = 1(z_{it}^* > 0)$$

y_{it}, $\mathbf{x}_{it}$ observed when $z_{it} = 1$

$$\text{Corr}[\varepsilon_{it},u_{it}] = \rho$$

Hausman and Wise's Attrition Model

$$y_{i1} = \mathbf{x}_{i1}'\boldsymbol{\beta} + \varepsilon_{i1} + u_i,\ \sigma^2 = \text{Var}[\varepsilon_{i1} + u_i] \qquad \text{(first period regression)}$$

$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta} + \varepsilon_{i2} + u_i \qquad \text{(second period regression, same } u_i\text{)}$$

$$z_{i2}^* = \delta y_{i2} + \mathbf{x}_{i2}'\boldsymbol{\theta} + \mathbf{w}_{i2}'\boldsymbol{\alpha} + v_{i2} \qquad \text{(attrition mechanism)}$$

$$z_{i2}^* = \mathbf{x}_{i2}'(\delta\boldsymbol{\beta} + \boldsymbol{\theta}) + \mathbf{w}_{i2}'\boldsymbol{\alpha} + \delta\varepsilon_{i2} + v_{i2} = \mathbf{r}_{i2}'\boldsymbol{\gamma} + h_{i2}.$$

$$z_{i2} = 1(z_{i2}^* > 0) \qquad \text{(attrition indicator observed in period 2)}$$

$$\rho_{12} = \text{Corr}[\varepsilon_{i1} + u_i, \varepsilon_{i2} + u_i] = \text{Var}[u_i] / \sigma^2$$

$$\rho_{23} = \text{Corr}[h_{i2}, \varepsilon_{i2} + u_i]$$

In each case, there are several different forms of the model which may be specified.

E53.3 Sample Selection Models with Fixed Effects

A sample selection model with fixed effects would appear as follows: The structural probit model would be

$$z_{it}^* = \alpha_i + \boldsymbol{\delta}'\mathbf{w}_{it} + u_{it}$$

$$z_{it} = \mathbf{1}(z_{it}^* > 0)$$

The primary regression equation is, then

$$y_{it} = \theta_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ [u_{it}, \varepsilon_{it}] \sim \text{BVN}[(0,0),1,\sigma,\rho]$$

y_{it}, $\mathbf{x}_{it}$ are observed only when $z_{it} = 1$.

Thus, the familiar selection model applies for each person in each period. This model is fit by a hybrid two step maximum likelihood procedure described below.

NOTE: If you wish to include time effects in this model, you must create them separately as dummy variables and include them in the model specifications. In order to implement the procedure described below, it is necessary to disable the automatic creation of the time dummies in this model.

The command for this model is completely self contained. Use

```
SELECT    ; Lhs = y, z (specify both dependent variables)
          ; Rhs = list of variables in x
          ; Rh2 = list of variables in w
          ; FEM
          ; Pds = specification of the panel $
```

It is not necessary to precede this with a **PROBIT** command, as the probit equation is fit at the same time as the selection model. The 'treatment effects' model, in which z_{it} appears in the regression and all observations are used,

$$y_{it} = \theta_i + \mu z_{it} + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ [u_{it}, \varepsilon_{it}] \sim \text{BVN}[(0,0),1,\sigma,\rho],$$

may be requested as in the cross section case by just including z in the Rhs list for the regression (if appropriate) and adding the request to the command:

; All

A special case in the data must be considered when fitting this model. If the probit model for individual i cannot be fit because z_{it} is always zero or one in every period, then the selection model cannot be either. Such groups must be skipped over in estimation of the model. This is the same condition that must be met for the probit model, but it is more likely to be a problem in this setting, as the selection is likely to be the same in every period. One possibility might be a model extension which treats selection as observation of the entire group or not, instead of period by period. This remains to be developed – the nature of the underlying correlation is complicated by this modification.

Results that are kept for this model are

Matrices:	*b*	= estimate of **β**
	varb	= asymptotic covariance matrix for estimate of **β**.
	alphafe	= estimated fixed effects
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function

Last Model: *b_variables*

Last Function: None

Asymptotic standard errors for the estimator in this model are computed by using bootstrapping. In order to request the computation of bootstrapped standard errors, add

; Nbt = the desired number of replications

Computation of the model is not always possible. As such, some of the bootstrap replications may fail. A trace of the replications appears in the model results. The example below illustrates. To compute the asymptotic standard errors without bootstrapping, use

; Nbt = 1

In this instance, the standard errors are computed as if the probit model estimates were known – that is, as if this were not a two step estimator.

E53.3.1 Standard Model Specifications

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Partial Effects displays marginal effects, same as **; Marginal Effects**.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Prob = name	saves probabilities as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.

E53.3.2 Application

The following illustrates this estimator with a random sample drawn so that there are 200 individuals and 15 periods of observation:

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-3000 $
MATRIX      ; ai = Rndm(200) ; ci = Rndm(200) $
CREATE      ; i = Trn(15,0) ; u = Rnn(0,1) ; e = .5*u+.5*Rnn(0,1) $
CREATE      ; z1 = Rnn(0,1) ; z2 = Rnn(0,1)
            ; d = (.5*z1+.5*z2+ai(i)+u) > 0 $
CREATE      ; x1 = Rnn(0,1) ; x2 = Rnn(0,1)  ; y = x1+x2+ci(i)+e $
SELECT      ; Lhs = y,d
            ; Rh1 = x1,x2 ; Rh2 = z1,z2
            ; FEM ; Pds = 15 ; Nbt = 10 $
```

These are the initial probit estimates computed to obtain the initial values.

```
--------------------------------------------------------------------------
Probit   Regression Start Values for D
Dependent variable                     D
Log likelihood function     -1895.34311
--------+-----------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------
      Z1|     .34436***      .02537    13.57  .0000      .29464    .39408
      Z2|     .32921***      .02505    13.14  .0000      .28011    .37831
Constant|    -.04731**       .02376    -1.99  .0465     -.09389   -.00073
--------+-----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------
Normal exit from iterations. Exit status=0.
--------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable                     D
Log likelihood function     -1179.81536
Estimation based on N =   3000, K = 202
Sample is 15 pds and    200 individuals
Skipped    0 groups with inestimable ai
PROBIT (normal)  probability model
Std. errors based on 10 bootstraps.
--------+-----------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------
        |Index function for probability
      Z1|     .52106***      .03391    15.36  .0000      .45459    .58752
      Z2|     .50736***      .03423    14.82  .0000      .44027    .57445
--------+-----------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------
Normal exit from iterations. Exit status=0.
Completed  2 bootstrap replications of 11.
Completed  3 bootstrap replications of 11.
Completed  4 bootstrap replications of 11.
Completed  5 bootstrap replications of 11.
Completed  6 bootstrap replications of 11.
Completed  7 bootstrap replications of 11.
Completed  8 bootstrap replications of 11.
Completed  9 bootstrap replications of 11.
Completed 10 bootstrap replications of 11.
--------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable                     D
Log likelihood function     -1239.78148
Sample is 15 pds and    200 individuals
Skipped    0 groups with inestimable ai
PROBIT (normal)  probability model
Std. errors based on 10 bootstraps.
```

```
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
      Z1|     .49606***      .03228    15.37  .0000      .43280    .55933
      Z2|     .53084***      .03304    16.07  .0000      .46609    .59559
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------


-----------------------------------------------------------------------------
FIXED EFFECTS Probit Model
Dependent variable                     D
PROBIT (normal)  probability model
Std. errors based on 10 bootstraps.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
      Z1|     .46132***      .03193    14.45  .0000      .39874    .52390
      Z2|     .46627***      .03309    14.09  .0000      .40141    .53113
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Completed 11 bootstrap replications of 11.

-----------------------------------------------------------------------------
FIXED EFFECTS Select Model
Dependent variable                     Y
Log likelihood function     -1859.55872
Skipped    0 groups with inestimable ai
Sample selection (by probit) model
Selection effects model based on D
Std. errors based on 10 bootstraps.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selected regression parameters
      X1|    1.02184***      .10561     9.68  .0000      .81484   1.22883
      X2|    1.02110***      .01178    86.65  .0000      .99800   1.04419
        |Regression standard deviation
   Sigma|     .66743***      .01252    53.30  .0000      .64288    .69197
        |Correlation coefficient
     Rho|     .78222***      .03344    23.39  .0000      .71669    .84776
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This warning applies to a bootstrap replication that failed. The estimator continues to attempt to complete the requested number of replications. In this case, as noted, it requires 11 attempts to complete the 10 replications.

```
Error   143: Models - estimated variance matrix of estimates is singular
Bootstrap rep. 10, attempt  1 failed. Continuing.
```

E53.3.3 Technical Details on FE Selection Models

The log likelihood function for this model including all parameters, for the *i*th individual in the sample is as follows:

$$\log L_i = \sum_{z_{it}=0} \log \boldsymbol{\Phi}\left(-\alpha_i - \boldsymbol{\delta}'\mathbf{w}_{it}\right) +$$

$$\sum_{z_{it}=1} \left[\begin{array}{l} \dfrac{-\log 2\pi}{2} - \log\sigma - \dfrac{(y_{it} - \theta_i - \boldsymbol{\beta}'\mathbf{x}_{it})^2}{2\sigma^2} + \\ \log \boldsymbol{\Phi}\left[\dfrac{(\alpha_i + \boldsymbol{\delta}'\mathbf{w}_{it}) + (\rho/\sigma)(y_{it} - \theta_i - \boldsymbol{\beta}'\mathbf{x}_{it})}{\sqrt{1-\rho^2}} \right] \end{array} \right]$$

The full log likelihood is then summed over all individuals. We first make the following reparameterizations:

$$\eta = 1/\sigma$$

$$\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta}$$

$$\kappa_i = \theta_i / \sigma$$

$$\tau = \rho/\sqrt{1-\rho^2}$$

This greatly simplifies the log likelihood:

$$\log L_i = \sum_{z_{it}=0} \log \boldsymbol{\Phi}\left(-\alpha_i - \delta'\mathbf{w}_{it}\right) +$$

$$\sum_{z_{it}=1} \left[\begin{array}{l} \dfrac{-\log 2\pi}{2} + \log\eta - (\eta y_{it} - \kappa_i - \boldsymbol{\gamma}'\mathbf{x}_{it})^2 + \\ \log \boldsymbol{\Phi}\left[\left(\sqrt{1+\tau^2}\right)(\alpha_i + \boldsymbol{\delta}'\mathbf{w}_{it}) + \tau(\eta y_{it} - \kappa_i - \boldsymbol{\gamma}'\mathbf{x}_{it}) \right] \end{array} \right]$$

In principle, this can now be maximized to provide fully efficient maximum likelihood estimates of the model's parameters. However, this would entail simultaneous estimation of two sets of fixed effects parameters. We do this in two steps instead, in a fashion similar to the Heckman style two step estimator for the cross section counterpart.

Step 1. The fixed effects probit estimator is estimated using the method described in Section R23.2.3. The model estimates are retained for Step 2.

Step 2. The log likelihood function is now conditioned on the probit estimates obtained at Step 1.

Let

$$\mu_{it} = \alpha_i + \delta' w_{it}$$

and let m_{it} denote the estimate of μ_{it} obtained by computing it at the Step 1 probit estimates. The conditional log likelihood is

$$\log L_C = \sum_{z_{it}=0} \log \boldsymbol{\Phi}\left(-m_{it}\right) \quad +$$

$$\sum_{z_{it}=1} \left[\begin{array}{l} \dfrac{-\log 2\pi}{2} + \log \eta - (\eta y_{it} - \kappa_i - \boldsymbol{\gamma}'\mathbf{x}_{it})^2 \\ + \log \boldsymbol{\Phi}\left[\left(\sqrt{1+\tau^2}\right) m_{it} + \tau(\eta y_{it} - \kappa_i - \boldsymbol{\gamma}'\mathbf{x}_{it}) \right] \end{array} \right]$$

At this step, the conditional log likelihood function is maximized with respect to the remaining parameters, η, κ_i, $\boldsymbol{\gamma}$, and τ. Note that the $z_{it} = 0$ observations are not used in obtaining this solution. In the treatment effects model, once again, only terms from the second part of the function are included, but the sign and form of the argument in $\Phi(.)$ is changed appropriately. If the treatment effects model is requested, then the sign of the term in the second part of the log likelihood is also changed accordingly, observation by observation. No other changes are needed internally.

Aside from the aforementioned incidental parameters problem, these estimates are consistent, albeit inefficient. However, since it is a two step estimator, the estimated asymptotic covariance matrix is inappropriate as it does not account for the randomness induced by estimation of the probit parameters used to compute m_{it}. In other applications we have used the Murphy and Topel estimator to complete this computation. In this case, however, that would require a full covariance matrix for the fixed effects parameters in the probit model and, moreover, require an exorbitant amount of computation. With contemporary computers, the latter consideration is generally going to be minor, but the former remains problematic.

As an alternative, we use the bootstrap method of obtaining an estimator of the asymptotic covariance matrix. Our approach is as follows: The number of bootstrap replications is set either at 20 or with

; Nbt = desired number of replications.

Within each replication, the bootstrap is drawn over the set of *N* individuals, not over the full sample. Suppose, for example, the sample contains 200 people, each observed 15 times (our earlier application). The bootstrap samples are then drawn from these 200 individuals, each with their 15 observations. Then, the entire two step procedure is computed for each replication – both the probit model and the selection model.

There is an additional complication. The log likelihood for the sample selection model is not globally concave. As a consequence, the iterations at Step 2 occasionally break down because the Hessian becomes indefinite. If this occurs during the initial estimation, the process is halted. However, if this occurs during a bootstrap replication the program tries again with a new bootstrap sample. This 'retry' is repeated up to 10 times. If after 10 tries it remains impossible to obtain a set of estimates, the routine gives up.

E53.4 Sample Selection Models with Random Effects

There is a lengthy literature on fixed and random effects in sample selection models. The fixed effects model was presented in the preceding section. The random effects model is cast in its simplest terms in Verbeek (1990), Zabel (1992) and Verbeek and Nijman (1992). The structural equations are:

Regression

$$y_{it} = \beta' \mathrm{x}_{it} + \varepsilon_{it} + c_i, \ \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated}$$

Selection Mechanism

$$z_{it}^* = \boldsymbol{\alpha}'\mathbf{w}_{it} + u_{it} + d_i$$

$$z_{it} = 1(z_{it}^* > 0), \ u_{it} \sim N[0,1]$$

Selection

$$y_{it}, \mathrm{x}_{it} \quad \text{observed only when } z_{it} = 1,$$

$$\mathrm{Corr}[\varepsilon_{it},u_{it}] = \rho$$

The random effects, (c_i, d_i) are assumed to be bivariate normally distributed with zero means, standard deviations σ_c and σ_d and correlation θ. 'Selectivity' comes in two forms here, through the correlation of the unique components, ε_{it} and u_{it}, and the correlation of the group specific components, c_i and d_i. Estimable parameters in this model are the slope parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, variance parameters σ_c and σ_d and the two correlation parameters.

NOTE: This model is fit by maximum simulated likelihood, not by two step least squares. There is no 'lambda' variable, $\phi(...)/\Phi(...)$ created or used during the estimation, so no coefficient for this variable will appear in the results.

E53.4.1 Including Group Means

A standard criticism of the random effects approach is that the group effects are likely to be correlated with the included variables. Zabel (1992) suggests that this can be remedied by including the group means of the variables in the models. The modified specification is

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \varepsilon_{it} + c_i, \ \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated}$$

$$z_{it}^* = \boldsymbol{\alpha}'\mathbf{w}_{it} + \boldsymbol{\gamma}'\overline{\mathbf{w}}_i + u_{it} + d_i$$

with the same stochastic specification.

E53.4.2 Treatment Effects

The 'treatment effects' model, in which z_{it} appears in the regression and all observations are used,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \gamma z_{it} + \varepsilon_{it} + c_i, \ \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated}$$

is requested as in the cross section case by just including z in the Rhs list for the regression (if appropriate) and adding the request to the command:

; All

E53.4.3 Commands

This model is fit as a random parameters model, using simulation rather than quadrature to do the estimation. It must be fit in three steps as shown below

Step 1. Compute the probit model to define selection mechanism

PROBIT ; Lhs = variable zit ; Rhs = one,... ; Hold $

Step 2. Selection model to produce good starting values

SELECT ; Lhs = variable yit ; Rhs = one,... ; MLE $

Step 3. Random effects model

SELECT ; Lhs = variable zit ; Rhs = one,...
; RPM ; Pds = panel specification
; Fcn = REM $

NOTE: Both equations must include a constant term, *one*.

The model part of the second **SELECT** command is the same as that in the first one. Zabel's modification is requested by adding

; Means

to the **SELECT** command. The default specification is uncorrelated group effects ($\theta = 0$). This may be relaxed by adding

; Correlated

to the command. Our experience suggests that the identification of θ is a bit weak. In the experiment below, in a large sample of $N = 300$, $T = 15$, in which the correlation is zero by construction in the original data, the estimate of it is, nonetheless, large and highly significant. However, convergence of the iterations could not be reached; the likelihood surface became quite flat in the dimension of θ in a range in which the derivatives with respect to the other parameters were fairly far from zero.

E53.4.4 Other Model Specifications

This model is estimated as a random parameters model with two random coefficients, the constants in the two equations. The set of options for the model specification are the same as for other random parameters models. See Chapter R24 for discussion. The results retained by this estimator are

Matrices:	*b*	= full coefficient vector
	varb	= full estimated asymptotic covariance matrix
Scalars:	*nreg*	= total number of observations
	kreg	= number of parameters estimated
	logl	= log likelihood
	s	= estimate of σ
	rho	= estimate of ρ

E53.4.5 Application

The model is applied to the data used to illustrate the fixed effects model. By construction, these data actually conform to the random effects model without the means included and with uncorrelated effects. Each formulation begins with the initial **PROBIT** followed by **SELECT**.

Basic Random Effects Formulation

```
PROBIT    ; Lhs = d ; Rhs = one,z1,z2 ; Hold $
SELECT    ; Lhs = y ; Rhs = one,x1,x2 ; MLE ; Par $
SELECT    ; Lhs = y ; Rhs = one,x1,x2
          ; RPM ; Pds = 15 ; Fcn = REM $
```

Random Effects with Group Means (First SELECT is the same)

```
SELECT    ; Lhs = y ; Rhs = one, x1,x2
          ; RPM ; Pds = 15 ; Fcn = REM ; Means $
```

Correlated Random Effects (First SELECT is the same)

```
SELECT    ; Lhs = y ; Rhs = one,x1,x2
          ; RPM ; Pds = 15 ; Fcn = REM ; Correlated ; Halton $
```

Group Means and Correlated Random Effects (First SELECT is the same)

```
SELECT    ; Lhs = y ; Rhs = one,x1,x2
          ; RPM ; Pds = 15 ; Fcn = REM ; Means ; Correlated ; Halton $
```

We show the full set of results for the second model, the REM with group means.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                  D
Log likelihood function     -1895.34311
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.04731**       .02376    -1.99  .0465     -.09389   -.00073
      Z1|     .34436***      .02537    13.57  .0000      .29464    .39408
      Z2|     .32921***      .02505    13.14  .0000      .28011    .37831
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on D                       |
| Selection rule is: Observations with D        =  1         |
| Results of selection:                                      |
|                   Data points     Sum of weights           |
| Data set              3000              3000.0             |
| Selected sample       1444              1444.0             |
+------------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step     least squares regression ...........
LHS=Y        Mean                 =         .26107
             Standard deviation   =        1.81942
             Number of observs.   =           1444
Model size   Parameters           =              4
             Degrees of freedom   =           1440
Residuals    Sum of squares       =        1992.04
             Standard error of e  =        1.17617
Fit          R-squared            =         .58181
             Adjusted R-squared   =         .58094
Standard error corrected for selection     1.20443
Correlation of disturbance in regression
and Selection Criterion (Rho)     =         .27787
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.03295         .09311     -.35  .7234     -.21544    .14953
      X1|     .99532***      .03211    31.00  .0000      .93238   1.05825
      X2|     .98370***      .03033    32.44  .0000      .92426   1.04314
  LAMBDA|     .33468***      .11660     2.87  .0041      .10614    .56322
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:  11 iterations. Status=0, F=    4178.508
-----------------------------------------------------------------------------
```

```
ML Estimates of Selection Model
Dependent variable                     Y
Log likelihood function     -4178.50797
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selection (probit) equation for D
Constant|    -.04739**       .02377    -1.99  .0462     -.09397   -.00080
      Z1|     .34587***      .02560    13.51  .0000      .29570    .39605
      Z2|     .32698***      .02463    13.28  .0000      .27872    .37525
        |Corrected regression, Regime 1
Constant|    -.03440         .09333     -.37  .7124     -.21732    .14852
      X1|     .99600***      .03179    31.33  .0000      .93369   1.05832
      X2|     .98376***      .03123    31.50  .0000      .92255   1.04497
SIGMA(1)|    1.20470***      .03039    39.64  .0000     1.14514   1.26426
RHO(1,2)|     .27940***      .09379     2.98  .0029      .09556    .46323
--------+--------------------------------------------------------------------
Random Coefficients  SelctREM Model
Dependent variable                     Y
Log likelihood function     -3142.96683
Sample is 15 pds and    200 individuals
Sample selection with random effects
Simulation based on    100 random draws
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selection corrected regression parameters
      X1|    1.01804***      .01769    57.56  .0000      .98338   1.05271
      X2|    1.01922***      .01609    63.33  .0000      .98768   1.05076
        |Correlation between regression and probit
      X1|    -.03590         .06865     -.52  .6010     -.17044    .09865
      X2|    -.40206***      .06548    -6.14  .0000     -.53040   -.27373
        |Disturbance standard deviation
      Z1|     .49125***      .02825    17.39  .0000      .43589    .54661
      Z2|     .45698***      .03124    14.63  .0000      .39575    .51822
        |Correlation between regression and probit
      Z1|     .31623***      .10835     2.92  .0035      .10387    .52860
      Z2|     .04251         .09871      .43  .6667     -.15096    .23597
        |Means for random parameters
One_Regr|    -.07960***      .02465    -3.23  .0012     -.12791   -.03128
One_Prbt|    -.03615         .02628    -1.38  .1689     -.08766    .01535
        |Scale parameters for dists. of random parameters
sOne_Reg|    1.08397***      .01613    67.21  .0000     1.05236   1.11558
sOne_Prb|    1.14806***      .03847    29.84  .0000     1.07266   1.22346
        |Disturbance standard deviation
   Sigma|    1.42515***      .01568    90.88  .0000     1.39441   1.45588
        |Correlation between regression and probit
     Rho|     .66677***      .03376    19.75  .0000      .60061    .73294
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E53.4.6 Technical Details on RE Selection Models

The log likelihood function for one group for the sample selection model is built up from the structure

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + c_i,\ \ \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated}$$

$$z_{it}^* = \boldsymbol{\alpha}'\mathbf{w}_{it} + u_{it} + d_i$$

$$z_{it} = 1(z_{it}^* > 0),\ u_{it} \sim N[0,1]$$

The contribution of the ith group to the log likelihood (which is then summed) is

$$\log L_i \mid c_i,d_i = \sum\nolimits_{z_{it}=0} \log \boldsymbol{\Phi}\left(-d_i - \boldsymbol{\alpha}'\mathbf{w}_{it}\right) + \sum\nolimits_{z_{it}=1} \left[\begin{array}{l} \dfrac{-\log 2\pi}{2} - \log\sigma - \dfrac{(y_{it} - c_i - \boldsymbol{\beta}'\mathbf{x}_{it})^2}{2\sigma^2} + \\ \log\boldsymbol{\Phi}\left[\dfrac{(d_i + \boldsymbol{\alpha}'\mathbf{w}_{it}) + (\rho/\sigma)(y_{it} - c_i - \boldsymbol{\beta}'\mathbf{x}_{it})}{\sqrt{1-\rho^2}}\right] \end{array}\right]$$

We reparameterize the log likelihood as follows:

$$\theta = 1/\sigma$$

$$\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta}$$

$$\tau = \rho / \sqrt{1-\rho^2}$$

We also isolate the two constant terms, α_0 and β_0 (γ_0 after the Olsen normalization) so that in the formulation below, the slope vectors do not contain constant terms. We also allow the group means for the nonconstant variables in **w** and **x** to appear in the vectors below, but there is no need to note them in particular in the derivation, so we leave them implicit. With these reparameterizations, the function becomes

$$\log L_i \mid c_i, d_i = \sum\nolimits_{z_{it}=0} \log\boldsymbol{\Phi}\left(-\alpha_0 - d_i - \boldsymbol{\alpha}'\mathbf{w}_{it}\right) + \sum\nolimits_{z_{it}=1} \left(\begin{array}{l} \dfrac{-\log 2\pi}{2} + \log\theta - \dfrac{(\theta y_{it} - (\gamma_0 + c_i) - \boldsymbol{\gamma}'\mathbf{x}_{it})^2}{2} \\ + \log\boldsymbol{\Phi}\left\{\begin{array}{l} \sqrt{1+\tau^2}\left[(\alpha_0 + d_i) + \boldsymbol{\alpha}'\mathbf{w}_{it}\right] \\ +\tau\left(\theta y_{it} - (\gamma_0 + c_i) - \boldsymbol{\gamma}'\mathbf{x}_{it}\right) \end{array}\right\} \end{array}\right)$$

The treatment effects model removes the first term and changes the sign of the argument in the CDF in the second term when $z_{it} = 0$, but no other changes are necessary. The unconditional log likelihood is found by integrating out the effects, c_i and d_i which we do with the simulation procedure described in Section R24.7. This is the form of our random parameters model as described there, in this case with exactly two random parameters. The original parameters are recovered after estimation, with standard errors obtained via the delta method.

E53.5 Random Parameters Sample Selection Models

The random parameters form of the sample selection model contains several structural equations.

Regression

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it} \quad \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated,}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}_\beta \mathbf{f}_i + \boldsymbol{\Gamma}_\beta \mathbf{v}_i$$

Selection Mechanism

$$z_{it}^* = \boldsymbol{\alpha}_i'\mathbf{w}_{it} + u_{it},$$

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \boldsymbol{\Delta}_\alpha \mathbf{g}_i + \boldsymbol{\Gamma}_\alpha \mathbf{h}_i,$$

$$z_{it} = 1(z_{it}^* > 0), \; u_{it} \sim N[0,1]$$

Observation Mechanism

$$y_{it}, \mathbf{x}_{it} \quad \text{observed when } z_{it} = 1$$

‘Selectivity’

$$(\varepsilon_{it},u_{it}) \sim N[(0,0),(\sigma^2,1,\rho\sigma)], \; \text{Corr}[\varepsilon_{it},u_{it}] = \rho$$

Our implementation of this model is the same as the one with all nonrandom parameters. Although the model here is fit by maximum likelihood, it is fit in two steps in the fashion of the two step least squares estimator described earlier in this chapter. In the first step, the probit model is fit and the results are stored for use by the selection model. In the second, the regression with selection is fit conditionally on the first step estimation.

NOTE: This model is fit by maximum simulated likelihood, not by two step least squares. There is no ‘lambda’ variable, $\phi(...)/\Phi(...)$ created or used during the estimation, so no coefficient for this variable will appear in the results.

E53.5.1 Treatment Effects

The ‘treatment effects’ model, in which z_{it} appears in the regression and all observations are used,

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \gamma_i z_{it} + \varepsilon_{it} \quad \varepsilon_{it} \sim N[0,\sigma^2], \text{ nonautocorrelated}$$

is requested as in the cross section case by just including *z* in the Rhs list for the regression (if appropriate) and adding the request to the command:

; All

E53.5.2 Commands

This must be fit in two parts as shown below: The probit model need not be a random parameters model; it can be fit as a standard model with nonrandom parameters if desired.

PROBIT ; Lhs = ... ; Rhs = ... ; Hold
[; RPM ; Pds = ... ; Fcn = ... ; Pts = ...] $
SELECT ; Lhs = ... ; Rhs = ...
; RPM ; Pds = ... ; Fcn = ... $

Zabel's modification is requested by adding

; Means

to the **SELECT** command. All other options for the random parameters are available, including

; Correlated	to allow random parameters to be freely correlated
; AR1	for autoregressive random effects
; RPM = **list**	of variables if means of parameters are heterogeneous
; Pts = n	for the number of replications
; Halton	to use Halton draws

and so on. (Details appear in Chapter R24.) Fitted values and residuals may be computed requesting

; Par	to keep individual specific parameter estimates.
; Keep = name	to retain fitted values
; Res = name	to retain residuals

These are computed using individual specific coefficient vectors.

E53.5.3 Results

The results retained by this estimator are

Matrices:	*b*	= full coefficient vector
	varb	= full estimated asymptotic covariance matrix
	beta_i	= individual specific parameters, if **; Par** is requested.
Scalars:	*nreg*	= total number of observations
	kreg	= number of parameters estimated
	logl	= log likelihood
	s	= estimate of σ
	rho	= estimate of ρ

E53.5.4 Application

There are many possible variants of the random parameters model. The following illustrates the simplest case, in which the template sample selection model is specified and coefficients in the selection model are random and uncorrelated.

```
PROBIT    ; Lhs  = d   ; Rhs = one,z1,z2 ; Hold ; RPM
          ; Pds  = 15 ; Fcn = one(n), z1(n), z2(n) ; Pts = 25 ; Halton $
SELECT    ; Lhs  = y   ; Rhs = one,x1,x2 ; RPM
          ; Pds  = 15 ; Fcn = one(n), x1(n), x2(n) ; Pts = 25 ; Halton $
```

```
-----------------------------------------------------------------------------
Probit   Regression Start Values for D
Dependent variable                   D
Log likelihood function     -1895.34311
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.04731**       .02376    -1.99  .0465     -.09389   -.00073
      Z1|     .34436***      .02537    13.57  .0000      .29464    .39408
      Z2|     .32921***      .02505    13.14  .0000      .28011    .37831
--------+--------------------------------------------------------------------
Random Coefficients  Probit   Model
Dependent variable                   D
Log likelihood function     -1498.85939
Restricted log likelihood   -1895.34311
Sample is 15 pds and    200 individuals
PROBIT (normal)  probability model
Simulation based on     25 Halton draws
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|    -.08282***      .02828    -2.93  .0034     -.13825   -.02739
      Z1|     .48350***      .03235    14.95  .0000      .42010    .54689
      Z2|     .47184***      .03281    14.38  .0000      .40753    .53615
        |Scale parameters for dists. of random parameters
Constant|    1.15242***      .04112    28.03  .0000     1.07183   1.23301
      Z1|     .12805***      .03142     4.07  .0000      .06646    .18964
      Z2|     .06748**       .03173     2.13  .0334      .00529    .12968
--------+--------------------------------------------------------------------
Ordinary     least squares regression ...........
(Results omitted)
-----------------------------------------------------------------------------
Random Coefficients  Selection  Model
Simulation based on     25 Halton draws
Standard errors corrected for 2 step est.
Selection effects model based on D
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|     .01375         .42229      .03  .9740     -.81393    .84142
      X1|    1.01654***      .03808    26.69  .0000      .94190   1.09118
      X2|    1.01010***      .01620    62.37  .0000      .97835   1.04184
        |Scale parameters for dists. of random parameters
Constant|     .96273***      .01561    61.68  .0000      .93213    .99332
      X1|     .04740**       .02094     2.26  .0236      .00636    .08843
      X2|     .05047***      .01726     2.92  .0034      .01665    .08430
        |Disturbance standard deviation
   Sigma|    1.43616***      .15245     9.42  .0000     1.13736   1.73497
        |Correlation between regression and probit
     Rho|     .56156         .52556     1.07  .2853     -.46853   1.59164
--------+--------------------------------------------------------------------
```

E53.5.5 Technical Details on the RP Selection Model

The log likelihood function for one group for the sample selection model is built up from the general random parameters structure

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it} \quad \varepsilon_{it} \sim \mathrm{N}[0,\sigma^2], \text{ nonautocorrelated,}\\
\boldsymbol{\beta}_i &= \boldsymbol{\beta} + \boldsymbol{\Delta}_\beta \mathbf{f}_i + \boldsymbol{\Gamma}_\beta \mathbf{v}_i\\
z_{it}^* &= \boldsymbol{\alpha}_i'\mathbf{w}_{it} + u_{it},\\
\boldsymbol{\alpha}_i &= \boldsymbol{\alpha} + \boldsymbol{\Delta}_\alpha \mathbf{g}_i + \boldsymbol{\Gamma}_\alpha \mathbf{h}_i,\\
z_{it} &= 1(z_{it}^* > 0),\ u_{it} \sim \mathrm{N}[0,1]
\end{aligned}
$$

The contribution of the *i*th group to the log likelihood (which is then summed) is

$$
\log L_i \mid \mathbf{v}_i,\mathbf{h}_i = \sum\nolimits_{z_{it}=0} \log \Phi\left(-\boldsymbol{\alpha}_i'\mathbf{w}_{it}\right) + \sum\nolimits_{z_{it}=1} \left(\begin{array}{l} \dfrac{-\log 2\pi}{2} - \log\sigma - \dfrac{(y_{it}-\boldsymbol{\beta}_i'\mathbf{x}_{it})^2}{2\sigma^2} + \\ \log\Phi\left[\dfrac{(\boldsymbol{\alpha}_i'\mathbf{w}_{it}) + (\rho/\sigma)(y_{it}-\boldsymbol{\beta}_i'\mathbf{x}_{it})}{\sqrt{1-\rho^2}}\right] \end{array} \right)
$$

The ancillary parameters, σ and ρ are assumed to be nonrandom as usual. This model is fit in two steps. In the first, the probit model is estimated as usual for the Heckman procedure or as a random parameters model. The results are retained for later use. If the probit model has been fit as a random parameters model, then the means of the parameter distributions are retained for use in the second step.

We then reparameterize the log likelihood as follows:

θ = $1/\sigma$

$\boldsymbol{\gamma}_i$ = $(1/\sigma)\boldsymbol{\beta}_i$ (this also scales $\boldsymbol{\beta}$, $\boldsymbol{\Delta}_\beta$ and $\boldsymbol{\Gamma}_\beta$)

τ = $\rho / \sqrt{1-\rho^2}$

a_{it} = $\boldsymbol{\alpha}'\mathbf{w}_{it}$ (where $\boldsymbol{\alpha}$ is the standard probit estimates or the means of the random parameters in the random parameters probit model.)

With these reparameterizations, the function becomes

$$\log L_i \mid \mathbf{v}_i,\mathbf{h}_i = \sum_{z_{it}=0} \log \boldsymbol{\Phi}\left(-a_{it}\right) + \sum_{z_{it}=1} \left(\begin{array}{l} \dfrac{-\log 2\pi}{2} + \log\theta - \dfrac{(\theta y_{it} - \boldsymbol{\gamma}_i'\mathbf{x}_{it})^2}{2} + \\ \log\boldsymbol{\Phi}\left\{\sqrt{1+\tau^2}\left[a_{it}\right] + \tau\left(\theta y_{it} - \boldsymbol{\gamma}_i'\mathbf{x}_{it}\right)\right\} \end{array} \right)$$

After the normalization, a_{it} (and the first term in the function) becomes irrelevant to the solution. The second term is then maximized using the template form of the random parameters model discussed in Chapter R24. When estimation is complete, the asymptotic covariance matrix is obtained by the delta method. Finally, the Murphy and Topel procedure is used to account for the presence of the estimated parameters in a_{it}. This particular form of the model is a fairly straightforward form of the random parameters structure.

E53.6 FIML Estimator for the RP Selection Model

The preceding section describes a two step random parameters model estimator. At the first step, the probit model is constructed. This may be a random parameters model or a fixed parameters model. At the second step, the random parameters sample selection model is estimated, taking the estimated probit model as fixed. The log likelihood that is maximized is

$$\log L_i \mid \mathbf{v}_i,\mathbf{h}_i = \sum_{z_{it}=0} \log \boldsymbol{\Phi}\left(-a_{it}\right) + \sum_{z_{it}=1} \left(\begin{array}{l} \dfrac{-\log 2\pi}{2} + \log\theta - \dfrac{(\theta y_{it} - \boldsymbol{\gamma}_i'\mathbf{x}_{it})^2}{2} + \\ \log\boldsymbol{\Phi}\left\{\sqrt{1+\tau^2}\left[a_{it}\right] + \tau\left(\theta y_{it} - \boldsymbol{\gamma}_i'\mathbf{x}_{it}\right)\right\} \end{array} \right)$$

where $a_{it} = \boldsymbol{\alpha}'\mathbf{w}_{it}$ from the probit model. Thus, a_{it} is taken as data, and α is not reestimated.

You may also estimate the full model with all parameters random simultaneously. The log likelihood function that is maximized is

$$\log L_i \mid \mathbf{v}_i,\mathbf{h}_i \;=\; \sum\nolimits_{z_{it}=0} \log\Phi\left(-\boldsymbol{\alpha}_i'\mathbf{w}_{it}\right) \;+$$

$$\sum\nolimits_{z_{it}=1}\left(\begin{array}{l}\dfrac{-\log 2\pi}{2} - \log\sigma - \dfrac{(y_{it}-\boldsymbol{\beta}_i'\mathbf{x}_{it})^2}{2\sigma^2} + \\ \log\Phi\left[\dfrac{(\boldsymbol{\alpha}_i'\mathbf{w}_{it}) + (\rho/\sigma)(y_{it}-\boldsymbol{\beta}_i'\mathbf{x}_{it})}{\sqrt{1-\rho^2}}\right]\end{array}\right)$$

All location parameters in both the probit model and in the regression model can be modeled as random. As before, we use the Olsen transformation to simplify the estimation;

$$\theta \;=\; 1/\sigma$$

$$\boldsymbol{\gamma}_i \;=\; (1/\sigma)\boldsymbol{\beta}_i \;\; \text{(this also scales } \boldsymbol{\beta}, \boldsymbol{\Delta}_\beta \text{ and } \boldsymbol{\Gamma}_\beta\text{)}$$

$$\tau \;=\; \rho \,/\, \sqrt{1-\rho^2}$$

This model is essentially the same as the one in the previous section. The difference is that the parameters of the distribution of $\boldsymbol{\alpha}_i$ are reestimated – it is a full information estimator. The command is

```
PROBIT or LOGIT ; … ; Hold $  (as usual for selection models)
SELECT          ; MLE ; Lhs = y,d
                ; Rhs = Rhs in the selection regression
                ; Rh2 = Rhs in the binary variable (probit or logit) equation
                ; RPM … as usual
                ; Pds  = setting is optional, 1 period is the default
                ; Pts  = setting if desired
                ; Halton if desired
                ; Fcn = settings for random parameters with
                        (type) for the regression parameters
                        [type] for the binary choice model parameters $
```

E53.7 The Hausman and Wise Attrition Model

The recent literature on sample selection contains numerous analyses of two period models, such as Kyriazidou (1997). They tend to focus on non- and semiparametric analyses. An early parametric contribution of Hausman and Wise (1979) is considered here. The model is specifically a two period model of attrition, which would seem to characterize many of the studies suggested in the current literature (which is why we consider it here). The model formulation is a two period random effects formulation:

$$y_{i1} = \mathbf{x}_{i1}'\boldsymbol{\beta} + \varepsilon_{i1} + u_i \quad \text{(first period regression)}$$

$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta} + \varepsilon_{i2} + u_i \quad \text{(second period regression).}$$

Attrition is likely in the second period (to begin the study, the individual must have been observed in the first period). The authors suggest that the probability that an observation is made in the second period varies with the value of y_{i2} as well as some other variables,

$$z_{i2}^* = \delta y_{i2} + \mathbf{x}_{i2}'\boldsymbol{\theta} + \mathbf{w}_{i2}'\boldsymbol{\alpha} + v_{i2.}$$

Attrition occurs if $z_{i2}^* \leq 0$, which produces a probit model,

$$z_{i2} = 1(z_{i2}^* > 0) \qquad \text{(attrition indicator observed in period 2).}$$

An observation is made in the second period if $z_{i2} = 1$, which makes this an early version of the familiar sample selection model. The reduced form of the observation equation is

$$\begin{aligned} z_{i2}^* &= \mathbf{x}_{i2}'(\delta\boldsymbol{\beta} + \boldsymbol{\theta}) + \mathbf{w}_{i2}'\boldsymbol{\alpha} + \delta\varepsilon_{i2} + v_{i2} \\ &= \mathbf{x}_{i2}'\boldsymbol{\pi} + \mathbf{w}_{i2}'\boldsymbol{\alpha} + h_{i2} \\ &= \mathbf{r}_{i2}'\boldsymbol{\gamma} + h_{i2}. \end{aligned}$$

The variables in the probit equation are all those in the second period regression plus any additional ones dictated by the application. The estimable parameters in this model are $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\sigma^2 = \text{Var}[\varepsilon_{it} + u_i]$ and two correlation coefficients,

$$\rho_{12} = \text{Corr}[\varepsilon_{i1} + u_i, \varepsilon_{i2} + u_i] = \text{Var}[u_i] / \sigma^2,$$

and

$$\rho_{23} = \text{Corr}[h_{i2}, \varepsilon_{i2} + u_i].$$

All disturbances are assumed to be normally distributed. (Readers are referred to the paper for motivation and details on this specification.)

The authors propose a full information maximum likelihood estimator. The estimator described here uses two steps. The parameters of the probit model are estimated first by maximum likelihood. Then the remaining parameters are estimated conditionally on these first step estimates. The Murphy and Topel adjustment is made after the second step. Further details are given at the end of this section.

E53.7.1 Commands

The Hausman and Wise estimator is obtained as follows: The data are not set up as a panel. All data appear on a single line, as one observation. There are three dependent variables and three sets of independent variables, as shown below.

```
PROBIT    ; Lhs = z
          ; Rhs = variables in probit and in second regression
          ; Hold $
SELECT    ; Lhs = first period y, second period y
          ; Rhs = first period regression
          ; Rh2 = second period regression $
```

NOTE: The Rhs and Rh2 lists must contain the same number of variables. These are two sets of observations on the same variables.

Data on y_{i2} and x_{i2} can be coded as missing values or anything else for observations which have attrition ($z = 0$). Values there will be ignored.

Other options are the standard ones for nonlinear optimization. There are no fitted values or marginal effects produced for this estimator, however. The results retained by the estimator are

Matrices:	*b*	= estimate of β
	varb	= estimate of asymptotic covariance matrix
Scalars:	*logl*	= log likelihood
	nreg	= number of observations
	*rho*12	= estimate of ρ_{12}
	*rho*23	= estimate of ρ_{23}
	s	= estimate of σ

E53.7.2 Application

We used the data from the previous example, and created a second period of data.

```
SAMPLE    ; 1-3000 $
CALC      ; Ran(12345) $
MATRIX    ; ai = Rndm(200) ; ci = Rndm(200) $
CREATE    ; i = Trn(15,0) $
CREATE    ; u = Rnn(0,1) ; e = .5*u+.5*Rnn(0,1) $
CREATE    ; z1 = Rnn(0,1) ; z2 = Rnn(0,1)
          ; d = (.5*z1+.5*z2+ai(i)+u) > 0 $
CREATE    ; x1 = Rnn(0,1) ; x2 = Rnn(0,1)
          ; y = x1+x2+ci(i)+e $
CREATE    ; x1a = Rnn(0,1) ; x2a = Rnn(0,1)
          ; ya = x1a+x2a+e+Rnn(0,1) $
PROBIT    ; Lhs = d ; Rhs = one,z1,z2,x1a,x2a ; Hold $
SELECT    ; Lhs = y,ya ; Rhs = one,x1,x2 ; Rh2 = one,x1a,x2a $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                     D
Log likelihood function     -1892.00484
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.04977**       .02381    -2.09  .0366     -.09644   -.00310
      Z1|     .34466***      .02540    13.57  .0000      .29489    .39443
      Z2|     .32991***      .02508    13.15  .0000      .28076    .37907
     X1A|    -.03439         .02418    -1.42  .1549     -.08177    .01300
     X2A|     .05144**       .02402     2.14  .0322      .00436    .09853
--------+--------------------------------------------------------------------
OLS Start values for attrition model..............
Ordinary     least squares regression ............
LHS=Y        Mean                 =        -.00870
             Standard deviation   =        1.84943
             Number of observs.   =           3000
Model size   Parameters           =              3
             Degrees of freedom   =           2997
Residuals    Sum of squares       =        .000000
             Standard error of e  =         .00000
Fit          R-squared            =         .58461
             Adjusted R-squared   =         .58433
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.01845         .02177     -.85  .3969     -.06113    .02423
      X1|    1.01748***      .02203    46.19  .0000      .97430   1.06065
      X2|    1.00345***      .02162    46.41  .0000      .96107   1.04583
--------+--------------------------------------------------------------------
Hausman and Wise Attrition Model
Dependent variable                     Y
Log likelihood function     -8881.37248
Estimation based on N =   3000, K =   6
First  period dependent variable =Y
Second period dependent variable =YA
Attrition indicator is D
Data means are for first period
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.04316**       .02098    -2.06  .0397     -.08429   -.00203
      X1|     .99566***      .02099    47.43  .0000      .95452   1.03680
      X2|    1.00692***      .02096    48.03  .0000      .96583   1.04801
   Sigma|    1.20586***      .01385    87.10  .0000     1.17872   1.23299
Rho(1,2)|     .34237***      .02256    15.18  .0000      .29816    .38659
Rho(2,3)|     .33678***      .03151    10.69  .0000      .27502    .39853
--------+--------------------------------------------------------------------
```

E53.7.3 Technical Details for the Hausman and Wise Attrition Model

The individual terms in the log likelihood for the model are

$$\text{Log } L_i = \frac{-\log 2\pi}{2} - \log\sigma - \frac{(y_{i1} - \mathbf{x}_{i1}'\boldsymbol{\beta})^2}{2\sigma^2} +$$

$$(z_{i2}) \times \left[\begin{array}{c} -\log\sqrt{2\pi\sigma^2(1-\rho_{12}^2)} - \dfrac{\left[(y_{i2} - \rho_{12}y_{i1}) - (\mathbf{x}_{i2} - \rho_{12}\mathbf{x}_{i1})'\boldsymbol{\beta}\right]^2}{2\sigma^2(1-\rho_{12}^2)} \\ + \log\boldsymbol{\Phi}\left[\dfrac{\mathbf{r}_{i2}'\delta + (\rho_{23}/\sigma)(y_{i2} - \mathbf{x}_{i2}'\boldsymbol{\beta})}{\sqrt{(1-\rho_{23}^2)}}\right] \end{array} \right] +$$

$$(1 - z_{i2}) \times \left[\log\left(1 - \boldsymbol{\Phi}\left[\frac{\mathbf{r}_{i2}'\boldsymbol{\delta} + (\rho_{12}\rho_{23}/\sigma((y_{i1} - \mathbf{x}_{i1}'\boldsymbol{\beta})}{\sqrt{(1-\rho_{12}\rho_{23}^2)}}\right]\right)\right]$$

Reparameterization of the log likelihood brings some large simplification. First, $\boldsymbol{\delta}$ has been estimated at the first step, so we carry only

$$a_{i2} = \mathbf{r}_{i2}'\boldsymbol{\delta}$$

in the log likelihood. Then, let c = -.5log2π and

$$\theta = 1/\sigma,$$

$$\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta},$$

$$T_{12} = \sqrt{1-\rho_{12}^2}\,, \qquad \tau_{12} = \rho_{12} / T_{12}$$

$$T_{23} = \sqrt{1-\rho_{23}^2}\,, \qquad \tau_{12} = \rho_{123} / T_{23}$$

$$T_{13} = \sqrt{1+\tau_{12}^2+\tau_{23}^2}\,, \quad \tau_{13} = \tau_{12}\,\tau_{23} / T_{13}$$

$$e_{i1} = \theta\, y_{i1} - \mathbf{x}_{i1}'\boldsymbol{\gamma}$$

$$e_{i2} = \theta\, y_{i2} - \mathbf{x}_{i2}'\boldsymbol{\gamma}$$

$$U_i = T_{12}\, e_{i2} - \tau_{12}\, e_{i1}$$

$$V_i = T_{23}a_{i2} + \tau_{13}e_{i2}$$

The log likelihood for observation i is now

$$\log L_i = c + \log\theta - .5e_{i1}^2 + z_{i2}(\, c + \log\theta - \log T_{12} - .5{*}U_i^2 + \log\Phi(V_i)] + (1-z_{i2})\log\Phi(-V_i)$$

This is maximized with respect to $\boldsymbol{\gamma}$, θ, τ_{12} and τ_{23}. Derivatives are complicated, but use familiar results for the normal distribution. The BHHH estimator is used for the asymptotic covariance matrix. The original parameters and their asymptotic covariance matrix are recovered using the delta method.

E54: Alternative Sample Selection Models

E54.1 Introduction

Many variants of the ‘sample selection’ model can be estimated with *LIMDEP*. (See Heckman (1979), Maddala (1983) and Greene (2011) for further discussion.) The basic structure is

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

$$z^* = \boldsymbol{\alpha}'\mathbf{w} + u,$$

$$\varepsilon,u \sim N[0,0,\sigma_\varepsilon^2,\sigma_u^2,\rho].$$

A bivariate classical normal (seemingly unrelated) regressions model applies to the structural equations. The standard deviations are σ_ε and σ_u, and the covariance is $\rho\sigma_\varepsilon\sigma_u$. If the data were randomly sampled from this bivariate population, the parameters could be estimated by least squares, or GLS combining the two equations. However, z^* is not observed. Its observed counterpart is z, which is determined by

$$z = 1 \text{ if } z^* > 0$$

and

$$z = 0 \text{ if } z^* \leq 0.$$

Moreover, values of y and $\mathbf{x}$ are only observed when z equals one. Thus, the model is two steps removed from the two equations seemingly unrelated regressions which would be simple to estimate. The essential feature of the model is that under the sampling rule, $E[y|\mathbf{x},z = 1]$ is not a linear regression. The development below presents estimators for the class of essentially nonlinear models that emerge from this specification.

This is the simplest form of this model. Several variants and estimators were considered in Chapter E52. Panel data estimators were developed in Chapter E53. In this chapter, we develop a number of models in which the first, linear equation, $y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$, is replaced with a nonlinear model in which the density of the random variable, rather than its conditional mean are specified;

$$f(y|\mathbf{x},\varepsilon) = g(y,\mathbf{x}, \varepsilon,\boldsymbol{\beta}).$$

The other assumptions are the same. The model is extended to a variety of settings, such as binary and multinomial choice models, count data models and a stochastic frontier model.

E54.2 Probit Model with Selection

In the bivariate probit setting (see Section E33.2), data on y_1 might be observed only when y_2 equals one. For example, in modeling loan defaults with a sample of applicants, default will only occur among applicants who are granted loans. Thus, in a bivariate probit model for the two outcomes, the observed default data are nonrandomly selected from the set of applicants. The model that might be used is

$$
\begin{aligned}
z_{i1} &= \boldsymbol{\beta}'\mathbf{x}_{i1} + \varepsilon_{i1},\ y_{i1} = \text{sgn}(z_{i1}),\\
z_{i2} &= \boldsymbol{\beta}'\mathbf{x}_{i2} + \varepsilon_{i2},\ y_{i2} = \text{sgn}(z_{i2}),\\
\varepsilon_{i1},\varepsilon_{i2} &\sim \text{BVN}(0,0,1,1,\rho),\\
(y_{i1},\mathbf{x}_{i1}) &\ \text{is observed only when } y_{i2} = 1.
\end{aligned}
$$

This is a type of sample selectivity model. The model was proposed by Wynand and van Praag (1981). An extensive application which uses choice based sampling as well is Boyes, Hoffman, and Low (1989). (See also Greene (1992 and 2011).) The sample selection model is obtained by adding **; Selection** to the **BIVARIATE PROBIT** (or just **BIVARIATE**) command. This model is fit in a single step, using full information maximum likelihood. Use

```
BIVARIATE   ; Lhs = y1, y2   (selection variable is second)
            ; Rhs = x1       (variables in selected model)
            ; Rh2 = x2       (variables in selection equation)
            ; Selection $
```

All other options and specifications are the same as for the model without selection. Except for the diagnostic table which indicates that this model has been chosen, the results for the selection model are exactly the same as for the basic model.

E54.2.1 Choice Based Sampling

Like other discrete choice models, you may use a choice based sampling correction with this model. You must provide a weighting variable which for this model will take only three different values. In each case, the weight is

$w_i(z_{i1},z_{i2})$ = *population proportion* / *sample proportion*.

The three cells in your data set for this selection model are $z_2 = 0$, ($z_2 = 1,z_1 = 0$) and ($z_2 = 1,z_1 = 1$). Your command is modified to account for the weighting as follows:

```
BIVARIATE ; Lhs  = y1, y2     (selection variable is second)
; Rhs = x1       (variables in selected model)
; Rh2 = x2       (variables in selection equation)
; Wts = wi
; Choice Based Sampling
; Selection $
```

E54.2.2 Application

The foregoing was applied in Greene (1992). The study analyzed usage and default patterns for a sample of individuals applying for and using a major credit card. Descriptive statistics for a subset of the variables in this data set of 13,444 observations (provided as data file credit.lpj) are as follows.

Variable	Mean	Standard Deviation	Minimum	Maximum	Cases
Dummy variable for whether individual holds the credit card					
CARDHLDR|	.780943	.413623	.000000	1.00000	13444
Dummy variable for whether individual cardholder defaulted on the credit card					
DEFAULT |	.074085	.261919	.000000	1.00000	13444
Number of major and minor derogatory reports on credit card usage					
MAJORDRG|	.462809	1.43272	.000000	22.0000	13444
MINORDRG|	.290539	.767620	.000000	11.0000	13444
Age in years and twelfths of a year when card was applied for					
AGE|	33.54046	10.06261	18.0	88.6667	13444
Income in $ per year – regular income and additional income					
INCOME|	30114.33	15035.36	600.0	99999.0	13444
ADDLINCM|	4126.17	9127.93	.000000	99999.0	13444
Ratio of average yearly expenditure to average total yearly income					
EXP_INC|	.070974	.103922	.0000882	2.03773	13444
Average yearly expenditure using the credit card					
AVGYREXP|	2357.658	3734.952	12.00000	121663.8	13444
Dummy variable indicates whether individual owns or rents their home					
OWNRENT |	.455965	.498076	.000000	1.00000	13444
Dummy variable indicates whether individual is self employed					
SELFEMPL|	.057944	.233646	.000000	1.00000	13444
Number of dependents in household, not including the individual					
DEPDNT |	1.01726	1.27910	.000000	9.00000	13444
Income per dependent, in $10,000 units					
INCPER |	21719.7	13591.2	362.500	150000.	13444
Months residing at current address when applied for the credit card					
CURNTADD|	55.3189	63.0897	.000000	576.000	13444
Dummy variable for whether the individual holds another major credit card					
CREDMAJR|	.813076	.389865	.000000	1.00000	13444
Number of credit accounts active at the time of card application					
TRADACCT|	6.42205	6.10691	.000000	50.0000	13444

The variable *cardhldr* is a binary variable which indicates whether the individual holds the major credit card whose vendor produced the overall data set. The probit equation is used to model *default*. The selection model that arises as the default is only observed for those with *cardhldr* = 1.

This application continues the original analysis. (The specification is different below.) The first set of estimates computes the bivariate probit model with selection. In fact, the sample is choice based. The list below shows the sample and true population proportions and the weights to be applied.

		Sample	Population	Weight
Card holder = 0		0.219	0.768	3.507
Card holder = 1	Default = 1	0.0949	0.0237	0.2497
	Default = 0	0.905	0.208	0.2298

The choice based sampling in these data is fairly drastic. The sample was constructed for the purpose of studying default, so it was heavily skewed toward defaulters, far in excess of observed rates. (We note, in the years since the study was done, the vendor has also drastically increased the acceptance rate.)

```
NAMELIST    ; card = one,age,income,ownrent,selfempl,curntadd $
NAMELIST    ; dflt  = one,income,avgyrexp,depdnt,incper,credmajr,tradacct $
BIVARIATE   ; Lhs = default,cardhldr
            ; Rh1= dflt ; Rh2 = card ; Selection
            ; Summary ; Partial Effects $
```

This set of instructions computes the weights for the choice based sampling estimator.

```
CALC        ; wc0 = 1-Xbr(cardhldr) ; pc0 = .768 $
REJECT      ; cardhldr = 0 $
CALC        ; wc11 = Xbr(default) ; pc11 = .232*.102
            ; wc10 = 1 - wc11 ; pc10 = .232*.898 $
SAMPLE      ; All $
CREATE      ; cbwt = (cardhldr = 0)*pc0 / wc0
                  + (cardhldr = 1)*(default = 1) * pc11 / wc11
                  + (cardhldr = 1)*(default = 0) * pc10 / wc10 $
```

This is the same model, now applying the weights. The results are substantially different, as might be expected.

```
BIVARIATE   ; Lhs = default,cardhldr
            ; Rh1= dflt ; Rh2 = card ; Selection
            ; Wts = cbwt ; Choice Based Sampling $
```

```
--------------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable                DEFCAR
Log likelihood function    -10049.00092
Estimation based on N =  13444, K =  14
Selection model based on CARDHLDR
Selected obs. 10499, Nonselected:  2945
--------+-----------------------------------------------------------------------
 DEFAULT|                  Standard            Prob.      95% Confidence
CARDHLDR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Index    equation for DEFAULT
Constant|    -.93050***      .05904   -15.76  .0000    -1.04621   -.81479
  INCOME|-.12203D-04***   .1999D-05    -6.10  .0000 -.16120D-04  -.82851D-05
AVGYREXP|-.89928D-05*     .4690D-05    -1.92  .0552 -.18185D-04  .19900D-06
  DEPDNT|     .04689**       .02213     2.12  .0341      .00351     .09027
  INCPER|-.12370D-05      .2208D-05     -.56  .5754 -.55649D-05  .30910D-05
CREDMAJR|    -.08457**       .04245    -1.99  .0464     -.16777    -.00136
TRADACCT|    -.01441***      .00299    -4.82  .0000     -.02028    -.00855
        |Index    equation for CARDHLDR
Constant|     .35085***      .04541     7.73  .0000      .26185     .43985
     AGE|    -.00269*        .00138    -1.94  .0519     -.00540     .00002
  INCOME| .16129D-04***   .8858D-06    18.21  .0000  .14393D-04  .17865D-04
 OWNRENT|     .16698***      .02687     6.21  .0000      .11431     .21965
SELFEMPL|    -.33527***      .04967    -6.75  .0000     -.43263    -.23792
CURNTADD|-.33633D-05         .00021     -.02  .9874 -.42220D-03  .41547D-03
        |Disturbance correlation
RHO(1,2)|     .85078        2.63801      .32  .7471    -4.31963    6.02119
--------+-----------------------------------------------------------------------
+---------------------------------+
|  Partial Effects for Ey1|y2=1    |
+----------+----------+----------+
|          |   Direct | Indirect |
| Variable | Efct  x1 | Efct  x2 |
+----------+----------+----------+
|   INCOME |   .00000 |   .00000 |
| AVGYREXP |   .00000 |   .00000 |
|   DEPDNT |   .00776 |   .00000 |
|   INCPER |   .00000 |   .00000 |
| CREDMAJR |  -.01399 |   .00000 |
| TRADACCT |  -.00238 |   .00000 |
|      AGE |   .00000 |   .00008 |
|  OWNRENT |   .00000 |  -.00523 |
| SELFEMPL |   .00000 |   .01051 |
| CURNTADD |   .00000 |   .00000 |
+----------+----------+----------+
```

(The partial effects related to *income* and *incper* are small because of the scale of the variable. The values are shown in the table below.)

```
-----------------------------------------------------------------------------
Partial derivatives of  E[y1|y2=1]  with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Effect shown is total of all parts above.
Estimate of E[y1|y2=1] = .085556
Observations used for means are  All Obs.
Total effects reported = direct+indirect.
--------+--------------------------------------------------------------------
 DEFAULT|     Partial      Standard            Prob.      95% Confidence
CARDHLDR|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  INCOME|-.25246D-05***   .3255D-06    -7.76  .0000 -.31626D-05  -.18865D-05
AVGYREXP|-.14881D-05*     .7756D-06    -1.92  .0551 -.30083D-05  .32180D-07
  DEPDNT|     .00776**      .00366      2.12  .0341      .00058    .01494
  INCPER|-.20468D-06      .3656D-06     -.56  .5756 -.92131D-06  .51195D-06
CREDMAJR|    -.01399**      .00703     -1.99  .0465     -.02777   -.00022
TRADACCT|    -.00238***     .00049     -4.84  .0000     -.00335   -.00142
     AGE| .84248D-04*     .4335D-04     1.94  .0520 -.71561D-06  .16921D-03
 OWNRENT|    -.00523***     .00086     -6.10  .0000     -.00691   -.00355
SELFEMPL|     .01051***     .00157      6.69  .0000      .00743    .01359
CURNTADD| .10539D-06      .6696D-05      .02  .9874 -.13019D-04  .13229D-04
--------+--------------------------------------------------------------------
These  are the  direct marginal  effects.
--------+--------------------------------------------------------------------
 DEFAULT|     Partial      Standard            Prob.      95% Confidence
CARDHLDR|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  INCOME|-.20192D-05***   .3289D-06    -6.14  .0000 -.26639D-05  -.13745D-05
AVGYREXP|-.14881D-05*     .7756D-06    -1.92  .0551 -.30083D-05  .32180D-07
  DEPDNT|     .00776**      .00366      2.12  .0341      .00058    .01494
  INCPER|-.20468D-06      .3656D-06     -.56  .5756 -.92131D-06  .51195D-06
CREDMAJR|    -.01399**      .00703     -1.99  .0465     -.02777   -.00022
TRADACCT|    -.00238***     .00049     -4.84  .0000     -.00335   -.00142
     AGE|        0.0    .....(Fixed Parameter).....
 OWNRENT|        0.0    .....(Fixed Parameter).....
SELFEMPL|        0.0    .....(Fixed Parameter).....
CURNTADD|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
These are the indirect  marginal effects.
--------+--------------------------------------------------------------------
 DEFAULT|     Partial      Standard            Prob.      95% Confidence
E[y1|x,z|      Effect        Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  INCOME|-.50539D-06***   .3307D-07   -15.28  .0000 -.57020D-06  -.44057D-06
AVGYREXP|        0.0    .....(Fixed Parameter).....
  DEPDNT|        0.0    .....(Fixed Parameter).....
  INCPER|        0.0    .....(Fixed Parameter).....
CREDMAJR|        0.0    .....(Fixed Parameter).....
TRADACCT|        0.0    .....(Fixed Parameter).....
     AGE| .84248D-04*     .4335D-04     1.94  .0520 -.71561D-06  .16921D-03
 OWNRENT|    -.00523***     .00086     -6.10  .0000     -.00691   -.00355
SELFEMPL|     .01051***     .00157      6.69  .0000      .00743    .01359
CURNTADD| .10539D-06      .6696D-05      .02  .9874 -.13019D-04  .13229D-04
--------+--------------------------------------------------------------------
```

```
+--------------------------------------------------------------+
| Analysis of dummy variables in the model. The effects are |
| computed using E[y1|y2=1,d=1] - E[y1|y2=1,d=0] where d is |
| the variable. Variances use the delta method.  The effect |
| accounts for all appearances of the variable in the model.|
+--------------------------------------------------------------+
|Variable      Effect   Standard error     t ratio          |
+--------------------------------------------------------------+
 CREDMAJR    -.014613     .007653           -1.909
 OWNRENT     -.005187     .000846           -6.130
 SELFEMPL     .013307     .002578            5.161

+-------------------------------------------------------+
| Joint Frequency Table for Bivariate Probit Model     |
| Predicted cell is the one with highest probability   |
+-------------------------------------------------------+
|                          CARDHLDR                     |
+-------------+-----------------------------------------+
|  DEFAULT    |       0          1         Total     |
|-------------+-------------+------------+------------+
|         0   |         0   |     9503   |     9503   |
|    Fitted   |   (     10) |  ( 10489)  |  ( 10499)  |
|-------------+-------------+------------+------------+
|         1   |         0   |      996   |      996   |
|    Fitted   |   (      0) |  (      0) |  (      0) |
|-------------+-------------+------------+------------+
|      Total  |         0   |    10499   |    10499   |
|    Fitted   |   (     10) |  ( 10489)  |  ( 10499)  |
|-------------+-------------+------------+------------+
| Counts based on  10499 selected of  13444 in sample |
+-------------------------------------------------------+
+----------------------------------------------------------+
| Bivariate Probit Predictions for DEFAULT  and CARDHLDR |
| Predicted cell (i,j) is cell with largest probability  |
| Neither DEFAULT  nor CARDHLDR predicted correctly        |
|                              0 of   13444 observations  |
| Only    DEFAULT  correctly predicted                     |
|         DEFAULT  = 0:      10 of    9503 observations  |
|         DEFAULT  = 1:       0 of     996 observations  |
| Only    CARDHLDR correctly predicted                     |
|         CARDHLDR = 0:       0 of       0 observations  |
|         CARDHLDR = 1:       0 of   10499 observations  |
| Both    DEFAULT  and CARDHLDR correctly predicted        |
|         DEFAULT  = 0 CARDHLDR = 0:       0 of       0  |
|         DEFAULT  = 1 CARDHLDR = 0:       0 of       0  |
|         DEFAULT  = 0 CARDHLDR = 1:    9493 of    9503  |
|         DEFAULT  = 1 CARDHLDR = 1:       0 of     996  |
+----------------------------------------------------------+
```

```
--------------------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               DEFCAR
Weighting variable                 CBWT
Log likelihood function     -6818.41100
Estimation based on N =  13444, K =  14
Selection model based on CARDHLDR
Selected obs. 10499, Nonselected:  2945
Std. errs corrected for choice based sample
--------+-----------------------------------------------------------------------
 DEFAULT|                  Standard            Prob.      95% Confidence
CARDHLDR|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Index    equation for DEFAULT
Constant|   -1.68985       1.98645     -.85  .3949    -5.58323   2.20353
  INCOME|-.60682D-05      .2645D-04    -.23  .8185 -.57905D-04  .45769D-04
AVGYREXP|-.81195D-05      .1246D-04    -.65  .5147 -.32543D-04  .16304D-04
  DEPDNT|     .04496        .05019      .90  .3704     -.05342    .14333
  INCPER|-.11524D-05      .4013D-05    -.29  .7740 -.90180D-05  .67131D-05
CREDMAJR|    -.07196        .12814     -.56  .5744     -.32312    .17919
TRADACCT|    -.01178        .01788     -.66  .5098     -.04682    .02325
        |Index    equation for CARDHLDR
Constant|   -1.27003       1.54792     -.82  .4119    -4.30390   1.76383
     AGE|    -.00239        .05642     -.04  .9662     -.11296    .10818
  INCOME| .15219D-04***   .5646D-05    2.70  .0070  .41521D-05  .26285D-04
 OWNRENT|     .17197        .72795      .24  .8132    -1.25478   1.59873
SELFEMPL|    -.33413       1.22403     -.27  .7849    -2.73319   2.06492
CURNTADD|-.60381D-04        .00306     -.02  .9843 -.60603D-02  .59395D-02
        |Disturbance correlation
RHO(1,2)|     .71564       1.78004      .40  .6877    -2.77317   4.20445
--------+-----------------------------------------------------------------------
```

E54.2.3 Technical Details

The log likelihood for the bivariate probit model with selection is

$$\text{Log - }L = \sum\nolimits_{y_2=1,y_1=1} \log\Phi_2[\boldsymbol{\beta}_1'\mathbf{x}_{i1},\boldsymbol{\beta}_2'\mathbf{x}_{i2},\rho] + \sum\nolimits_{y_2=1,y_1=0} \log\Phi_2[-\boldsymbol{\beta}_1'\mathbf{x}_{i1},\boldsymbol{\beta}_2'\mathbf{x}_{i2},-\rho]$$
$$+ \sum\nolimits_{y_2=0} \log\Phi[-\boldsymbol{\beta}_2'\mathbf{x}_{i2}].$$

The necessary first and second derivatives are given in Section E33.2.9.

NOTE: This is one of several sample selection models estimated by maximum likelihood with *LIMDEP*. In this setting, there is no 'lambda' variable as there is in the regression model with sample selection. Heckman's (1979) selection correction variable applies to the linear regression model estimated with two step least squares, but not generally to models fit by maximum likelihood. For testing for selection effects, the appropriate approach is to test the hypothesis of no effects, which results if ρ equals zero.

NOTE: You may code y_1 as 0.0 for the nonselected (nonobserved) observations in this model. The correct value to use (or ignore) is determined by the program during estimation.

E54.3 Ordered Probit Model

The following describes an ordered probit counterpart to the standard sample selection model. This is only available for the ordered probit specification, not the ordered logit, Gompertz, etc. The structural equations are, first, the main equation, the ordered choice model,

$$
\begin{aligned}
y_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \ \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}), E[\varepsilon_i] = 0, \text{Var}[\varepsilon_i] = 1,\\
y_i &= 0 \text{ if } y_i \le \mu_0,\\
&= 1 \text{ if } \mu_0 < y_i \le \mu_1,\\
&= 2 \text{ if } \mu_1 < y_i \le \mu_2,\\
&\ldots\\
&= J \text{ if } y_i > \mu_{J\text{-}1}.
\end{aligned}
$$

Second is the selection equation, a univariate probit model,

$$
\begin{aligned}
d_i^* &= \boldsymbol{\alpha}'\mathbf{z}_i + u_i,\\
d_i &= 1 \text{ if } d_i^* > 0 \text{ and } 0 \text{ otherwise,}
\end{aligned}
$$

The observation mechanism is

$$
\begin{aligned}
&[y_i,\mathbf{x}_i] \quad \text{is observed if and only if } d_i = 1.\\
&\varepsilon_i, u_i \sim N_2[0,0,1,1,\rho]; \text{ there is ‘selectivity’ if } \rho \text{ is not equal to zero.}
\end{aligned}
$$

This model requires two passes to estimate. In the first, you fit a probit model for the selection variable, d. You then pass these values to the ordered probit model using a standard command for this operation, the **; Hold** parameter in the probit command. The two commands would be as follows: (This model is requested in the same fashion as *LIMDEP*’s other sample selectivity models.) Estimate the first stage probit model and hold the results for next step in the estimation.

```
PROBIT      ; Rhs = z list ; Lhs = d ; Hold $
```

Second, estimate the ordered probit model with selectivity,

```
ORDERED     ; ... as usual ; Selection  $
```

You need not make any other changes in the ordered probit command.

The second step reestimates $\boldsymbol{\alpha}$ from the probit model along with $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$, obtaining a FIML set of estimates for all parameters including ρ. The ordered probit command results in two full rounds of estimation. In the first round, the model is estimated as if there were no selection. This provides the remaining starting values. The starting value for ρ is zero. Then, in the second round, the FIML estimates are computed. This model is rather difficult to estimate, and it is best to allow *LIMDEP* to use its own starting values. (In spite of this, nonconvergence can be a problem. When problems arise, be sure first to check the scaling of the independent variables.)

NOTE: This model is *not* fit by computing a 'lambda' variable, $\lambda_i = \phi(\boldsymbol{\alpha}'\mathbf{z}_i)/ \Phi(\boldsymbol{\alpha}'\mathbf{z}_i)$ from the results of the first step probit and including it in the ordered probit at the second. It is estimated by maximizing the likelihood function shown at the end of this section with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and ρ. There will be no coefficient shown for such a variable in the estimation results, though the estimated ρ is shown.

NOTE: (This is another frequently asked question.) All observations in the sample are used in fitting this model, not just the ones for which $d = 1$. The observations for which $d = 0$ contribute to the probit part of the log likelihood. The remainder contribute both to the probit and the ordered probit.

The **; Rst = ...** and **; CML:** options for imposing restrictions can be used freely with this model to constrain $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The parameter vector is

$$\boldsymbol{\Theta} = [\beta_1,...,\beta_K,\alpha_1,...,\alpha_L,\mu_1,...,\mu_{J-1},\rho].$$

You may also give your own starting values with **; Start = list ...**, though the internal values will usually be preferable.

All results kept for the basic model are also kept; *b* and *varb* still include only $\boldsymbol{\beta}$, but **; Par** adds all of $[\boldsymbol{\mu},\boldsymbol{\alpha},\rho]$ to the parameter vector. This model adds two additional scalars:

rho = estimate of ρ,
varrho = estimate of asymptotic variance of estimated ρ.

NOTE: The estimates of $\boldsymbol{\alpha}$ update the estimates you stored with **; Hold** when you fit the probit model. Thus, for example, if you were to follow your **ORDERED** command immediately with the identical command, the starting values used for $\boldsymbol{\alpha}$ would be the MLEs from the prior ordered probit command, not the ones from the original probit model that you fit earlier. Also, if you were to follow this model command with a selection model command, this estimate of $\boldsymbol{\alpha}$ would be used there, as well.

With the corrected estimates of $[\boldsymbol{\beta},\boldsymbol{\mu}]$ in hand, predictions for this model are computed in the same manner as for the basic model without selection. The only difference is that no prediction for y is computed in the selection model if $d = 0$.

E54.3.1 Application

The following illustrates the model with some simulated data which satisfy the assumptions of the specified model:

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-500 $
CREATE      ; x1 = Rnu(1,4) ; x2 = Rnd(2) - 1 $
CREATE      ; z1 = Rnn(0,1) ; z2 = Rnn(0,1) ; u = Rnn(0,2) $
CREATE      ; d  = (z1 + z2 + u) > 0 $
CREATE      ; e  = u + Rnn(0,3) ; y = 1 + .5 * x1 + 1.2 * x2 + e $
RECODE      ; y ; -25/2.5 = 0 ; 2.501/3 = 1 ; 3.001/4 = 2 ; 4.01/100 = 3 $
PROBIT      ; Lhs = d ; Rhs = one,z1,z2 ; Hold $
ORDERED     ; Lhs = y ; Rhs = one,x1,x2 ; Select ; Partial Effects $
```

This is the initially estimated probit equation. The coefficients below are used as the starting values for the ordered probit with selection. At this point, this is the model that is used in subsequent sample selection models.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                     D
Log likelihood function      -302.36938
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|      .04460         .05935      .75  .4524     -.07173    .16093
      Z1|      .44979***      .06353     7.08  .0000      .32528    .57430
      Z2|      .40994***      .06890     5.95  .0000      .27490    .54499
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

This is the ordered probit model fit without regard to the sample selection issue. These are used as starting values for the MLE. The initial value for ρ is zero.

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                     Y
Log likelihood function      -578.78276
Underlying probabilities based on Normal
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|     -.13761         .24213     -.57  .5698     -.61219    .33696
      X1|      .17417**       .08453     2.06  .0394      .00849    .33986
      X2|      .45932***      .14938     3.07  .0021      .16654    .75210
        | Threshold parameters for index
   Mu(1)|      .14970***      .03905     3.83  .0001      .07316    .22624
   Mu(2)|      .49683***      .06502     7.64  .0000      .36940    .62426
--------+--------------------------------------------------------------------
```

```
+----------------------------------------------------------------------+
|                 CELL FREQUENCIES FOR ORDERED CHOICES                 |
+----------------------------------------------------------------------+
|               Frequency         Cumulative  < =    Cumulative  > =   |
|Outcome       Count    Percent   Count    Percent   Count    Percent  |
|----------- ------- ---------  ------- ---------  ------- ---------  |
|Y=00             76    30.0395       76    30.0395      500  100.0000 |
|Y=01             13     5.1383       89    35.1779      424   69.9605 |
|Y=02             33    13.0435      122    48.2213      411   64.8221 |
|Y=03            131    51.7787      500  100.0000      131   51.7787 |
+----------------------------------------------------------------------+
```

This is the objective; FIML estimates of the ordered probit model and, simultaneously, the probit model. The 'Selection equation' below is the reestimated probit model. This model is stored for use by later sample selection models.

```
Normal exit:  16 iterations. Status=0, F=    573.5784
-----------------------------------------------------------------------------
Ordered Probit Model with Selection.
Dependent variable                     Y
Log likelihood function      -573.57840
--------+--------------------------------------------------------------------
       D|                Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Index function for probability
Constant|    -.54284**        .21559    -2.52  .0118     -.96538   -.12030
      X1|     .13625*         .07301     1.87  .0620     -.00684    .27935
      X2|     .40042***       .13360     3.00  .0027      .13857    .66226
        | Threshold parameters for index
   Mu(1)|     .12773***       .03751     3.41  .0007      .05421    .20124
   Mu(2)|     .42892***       .07020     6.11  .0000      .29134    .56651
        | Selection equation
Constant|     .04349          .05929      .73  .4633     -.07273    .15970
      Z1|     .41957***       .06248     6.72  .0000      .29711    .54202
      Z2|     .44936***       .06940     6.48  .0000      .31335    .58538
        | Cor[u(probit),e(ordered probit)]
Rho(u,e)|     .69491***       .14488     4.80  .0000      .41095    .97888
--------+--------------------------------------------------------------------
Partial effects of variables on P[Y        =  0|D        = 1]
--------+--------------------------------------------------------------------
       D|     Partial      Standard            Prob.      95% Confidence
       Y|      Effect         Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Direct partial effect in ordered choice equation
      X1|    -.03444*         .01807    -1.91  .0567     -.06986    .00098
      X2|    -.10121***       .03236    -3.13  .0018     -.16463   -.03779
        | Indirect partial effect in sample selection equation
      Z1|     .06577***       .01282     5.13  .0000      .04064    .09090
      Z2|     .07044***       .01309     5.38  .0000      .04478    .09610
        | Full partial effect = direct effect + indirect effect
--------+--------------------------------------------------------------------
Partial effects of variables on P[Y        =  1|D        = 1]
--------+--------------------------------------------------------------------
        | Direct partial effect in ordered choice equation
      X1|     .00281          .00348      .81  .4196     -.00401    .00962
      X2|     .00825*         .00470     1.75  .0794     -.00097    .01746
```

```
        | Indirect partial effect in sample selection equation
      Z1|     .01137***       .00259     4.39  .0000      .00630    .01645
      Z2|     .01218***       .00274     4.44  .0000      .00681    .01755
        | Full partial effect = direct effect + indirect effect
--------+-------------------------------------------------------------------
Partial effects of variables on P[Y        =  2|D        = 1]
--------+-------------------------------------------------------------------
        | Direct partial effect in ordered choice equation
      X1|     .00823          .00626     1.31  .1889     -.00405    .02050
      X2|     .02418**        .01006     2.40  .0163      .00445    .04390
        | Indirect partial effect in sample selection equation
      Z1|     .02669***       .00701     3.81  .0001      .01295    .04043
      Z2|     .02859***       .00746     3.83  .0001      .01396    .04322
        | Full partial effect = direct effect + indirect effect
--------+-------------------------------------------------------------------
Partial effects of variables on P[Y        =  3|D        = 1]
--------+-------------------------------------------------------------------
        | Direct partial effect in ordered choice equation
      X1|     .02341*         .01257     1.86  .0627     -.00124    .04805
      X2|     .06879***       .02341     2.94  .0033      .02290    .11467
        | Indirect partial effect in sample selection equation
      Z1|     .05875***       .01898     3.10  .0020      .02155    .09595
      Z2|     .06292***       .02096     3.00  .0027      .02185    .10400
        | Full partial effect = direct effect + indirect effect
--------+-------------------------------------------------------------------
```

E54.3.2 Technical Details for the Selection Model

In the sample selection model, $[\varepsilon,u]$ are assumed to have a bivariate standard normal distribution with correlation ρ. Then, the probabilities in the log likelihood are:

For observations with $d_i = 0$, Prob = Prob[$d = 0$] = univariate normal CDF.

For observations with $d_i = 1$, Prob = Prob[y_i* in particular range and $d = 1 \mid \rho$]

= bivariate normal probability.

The log likelihood for the model with sample selection is

$$\log L = \Sigma_{d=0} \log \Phi(-\boldsymbol{\alpha}'\mathbf{x}_2) + \Sigma_{d=1} \log \{\Phi_2[a_j,\boldsymbol{\alpha}'\mathbf{z},\rho] - \Phi_2[a_{j-1},\boldsymbol{\alpha}'\mathbf{z},\rho]\}$$

where $\Phi(\bullet)$ = standard normal CDF,

$\Phi_2(\bullet,\bullet,\bullet)$ = bivariate standard normal CDF,

$a_j = \mu_j - \boldsymbol{\beta}'\mathbf{x}$,

$a_{j-1} = \mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}$,

and j = the value taken by y_i for that observation.

The same convention used above is maintained for the μs. The first derivatives are tedious but straightforward. They can be derived by applying the formulas given in Chapter E33 for the bivariate probit model. The derivation is a bit simpler here because for the differentiation of the bivariate CDF, q_1 and q_2 are both +1.

E54.4 Poisson and Negative Binomial Regression Models with Selection

Extending the selectivity model to models for counts, such as the Poisson and negative binomial requires a change in approach from the models of the previous sections. Since there is no natural joint normality assumption that ties the count model to the selection model, a different approach is needed. We use the following structure. (See Terza (2010). The mathematical detail for this model is developed in full later in this section.) The Poisson and negative binomial specifications are modified as follows:

$$
\begin{aligned}
z_i^* &= \boldsymbol{\alpha}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1] \\
z_i &= \mathbf{1}(z_i^* > 0) \qquad \text{(probit selection equation)} \\
\lambda_i | \varepsilon_i &= \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) \\
y_i | \varepsilon_i &\sim \text{Poisson}(\lambda_i | \varepsilon_i) \qquad \text{(count model for outcome)} \\
[u_i,\varepsilon_i] &\sim N[(0,1),(1,\rho\sigma,\sigma^2)] \\
y_i,\mathbf{x}_i & \quad \text{are observed only when } z_i = 1.
\end{aligned}
$$

Thus, $y \mid \varepsilon$ is distributed as Poisson with mean (and variance) $E[y|\varepsilon] = \exp(\boldsymbol{\beta}'\mathbf{x} + \varepsilon)$. The distribution in the selected population is nonPoisson, but this does preserve its discreteness. The force of the sample selection is exerted on the mean of the discrete variable (and its variance). The estimator is full information maximum likelihood. (The negative binomial model is considered below.)

In the standard regression framework, the development proceeds by modeling the joint distribution of u_i and the disturbance in the regression model, which would correspond to

$$\varepsilon_i = y_i - E[y_i|\mathbf{x}_i].$$

The familiar Heckman model hinges on joint normality of $[u_i,\varepsilon_i]$, which is clearly untenable here – since y_i is discrete, its deviation from the conditional mean function could not be normally distributed. The approach taken is to introduce the unobservable factors in the mean of the count variable, then use a form of the selection approach to model it through the covariance of u and ε, as is done elsewhere. The change in this model is that the linear techniques used in Chapter E52 are inappropriate here.

E54.4.1 Full Information Maximum Likelihood Estimation

A full information maximum likelihood estimator for the sample selection model is requested with

PROBIT ; Lhs = ... ; Rhs = ... ; Hold $
POISSON ; Lhs = ... ; Rhs = ... ; Selection ; MLE $
or NEGBIN ; Lhs = ... ; Rhs = ... ; Selection ; MLE $

The computations are based on the heterogeneity model of Section E42.3. This must be preceded by the probit model in order to define the full set of variables in the model and to provide the starting values for the iterations.

All options that are useable for the Poisson model are supported here as well, including

Optimization:	**; Maxit = n**	to set maximum restrictions
	; Alg = name	to select algorithm (you generally should not change this)
	; Tlf [= value]	to set tolerance for convergence criteria
	; Output = value	to control intermediate output
	; Hpt = n	to specify number of nodes for Hermite quadrature
Constraints:	**; Rst = list**	to specify fixed value and equality restrictions
	; CML: spec	
	; Test: spec	to define Wald tests
Output:	**; Partial Effects**	
	; Covariance Matrix	to display the estimated asymptotic covariance matrix, same as **; Printvc**
	; List	to display predicted values
	; Keep = name	to retain fitted values
	; Res = name	to retain residuals
	; Parameters	to retain estimates of σ and ρ in *b* and *varb*

and so on for other program options are all supported. Output for this model will include the initial Poisson regression followed by the FIML results, then any optional output you have requested, such as a list of fitted values.

NOTE: This estimator reestimates the parameters of the probit model, and replaces the estimates that were initially retained with **; Hold** on the **PROBIT** command. See the example below.

WARNING: The negative binomial model with sample selection is quite volatile, and without prior scaling of the data (and a good fit of the model and the data), the numerical properties of the estimator appear to be somewhat unstable.

The results that are retained include

Matrices: *b* and *varb*, as usual
include Poisson slopes followed by probit parameters
σ then ρ with **; Parameters** option

Scalars:
logl = log likelihood
kreg = number of parameters in $[\boldsymbol{\beta}', \boldsymbol{\gamma}', \sigma, \rho]'$
nreg = number of observations, total, not just selected
s = estimate of σ
rho = estimate of ρ

E54.4.2 An Incidental Truncation Model

Winkelmann (2008, pp. 153-154) describes a model (attributed to Crepon and Duguet (1997)) which is labeled the 'incidental truncation' model. This is a case in which the binary variable is correlated with the Poisson outcome, and directly affects it, in a form similar to the ZIP models discussed below. In this model, the data are observed when $z_i = 0$, but $z_i = 0$ implies that $y_i = 0$. The difference between this and the ZIP model is only the correlation between the two latent disturbances. The structure is actually a small modification of the model we have considered above.

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1],$$

$$z_i = \mathbf{1}(z_i^* > 0).$$

Thus, a probit model applies to the indicator, z_i. The following applies to the observed y_i:

$$y_i^* \sim \text{Poisson }(\lambda_i|\varepsilon_i) \text{ is a latent variable distributed as Poisson,}$$

$$\lambda_i|\varepsilon_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i),$$

$$y_i = y_i^* \text{ and } \mathbf{x}_i \text{ are observed when } z_i = 1,$$

$$y_i = 0 \text{ when } z_i = 0, \mathbf{x}_i \text{ is still observed when } z_i = 0.$$

For the sample selection model, the joint density of the observed response variables y_i and z_i is of the form

$$\mathbf{1}(z_i = 1) \times \{\text{Prob}(z_i = 1) \times \text{Poisson probability}\} + \mathbf{1}(z_i = 0)\times\text{Prob}(z_i = 0)$$

while for the incidental truncation model, the joint density is of the form

$$\text{Prob}(z_i = 1) \times \text{Poisson probability} + \mathbf{1}(z_i = 0)\times\text{Prob}(z_i = 0).$$

This model is requested by adding

; All

to the **POISSON** command given earlier.

PROBIT ; Lhs = ... ; Rhs = ... ; Hold $
POISSON ; Lhs = ... ; Rhs = ... ; Selection ; MLE ; All $

All other aspects are the same as in the model described earlier.

E54.4.3 Imposing Restrictions and Fixing ρ

The parameter vector is [β′, γ′, σ, ρ]′. Use this if you wish to impose constraints. For example, to fix the value of ρ at -.5 (as we do below in an example), you could use the following:

```
NAMELIST    ; xp = Rhs variables in probit equation
            ; xr = Rhs variables in Poisson model $
CALC        ; kp = Col(xp) ; kr = Col(xr) $
PROBIT      ; Lhs = ... ; Rhs = xp ; Hold $
POISSON     ; Lhs = the dependent variable
            ; Rhs = xr
            ; Selection ; MLE
            ; Rst = kr_b, kp_c, sg, -.5 $
```

You can use this device to test for a selectivity effect as well. The simple t and likelihood ratio tests can be carried out based on the value of ρ that is estimated. But, the t test requires estimation of the full model while the LR test requires assembling estimates of the pair of models and collecting three terms:

```
PROBIT      ; ... ; Hold $
POISSON     ; ... estimate full model by FIML $
CALC        ; lfiml = logl $
CALC        ; lprobit = logl  $
REJECT      ; the Lhs variable for probit model = 0 $
POISSON     ; ... Poisson model without selection, on selected observations $
CALC        ; lpois = logl
            ; List
            ; lm = 2*(lfiml - lprobit - lpois)
            ; 1 - Chi(lm,1)  $
```

The LM test should be the simplest to carry out. In the earlier example, just change our -.5 to 0, and add **; Maxit = 0** to the command. An example appears below.

E54.4.4 Application

The variable *cardhldr* is a binary variable which indicates whether the individual holds the major credit card whose vendor produced the overall data set; *inc_per*, which is the ratio of household income to number of dependents. The probit equation is used to model *cardhldr* as a function of *age*, *income*, and *inc_per*. The count variable analyzed here is *majordrg*, the number of major derogatory credit reports (long defaults) reported in the first year of credit card usage. The selection corrected Poisson model is then fit using only those observations for which *cardhldr* equals 1. Since *majordrg* is by far the dominant determinant of whether an application for a credit card will be accepted, one would expect the effect of the selection to be substantive in these data. The other variables used in the count model are *age*, *income*, *ownrent* and *avgexp*. The first command fits an unrestricted model. In the second, the correlation coefficient is fixed at -.5.

The commands are:

NAMELIST ; xp = one,age,income,incper $
NAMELIST ; xr = one,age,income,ownrent,avgyrexp $
PROBIT ; Lhs = cardhldr ; Rhs = xp ; Hold $
POISSON ; Lhs = majordrg ; Rhs = xr
; MLE ; Selection ; Partial Effects $
PROBIT ; Quietly ; Lhs = cardhldr ; Rhs = xp ; Hold $
POISSON ; Lhs = majordrg ; Rhs = xr
; MLE ; Selection ; Rst = 5_b,4_c,sc,-.5 $

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable             CARDHLDR
Log likelihood function     -6873.93812
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
CARDHLDR|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .23319***      .04775     4.88  .0000      .13961    .32678
     AGE|    -.00045         .00129     -.35  .7251     -.00297    .00207
  INCOME|     .13251***      .01067    12.42  .0000      .11159    .15343
  INCPER|     .08520***      .01085     7.85  .0000      .06393    .10647
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Unrestricted Poisson Regression Start Value
Dependent variable             MAJORDRG
Log likelihood function     -4875.30997
Estd sigma for heterogeneity =     .355
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
MAJORDRG|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -3.15941***      .09107   -34.69  .0000    -3.33790  -2.98092
     AGE|     .02109***      .00252     8.38  .0000      .01616    .02602
  INCOME|     .11903***      .01446     8.23  .0000      .09070    .14737
 OWNRENT|    -.03243         .05812     -.56  .5769     -.14635    .08149
AVGYREXP|     .24865***      .02746     9.05  .0000      .19483    .30248
--------+--------------------------------------------------------------------
Line search at iteration 45 does not improve fn. Exiting optimization.
-----------------------------------------------------------------------------
Poisson  Model with Sample Selection.
Dependent variable             MAJORDRG
Log likelihood function    -11212.97881
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit =     -6873.9381
LogL for initial Poisson=     -4875.3100
Means for Psn/Neg.Bin. use selected data.
Means for Probit based on all observations.
```

MAJORDRG	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters of Poisson/Neg. Binomial Probability					
Constant	-3.13246***	.17071	-18.35	.0000	-3.46705	-2.79787
AGE	.02594***	.00399	6.49	.0000	.01811	.03377
INCOME	-.10181**	.03956	-2.57	.0101	-.17934	-.02428
OWNRENT	.00137	.06968	.02	.9844	-.13519	.13793
AVGYREXP	.46542***	.07656	6.08	.0000	.31536	.61548
	Parameters of Probit Selection Model					
Constant	.23623***	.04585	5.15	.0000	.14638	.32609
AGE	-.86380D-05	.00127	-.01	.9946	-.25001D-02	.24828D-02
INCOME	.14002***	.00924	15.15	.0000	.12191	.15814
INCPER	.06679***	.00873	7.65	.0000	.04969	.08390
	Standard Deviation of Heterogeneity					
Sigma	2.46562***	.22069	11.17	.0000	2.03307	2.89817
	Correlation of Heterogeneity & Selection					
Rho	-.97406***	.01745	-55.82	.0000	-1.00826	-.93986

```
Partial derivatives of expected val. with
respect to the vector of characteristics.
They are computed at the means of the var-
iables.  Separate effects are shown first
followed by the sum of the two effects for
variables which appear in both Poisson and
Probit models.  Estimated value of E[y|D=1]
using sample mean =   .13436.
Note, std. errs. assume fixed rho & sigma.
```

MAJORDRG	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters of Poisson/Neg. Binomial Probability					
AGE	.00349***	.00054	6.49	.0000	.00243	.00454
INCOME	-.01368**	.00531	-2.57	.0101	-.02410	-.00326
OWNRENT	.00018	.00936	.02	.9844	-.01816	.01853
AVGYREXP	.06253***	.01029	6.08	.0000	.04237	.08269
	Parameters of Probit Selection Model					
AGE	-.22827D-05	.00034	-.01	.9946	-.66071D-03	.65614D-03
INCOME	.03700***	.01261	2.93	.0033	.01229	.06171
INCPER	.01765***	.00575	3.07	.0021	.00638	.02892
	Combined effect of two terms					
AGE	.00348***	.00048	7.20	.0000	.00254	.00443
INCOME	.02332**	.00948	2.46	.0139	.00474	.04190

```
Poisson  Model with Sample Selection.
Dependent variable              MAJORDRG
Log likelihood function    -11224.44055
Restricted log likelihood  -11749.24809
Chi squared [   2 d.f.]      1049.61508
Significance level               .00000
McFadden Pseudo R-squared      .0446673
Estimation based on N =  13444, K =  10
Inf.Cr.AIC  =  22468.9 AIC/N =    1.671
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit =    -6873.9381
LogL for initial Poisson=    -4875.3100
```

MAJORDRG	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters of Poisson/Neg. Binomial Probability					
Constant	-3.79785***	.14884	-25.52	.0000	-4.08958	-3.50612
AGE	.02418***	.00359	6.73	.0000	.01714	.03122
INCOME	.07714***	.02208	3.49	.0005	.03386	.12042
OWNRENT	-.01769	.06989	-.25	.8002	-.15467	.11929
AVGYREXP	.41203***	.05039	8.18	.0000	.31327	.51079
	Parameters of Probit Selection Model					
Constant	.22217***	.04621	4.81	.0000	.13160	.31274
AGE	-.00033	.00129	-.26	.7985	-.00285	.00219
INCOME	.13466***	.00978	13.77	.0000	.11549	.15384
INCPER	.08566***	.00960	8.92	.0000	.06684	.10448
	Standard Deviation of Heterogeneity					
Sigma	1.40050***	.06452	21.71	.0000	1.27404	1.52695
	Correlation of Heterogeneity & Selection					
Rho	-.50000	.....(Fixed Parameter).....				

Technical Details on FIML Estimation

The log likelihood function for the full model is the joint density for the observed data. When z_i equals one, $(y_i,\mathbf{x}_i,z_i,\mathbf{w}_i)$ are all observed. We seek $P[y_i, z_i=1|\mathbf{x}_i,\mathbf{w}_i]$. To obtain it, proceed as follows:

$$P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i] = \int_{-\infty}^{\infty} P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i]\, f(\varepsilon_i)d\varepsilon_i$$

$$= E_\varepsilon\{P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\boldsymbol{\varepsilon}_i]\}.$$

Conditioned on ε_i, z_i and y_i are independent. Therefore,

$$P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i] = P[y_i|\mathbf{x}_i,\varepsilon_i]\text{Prob}[z_i=1|\mathbf{w}_i,\varepsilon_i].$$

The first part, $P[y_i\,|\mathbf{x}_i,\varepsilon_i]$ is the conditional Poisson distribution with heterogeneity defined earlier. By joint normality, $f(u_i|\varepsilon_i) = N[(\rho/\sigma)\varepsilon_i\,, (1-\rho^2)]$. Therefore, $\text{Prob}[z_i=1|\mathbf{w}_i,\varepsilon_i]$ is

$$\text{Prob}[z_i=1|\mathbf{w}_i,\varepsilon_i] = \Phi\left([\boldsymbol{\alpha}'\mathbf{w}_i + (\rho/\sigma)\varepsilon_i]/\sqrt{1-\rho^2}\right).$$

Combining terms and using the earlier approach, the unconditional probability is

$$P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i] =$$

$$\int_{-\infty}^{\infty} \frac{\exp[-\lambda_i(\varepsilon)]\lambda_i(\varepsilon)^{y_i}}{y_i\,!}\,\Phi\left([\boldsymbol{\alpha}'\mathbf{w}_i + (\rho/\sigma)\varepsilon_i]/\sqrt{1-\rho^2}\right)\frac{1}{\sigma\sqrt{2\pi}}\exp[-\varepsilon_i^2/(2\sigma^2)]d\varepsilon_i\,.$$

Let $v = \varepsilon/(\sigma\sqrt{2})$, $\theta = \sigma\sqrt{2}$, $\tau = \sqrt{2}\,[\rho/\sqrt{1-\rho^2}\,]$, and $\boldsymbol{\gamma} = [1/\sqrt{1-\rho^2}\,]\boldsymbol{\alpha}$.

(Thus, the reverse transformations are

$$\rho^2 = [\tau^2/(2+\tau^2)],\ \text{Sgn}(\rho) = \text{Sgn}(\tau),\ \text{and}\ \sigma = \theta/\sqrt{2}\,.)$$

After making the change of variable and reparameterizing the probability as before, we obtain

$$P[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i] = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v^2)\ \frac{\exp[-\lambda_i(v)]\lambda_i(v)^{y_i}}{y_i\,!}\Phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_i)\ dv_i$$

where $\lambda_i(v) = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta v)$. This is approximated with Hermite quadrature since no closed form exists. When z_i equals zero, only $(z_i,\mathbf{w}_i)$ are observed. The contribution to the likelihood function is

$$\text{Prob}[z_i = 0|\mathbf{w}_i\ ] = E_\varepsilon[1 - \text{Prob}[u_i > -\boldsymbol{\alpha}'\mathbf{w}_i|\mathbf{w}_i,\varepsilon_i]] = E_\varepsilon[\text{Prob}[u_i \leq -\boldsymbol{\alpha}'\mathbf{w}_i|\mathbf{w}_i,\varepsilon_i]].$$

This provides the probability needed to construct the likelihood function.

$$\text{Prob}[z_i = 0|\mathbf{w}_i,\varepsilon_i] = 1 - \Phi[\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_i/(\sqrt{2}\,\sigma)]$$

so

$$\text{Prob}[z_i = 0|\mathbf{w}_i] = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v^2)\ \Phi[-(\boldsymbol{\gamma}'\mathbf{w} + \tau v)]dv.$$

Hermite quadrature is used to evaluate the integral.

Maximum likelihood estimates of $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \tau]$ are obtained by maximizing

$$\log L = \Sigma_{z=0} \log\text{Prob}[z_i=0|\mathbf{w}] + \Sigma_{z=1} \log P[y_i,z_i=1|\mathbf{x},\mathbf{w}].$$

The approximate function is

$$\log L = \sum_{i=obs.\ with\ z_i=1} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h\frac{\exp(-\lambda_i(v_h))\lambda_i(v_h)^{y_i}}{y_i\,!}\Phi\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h\right)\right]$$

$$+ \sum_{i=obs.\ with\ z_i=0} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h\Phi\left(-\boldsymbol{\gamma}'\mathbf{w}_i - \tau v_h\right)\right]$$

where v_h and ω_h are the nodes and weights for the quadrature and

$$\lambda_i(v_h) = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta v_h).$$

The BHHH estimator of the asymptotic covariance matrix for the parameter estimates is a natural choice given the complexity of the function. The first derivatives must be approximated as well. For convenience, let

$$P_{ih} = P(y_i,\lambda_i(v_h)) = \frac{\exp(-\lambda_i(v_h))\lambda_i(v_h)^{y_i}}{y_i!}$$

$$\lambda_{ih} = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta v_h)$$

$$\Phi_{ih} = \Phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h\ )\ \text{(normal CDF)}$$

and

$$\phi_{ih} = \phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h\ )\ \text{(normal density).}$$

To save some notation, denote the individual terms summed in the log likelihood as log L_i. We also take advantage of the result that $\partial P(.,.)/\partial z = P\times\partial \log P(.,.)/\partial z$ for any argument z which appears in the function. Then,

$$
\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\Phi_{ih}(y_i-\lambda_{ih})\,\mathbf{x}_i \\
\frac{\partial \log L}{\partial \theta} &= \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\Phi_{ih}(y_i-\lambda_{ih})\,v_h \\
\frac{\partial \log L}{\partial \boldsymbol{\gamma}} &= \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\phi_{ih}\,\mathbf{w}_i - \sum_{z_i=0} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \phi_{ih}\,\mathbf{w}_i \\
\frac{\partial \log L}{\partial \tau} &= \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\phi_{ih}\,v_h - \sum_{z_i=0} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \phi_{ih}\,v_h
\end{aligned}
$$

Estimates of the structural parameters, (α,ρ,σ) and their standard errors are computed using the delta method.

The incidental truncation model requires only minor modification of the preceding. The approximate log likelihood for that model is

$$
\begin{aligned}
\log L &= \sum_{all\ obs.} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \frac{\exp(-\lambda_i(v_h))\lambda_i(v_h)^{y_i}}{y_i!}\Phi\left(\boldsymbol{\gamma}'\mathbf{w}_i+\tau v_h\right)\right] \\
&\quad + \sum_{i=obs.\ with\ z_i=0} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \Phi\left(-\boldsymbol{\gamma}'\mathbf{w}_i-\tau v_h\right)\right].
\end{aligned}
$$

This merely changes the observations included in the first summation. Other changes in the subsequent results are likewise minor.

E54.5 Multinomial Logit Model

The multinomial logit model can be extended in the same fashion as the binomial logit model. As before, the first step is to incorporate the unobservable heterogeneity in the multinomial logit model in a consistent fashion, then extend the selection model. The basic probability model for choice among $J + 1$ alternatives is based on a random utility model,

$$U_{ij} = \boldsymbol{\beta}_j'\mathbf{x}_i + \varepsilon_{ij}$$

where ε_{ij}, $j = 0,...,J$ have independent type 1 extreme value distributions. This produces the familiar multinomial logit model

$$
\begin{aligned}
\text{Prob}(y_i = j|\mathbf{x}_i) &= \text{Prob}(U_{ij} > U_{ik})\ \forall\ k \neq j \\
&= \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_i)}{\Sigma_{m=0}^{J}\exp(\boldsymbol{\beta}_m'\mathbf{x}_i)},\ j = 0,1,...,J,\ \boldsymbol{\beta}_0 = \mathbf{0}.
\end{aligned}
$$

We introduce the individual heterogeneity into the model by augmenting the utility functions with the common individual term, v_i, so that

$$U_{ij} \mid v_i \;=\; \boldsymbol{\beta}_j'\mathbf{x}_i + \theta_j v_i + \varepsilon_{ij},\; v_i \sim \mathrm{N}[0,1].$$

Then, the conditional probabilities are

$$\begin{aligned}\mathrm{Prob}(y_i = j|\mathbf{x}_i,v_i) &= \mathrm{Prob}(U_{ij} > U_{ik}|v_i)\ \forall\ k \neq j \\ &= \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_i + \theta_j v_i)}{\Sigma_{m=0}^{J}\exp(\boldsymbol{\beta}_m'\mathbf{x}_i + \theta_m v_i)},\ j = 0,1,...,J,\ \boldsymbol{\beta}_0 = \mathbf{0}, \theta_0 = 0.\end{aligned}$$

As before, the selection mechanism is

$$\begin{aligned} z_i^* &= \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ u_i \sim N[0,1],\ z_i = 1(z_i^* > 0) \\ (y_i,\mathbf{x}_i) &\ \text{is observed only when } y_{i2} = 1 \\ (u_i,v_i) &\sim \mathrm{BVN}[(0,0),(1,\rho,1)] \end{aligned}$$

This model is estimated using maximum simulated likelihood.

An example appears below. Estimation proceeds in three steps. First, the starting values for the uncorrected multinomial logit model are obtained by simple linear regression of the choice binary variables, $A_{ij} = 1(y_i = j)$, $j = 1,...,J$ on $\mathbf{x}_i$. You can display these results by adding **; OLS** to your **MLOGIT** command, but we emphasize these OLS results are not useful for anything but computing starting values. Then, the multinomial logit model is computed ignoring the selection. (This step and the OLS results are based on the observations for which z_i equals one.) These results are not displayed. When these iterations are complete, the solver returns immediately to the iterations to compute the parameters of the full model. This intermediate step is used to improve the starting values. The final results are then displayed. You can also compute marginal effects, probabilities, etc. with the model in the same fashion as with the basic model without selectivity.

The commands for estimating this model are

```
PROBIT     ; Lhs = zi ; Rhs = variables in w ; Hold $
MLOGIT     ; Lhs = yi ; Rhs = variables in x  ; Selection $
```

All other parts of the command and optional features are the same as in the uncorrected case.

To illustrate this model, we have used the health care data employed in numerous earlier examples. Here, we have modeled the self reported health satisfaction variable (which is more naturally an ordered choice, but this is purely for a numerical example) as a multinomial logit outcome. The selection variable is whether or not the individual has visited the doctor. In order to simplify the application, we have reduced the sample size and truncated the distribution of outcomes by discarding observations with reported value greater than five. The commands and output are as follows:

```
REJECT     ; _groupti < 7 $
REJECT     ; hsat  > 5 $  (This leaves 1,939 observations in the sample.)
PROBIT     ; Lhs = doctor ; Rhs = one,age,married ; Hold $
MLOGIT     ; Lhs = hsat ; Rhs = one,hhninc,female
           ; Selection ; Partial Effects
           ; Pts = 25 ; Halton $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable               DOCTOR
Log likelihood function      -956.23551
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .09864         .17848      .55  .5805     -.25118    .44845
     AGE|     .01752***      .00366     4.78  .0000      .01034    .02470
 MARRIED|    -.06564         .08757     -.75  .4535     -.23728    .10600
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:   4 iterations. Status=0, F=    2185.501
Normal exit:  27 iterations. Status=0, F=    9379.743
-----------------------------------------------------------------------------
Sample Selection/Multinomial Logit
Dependent variable                 HSAT
Log likelihood function     -9379.74304
Sample observations selected:DOCTOR  =1
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Characteristics in numerator of Prob[HSAT=1  ]
Constant|    -.73557         .48874    -1.51  .1323    -1.69348    .22234
  HHNINC|     .26479        1.33280      .20  .8425    -2.34744   2.87702
  FEMALE|     .10919         .36555      .30  .7652     -.60727    .82565
        | Characteristics in numerator of Prob[HSAT=2  ]
Constant|    -.31745         .34419     -.92  .3564     -.99205    .35715
  HHNINC|    2.04040**       .87440     2.33  .0196      .32660   3.75421
  FEMALE|     .20314         .28148      .72  .4705     -.34855    .75483
        | Characteristics in numerator of Prob[HSAT=3  ]
Constant|     .53281*        .31114     1.71  .0868     -.07702   1.14264
  HHNINC|    1.27974         .83194     1.54  .1240     -.35083   2.91030
  FEMALE|     .07637         .25708      .30  .7664     -.42750    .58024
        | Characteristics in numerator of Prob[HSAT=4  ]
Constant|     .49256         .30491     1.62  .1062     -.10505   1.09017
  HHNINC|    1.94458**       .80461     2.42  .0157      .36758   3.52159
  FEMALE|     .06189         .25180      .25  .8058     -.43162    .55541
        | Characteristics in numerator of Prob[HSAT=5  ]
Constant|    1.73810***      .26891     6.46  .0000     1.21104   2.26516
  HHNINC|    1.12489         .73288     1.53  .1248     -.31153   2.56131
  FEMALE|     .24064         .23129     1.04  .2981     -.21268    .69396
        | Utility weights on latent heterogeneity
Theta_00|        0.0    .....(Fixed Parameter).....
Theta_01|     .30356         .19640     1.55  .1222     -.08137    .68848
Theta_02|    -.11145         .15471     -.72  .4713     -.41468    .19179
Theta_03|    -.04753         .14733     -.32  .7470     -.33629    .24124
Theta_04|     .02616         .14186      .18  .8537     -.25189    .30421
Theta_05|    -.00108         .13334     -.01  .9935     -.26242    .26025
        | Reestimated Probit Selection Equation
Constant|     .09833         .18919      .52  .6032     -.27247    .46913
     AGE|     .01753***      .00395     4.44  .0000      .00979    .02527
 MARRIED|    -.06568         .08765     -.75  .4536     -.23747    .10611
        | Correlation Between Heterogeneity and Selection
Rho(e,u)|     .00267         .03243      .08  .9344     -.06090    .06624
```

```
--------------------------------------------------------------------------------
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are  All Obs.
A full set is given for the entire set of
outcomes, HSAT     =  0 to HSAT     =   5
Probabilities at the mean values of X are
  0= .054 1= .030 2= .086 3= .147 4= .174
  5= .508
--------+-----------------------------------------------------------------------
        |     Partial                          Prob.      95% Confidence
    HSAT|      Effect    Elasticity      z     |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        | Marginal effects on Prob[HSAT=0  ]
  HHNINC|    -.06933*       -.42594    -1.92  .0547     -.14006    .00141
  FEMALE|    -.00893        -.08945     -.77  .4397     -.03156    .01371
        | Marginal effects on Prob[HSAT=1  ]
  HHNINC|    -.03054        -.33804     -.92  .3566     -.09548    .03439
  FEMALE|    -.00168        -.03032     -.19  .8456     -.01858    .01522
        | Marginal effects on Prob[HSAT=2  ]
  HHNINC|     .06546         .25140     1.46  .1455     -.02267    .15359
  FEMALE|     .00328         .02055      .23  .8210     -.02513    .03169
        | Marginal effects on Prob[HSAT=3  ]
  HHNINC|    -.00049        -.00111     -.01  .9939     -.12596    .12498
  FEMALE|    -.01303        -.04810     -.72  .4708     -.04845    .02239
        | Marginal effects on Prob[HSAT=4  ]
  HHNINC|     .11534*        .21959     1.72  .0861     -.01639    .24707
  FEMALE|    -.01801        -.05594     -.93  .3535     -.05606    .02004
        | Marginal effects on Prob[HSAT=5  ]
  HHNINC|    -.08044        -.05252     -.89  .3733     -.25753    .09665
  FEMALE|     .03837         .04086     1.50  .1343     -.01186    .08859
--------+-----------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

Marginal Effects Averaged Over Individuals
--------+---------+---------+---------+---------+---------+---------+
Variable|  HSAT=0 |  HSAT=1 |  HSAT=2 |  HSAT=3 |  HSAT=4 |  HSAT=5 |
--------+---------+---------+---------+---------+---------+---------+
ONE     |  -.0549 |  -.0524 |  -.1134 |  -.0676 |  -.0874 |   .3757 |
HHNINC  |  -.0704 |  -.0307 |   .0652 |  -.0004 |   .1153 |  -.0792 |
FEMALE  |  -.0091 |  -.0017 |   .0033 |  -.0129 |  -.0180 |   .0383 |
--------+---------+---------+---------+---------+---------+---------+

Averages of Individual Elasticities of Probabilities
--------+---------+---------+---------+---------+---------+---------+
Variable|  HSAT=0 |  HSAT=1 |  HSAT=2 |  HSAT=3 |  HSAT=4 |  HSAT=5 |
--------+---------+---------+---------+---------+---------+---------+
ONE     |  -.9949 | -1.7305 | -1.3124 |  -.4621 |  -.5024 |   .7432 |
HHNINC  |  -.4292 |  -.3416 |   .2457 |  -.0059 |   .2140 |  -.0571 |
FEMALE  |  -.0877 |  -.0308 |   .0180 |  -.0479 |  -.0555 |   .0376 |
--------+---------+---------+---------+---------+---------+---------+
```

E54.6 Sample Selected Stochastic Frontier Model

This model does not yet appear in the literature, and is new with this release of *LIMDEP*. The model is a familiar sample selection form

$$z^* = \boldsymbol{\alpha}'\mathbf{w} + \omega, \qquad z = 1(z^* > 0)$$

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u$$

$$u = |U| \text{ with } U \sim N[0,\sigma_u^2]$$

$$(v,\omega) \sim \text{Bivariate normal with } [(0,0),(\sigma_v^2, \rho\sigma_v, 1)]$$

and $(y,\mathbf{x})$ only observed when $z = 1$.

(It is necessary to deviate from the common notation of this chapter because the frontier function literature also has a common notation for the components of these models, that conflicts with our usage in this chapter. The difference will be immaterial.) The selection mechanism operates through the heterogeneity component of the production model, v, not the inefficiency, u. (Thus, 'observation' – being in the sample – is not viewed as a function of the level of inefficiency.)

The model is fit by maximum simulated likelihood. To request it, use

```
PROBIT     ; Lhs = d ; Rhs = variables in w ; Hold $
FRONTIER   ; Lhs = y ; Rhs = variables in x ; Selection $
```

The model must be the base case, half normal model, with no panel data application, no truncation, or heteroscedasticity, etc. Other aspects of the frontier model, in particular,

```
; Eff = JLMS estimates of u
```

operate in the usual way.

You may control the simulations with **; Halton** and **; Pts** for the simulation. The estimation method is developed in detail in Section E54.5 below.

In the example below, we use a contrived selection mechanism with the dairy farm data used to demonstrate the stochastic frontier models in Chapters E62-E64. The variables in the model are output, milk production, and four inputs, cows, land, feed and labor, in log form. We created *zi* as simply a dummy variable that splits the sample into large and small farms and used a logit model based simply on the number of cows. The second results do not correct for selection. The commands are

```
CREATE     ; group = yit > 11.1$
PROBIT     ; Lhs = group ; Rhs = one,cows ; Hold ; Quiet $
FRONTIER   ; Lhs = yit ; Rhs = one,x1,x2,x3,x4
           ; Selection
           ; Halton ; Pts = 15 ; tlg = 1.d-4 $
FRONTIER   ; Lhs = yit ; Rhs = one,x1,x2,x3,x4 $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                   YIT
Log likelihood function        607.89950
Estimation based on N =   1482, K =   8
Variances: Sigma-squared(v)=      .01713
           Sigma-squared(u)=      .00243
           Sigma(u)        =      .04929
           Sigma(v)        =      .13090
           Sigma           =      .13987
           Lambda          =      .37658
Sample Selection/Frontier Model
Murphy/Topel Corrected VC Matrix
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           491.68632
Chi-sq=2*[LogL(SF)-LogL(LS)] =  232.426
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.6143***      .01634   710.69  .0000     11.5823   11.6463
      X1|     .65235***      .02271    28.72  .0000      .60783    .69687
      X2|     .02442**       .01232     1.98  .0475      .00027    .04856
      X3|     .03930***      .01326     2.96  .0030      .01332    .06528
      X4|     .41258***      .01078    38.26  .0000      .39145    .43371
Sigma(u)|     .04929**       .01973     2.50  .0125      .01062    .08797
Sigma(v)|     .13090***      .00360    36.37  .0000      .12385    .13795
Rho(w,v)|     .80390***      .06290    12.78  .0000      .68061    .92718
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                   YIT
Log likelihood function        822.68831
Estimation based on N =   1482, K =   7
Variances: Sigma-squared(v)=      .01075
           Sigma-squared(u)=      .02425
           Sigma(v)        =      .10371
           Sigma(u)        =      .15573
Sigma = Sqr[(s^2(u)+s^2(v)]=      .18710
Gamma = sigma(u)^2/sigma^2 =      .69277
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           809.67610
Chi-sq=2*[LogL(SF)-LogL(LS)] =   26.024
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.7014***       .00447  2614.87  .0000     11.6926   11.7101
      X1|     .58369***       .01887    30.93  .0000      .54670    .62068
      X2|     .03555***       .01113     3.20  .0014      .01375    .05736
      X3|     .02256*         .01281     1.76  .0783     -.00256    .04768
      X4|     .44948***       .01035    43.42  .0000      .42919    .46977
        |Variance parameters for compound error
  Lambda|    1.50164***       .08748    17.17  .0000     1.33019   1.67310
   Sigma|     .18710***       .00011  1698.90  .0000      .18688    .18732
--------+--------------------------------------------------------------------
```

E54.7 Tobit Model with Selectivity

The sample selection model detailed in Chapter E52 is extended to the tobit model. The model is

$$y^* \;=\; \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

$$y \;=\; 0 \text{ if } y^* \leq 0,\ y = y^* \text{ otherwise, or } y = \max(0,y^*), \text{(tobit)}$$

$$z^* \;=\; \boldsymbol{\alpha}'\mathbf{w} + u,$$

$$z \;=\; 1 \text{ if } z^* > 0, \text{ or } 0 \text{ if } z^* \leq 0, \text{ or } z = 1[z^* > 0], \qquad \text{(probit)}$$

$$[y\ ,\mathbf{x}\ ] \quad \text{are observed only when } z = 1, \qquad \text{(sampling)}$$

This model is a mixture of censoring and a type of truncation. The procedure for estimating this model follows the standard set of steps for selectivity models given in Section E52.2.2. The standard procedure for estimating a sample selectivity model in *LIMDEP* is:

Step 1. Estimate the parameters of the probit model first and **; Hold** them aside for the next step in the procedure.

Step 2. Using the probit results from Step 1, fit the main equation of the model.

The tobit estimator to be described here is a full information maximum likelihood estimator. Nonetheless, at the beginning of Step 2, a second step least squares regression is computed in order to obtain the starting values for the MLE. These are corrected for selection, to a degree, *but they are still inconsistent*. The results given at this point are obtained by least squares, and, as such, are inconsistent in the same manner that the OLS coefficients are inconsistent in the basic tobit model. As noted, these are just starting values for the iterations. The MLE is consistent and efficient.

The commands are

PROBIT ; Lhs = z ; Rhs = list for w ; Hold $
SELECT ; Tobit ; MLE ; Lhs = y ; Rhs = list for x $

Note that the command for the tobit model in this case is **SELECT**, not **TOBIT**.

NOTE: As in the MLE for the selection model, there is no 'lambda' variable computed for this model. The estimator is not least squares. When a sample selection model is fit by maximum likelihood, there is no selection 'correction' variable added to the model.

The model parameters estimated by MLE are $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, σ, ρ. The probit coefficients precede the regression parameters in the parameter vector. You may provide your own starting values for the iterations with

; Start = ... list

Fixed value and equality restrictions may be imposed with

; Rst = ... list

The first set of output from the **SELECT** command is the standard output from the two step least squares estimation of this model. The final output includes the log likelihood and an indication of the parts of the parameter vector. The parameter vector shown is $[\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_1,\rho]$. Remaining output is the same as for the selection model. The retrievable results from this estimator are as follows:

Matrices: *b* and *varb* as usual. These contain $[\boldsymbol{\alpha},\boldsymbol{\beta},\sigma,\rho]$. Do not use **; Par**.
*bsr*1 = all of *b* except α

Scalars: *logl, nreg, rho, varrho, s, ybar, sy, sigma*1

Last Model: *a_variables*, *b_variables*, *r*12, *sigma*

E54.7.1 Predictions from the Selection Model

The tobit model with sample selection uses the linear prediction of the underlying latent variable for the fitted values. This is

$$E[y^* \mid z = 1] = \boldsymbol{\beta}'\mathbf{x} + \rho\sigma\lambda$$

where $$\lambda = \phi(\boldsymbol{\alpha}'\mathbf{w}) / \Phi(\boldsymbol{\alpha}'\mathbf{w}).$$

This is the value that is displayed and kept with **; List** and **; Keep**. Other parts of the fitted values listing are the same as for the basic tobit model. These predictions are based on the linear, single equation specification, not the tobit specification, and they do not exploit the correlation between the primary equation and the selection model. As such, they can be improved with some further manipulation. For the observed variable in the tobit model, ignoring the selectivity,

$$\begin{aligned} E[y|\mathbf{x}] &= \text{Prob}(y > 0|\ \mathbf{x}) \times E[y|y > 0, \mathbf{x}] \\ &= \Phi[\boldsymbol{\beta}'\mathbf{x}/\sigma] \times [\boldsymbol{\beta}'\mathbf{x} + \sigma\lambda], \end{aligned}$$

where $$\lambda = \phi[\boldsymbol{\beta}'\mathbf{x}/\sigma] / \Phi[\boldsymbol{\beta}'\mathbf{x}/\sigma].$$

For the tobit model with selection, we need, instead,

$$E[y|\ \mathbf{x}, \text{selection}] = \text{Prob}(y > 0 \mid z = 1) \times E[y \mid \mathbf{x}, y > 0, z = 1].$$

The probability can be found from the bivariate normal distribution:

$$\text{Prob}(y > 0|z = 1) = \Phi_2[\boldsymbol{\beta}'\mathbf{x}/\sigma,\boldsymbol{\alpha}'\mathbf{w},\rho] / \Phi(\boldsymbol{\alpha}'\mathbf{w}).$$

The conditional mean function is more involved. We use a general result for truncation in a bivariate normal distribution. For present purposes, it would be as follows:

$$E[y|y > 0, z = 1] = \boldsymbol{\beta}'\mathbf{x} + E[\varepsilon|\varepsilon > -\boldsymbol{\beta}|\mathbf{x}, u > -\boldsymbol{\alpha}'\mathbf{w}].$$

To simplify the notation, write this as

$$E[\varepsilon|\varepsilon > -\boldsymbol{\beta}'\mathbf{x}, u > -\boldsymbol{\alpha}'\mathbf{w}] = \sigma E[q|q > h, u > k],$$

where $q = \varepsilon/\sigma$,

$h = -\boldsymbol{\beta}'\mathbf{x}/\sigma$,

and $k = -\boldsymbol{\alpha}'\mathbf{w}$.

Maddala (1983) gives an expression for this conditional mean of a bivariate standard normal distribution (0,0,1,1,ρ). Let Φ_2 denote the bivariate normal probability and

$$\delta = -1 / (1 - \rho^2)^{1/2}.$$

Then, $E[q \mid q > h, u > k] = \{\phi(h)\Phi[\delta(k - \rho h)] + \rho\phi(k)\Phi[\delta(h - \rho k)]\}/\Phi_2.$

Thus, $E[y \mid z = 1] = \Phi_2\boldsymbol{\beta}'\mathbf{x} + \sigma\{\phi(h)\Phi[\delta(k - \rho h)] + \rho\phi(k)\Phi[\delta(h - \rho k)]\}.$

The program below can be used for this computation: We first set up the data and fit the model.

```
NAMELIST    ; x = variables in tobit model ; w = variables in probit $
CREATE      ; y = dependent variable in regression
            ; z = dependent variable in probit equation $
PROBIT      ; Lhs = z ; Rhs = w ; Hold $
SELECT      ; Lhs = y ; Rhs = x  ; Tobit ; MLE $
```

Do the following only if data for the predictions are unavailable for the limit (*d*=0) observations.

```
REJECT      ; z = 0 $
```

Determine the size and location of parameter vectors in *b.*

```
CALC        ; ka = Col(w) ; kb = ka + Col(x) ; jb = ka+1 $
```

Extract subvectors of saved parameter vector. Scalars *s* and *rho* already contain σ and ρ needed below.

```
MATRIX      ; alpha = b(1 : ka) ; beta = b(jb : kb) $
```

This simplifies the bivariate normal calculation. Then, set up the variables for the bivariate normal.

```
CALC        ; delta = -1 / Sqr (1 - rho^2) $
CREATE      ; h = -beta'x/s ; mh = -h ; k = -alpha'w ; mk = -k $
NAMELIST    ; hk = mh, mk $
```

Compute the conditional mean function.

```
CREATE      ; phi2 = Bvn(hk,rho)
            ; ey = phi2 * beta'x
                 + s  * (N01(h) * Phi(delta*(k-rho*h))
                 + rho*N01(k) * Phi(delta*(h-rho*k))) $
```

The result can now be inspected or saved in a file.

```
LIST        ; ey $
```

E54.7.2 Application

To illustrate this model, we have fit an hours equation using the Mroz labor supply data analyzed in Section E52.2.2. Here, we use additional information about the determinants of labor force participation.

```
PROBIT      ; Lhs = lfp ; Rhs = one,kids,faminc,cit
            ; Hold $
SELECT      ; Lhs = whrs ; Rhs = one,kl6,k618,wa,we
            ; Tobit ; MLE $
```

This command sequence produces the following results:

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    LFP
Log likelihood function        -510.23024
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
     LFP|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .04610        .12714      .36  .7169     -.20308    .29529
    KIDS|    -.10867        .10071    -1.08  .2806     -.30607    .08872
  FAMINC| .11344D-04***   .3945D-05    2.88  .0040  .36109D-05  .19077D-04
     CIT|    -.09083        .09943     -.91  .3610     -.28571    .10406
--------+--------------------------------------------------------------------
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on LFP                     |
| Selection rule is: Observations with LFP      =  1         |
| Results of selection:                                      |
|                  Data points     Sum of weights            |
| Data set             753              753.0                |
| Selected sample      428              428.0                |
+------------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step      least squares regression ............
LHS=WHRS      Mean                    =     1302.92991
              Standard deviation      =      776.27438
              Number of observs.      =            428
Model size    Parameters              =              6
              Degrees of freedom      =            422
Residuals     Sum of squares          =    .228070E+09
              Standard error of e     =      735.15314
Fit           R-squared               =         .10104
              Adjusted R-squared      =         .09039
Model test    F[  5,    422] (prob) =      9.5(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Standard error corrected for selection  1831.13623
Correlation of disturbance in regression
and Selection Criterion (Rho)         =       -1.00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    4044.67***    1038.178     3.90  .0001     2009.88   6079.46
     KL6|   -279.240*      154.1284    -1.81  .0700    -581.326    22.846
    K618|   -95.5417*      55.43541    -1.72  .0848   -204.1931   13.1097
      WA|   -12.4152        8.78275    -1.41  .1575    -29.6291    4.7986
      WE|   -44.3066       27.41159    -1.62  .1060    -98.0323    9.4191
  LAMBDA|   -2178.31**     1108.233    -1.97  .0493    -4350.41     -6.21
--------+--------------------------------------------------------------------
Normal exit:  49 iterations. Status=0, F=    3949.050
-----------------------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                   WHRS
Log likelihood function     -3949.04956
Estimation based on N =    753, K =  11
Inf.Cr.AIC  =   7920.1 AIC/N =   10.518
LHS is CENSORED. Tobit Model fit by MLE.
FIRST  4 estimates are probit equation.
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    WHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selection (probit) equation for LFP
Constant|    -.00712         .12085     -.06  .9530     -.24398    .22974
    KIDS|    -.11741         .10046    -1.17  .2425     -.31431    .07949
  FAMINC| .14538D-04***    .3395D-05    4.28  .0000  .78842D-05  .21191D-04
     CIT|    -.11780         .09641    -1.22  .2218     -.30676    .07116
        |Corrected regression, Regime 1
Constant|    2496.30***    484.8102     5.15  .0000     1546.09   3446.51
     KL6|   -321.906***    80.39344    -4.00  .0001    -479.474  -164.338
    K618|   -114.815***    34.01524    -3.38  .0007    -181.483   -48.146
      WA|   -9.31341*       5.44718    -1.71  .0873   -19.98968   1.36286
      WE|   -27.3182       18.02920    -1.52  .1297    -62.6548    8.0183
SIGMA(1)|    803.614***    112.7699     7.13  .0000     582.589  1024.638
RHO(1,2)|    -.46809         .34061    -1.37  .1694    -1.13568    .19950
--------+--------------------------------------------------------------------
```

E54.7.3 Technical Details on Estimation

The log likelihood function for the tobit model with sample selection is as follows:

$$\begin{aligned}\log L = & \Sigma_{z=0} \log\Phi(-\boldsymbol{\alpha}'\mathbf{w}) \\ & + \Sigma_{z=1,y=0} \log\Phi_2[-\boldsymbol{\beta}'\mathbf{x}/\sigma,\boldsymbol{\alpha}'\mathbf{w},-\rho] \\ & + \Sigma_{z=1,\,y>0} -\tfrac{1}{2}[\log 2\pi + \log\sigma + (\varepsilon_i/\sigma)^2] + \log\Phi[r_i/(1-\rho^2)^{1/2}],\end{aligned}$$

where $\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}$,

$r_i = \boldsymbol{\alpha}'\mathbf{w} + \rho\varepsilon_i/\sigma$,

Let $\delta = 1/(1-\rho^2)^{1/2}$.

Derivatives in the three parts of the log likelihood are defined as:

$$\begin{aligned} g_\rho &= \partial\log L_i/\rho, \\ g_\sigma &= \partial\log L_i/\sigma, \\ \mathbf{d}_\alpha &= \partial\log L_i/\partial(\boldsymbol{\alpha}'\mathbf{w}), \\ \mathbf{d}_\beta &= \partial\log L_i/\partial(\boldsymbol{\beta}'\mathbf{x}). \end{aligned}$$

For the three parts of the log likelihood function, in the order above:

$$\begin{aligned} \mathrm{g}_\rho &= 0, \\ \mathrm{g}_\sigma &= 0, \\ \mathbf{d}_\alpha &= -\phi/\Phi \text{ from the first normal CDF term}, \\ \mathbf{d}_\beta &= 0. \end{aligned}$$

The second set of terms are from the bivariate probit model presented in Chapter E33.

$$\begin{aligned} g_\rho &= \phi_{00}/\Phi_{00} \text{ (bivariate normal density over CDF)}, \\ g_\rho &= (\phi/\Phi)[\delta\varepsilon_i/\sigma_1 + \rho u_i\delta^3], \\ g_\sigma &= -1/\sigma_1 + (\varepsilon_i/s_1)^2/\sigma_1 - (\phi/\Phi)\rho\delta\varepsilon_i/\sigma_1^2, \\ \mathbf{d}_\alpha &= (\phi/\Phi)\delta, \\ \mathbf{d}_\beta &= \varepsilon_i/\sigma_1^2 - (\phi/\Phi)\rho\delta/\sigma_1. \end{aligned}$$

Terms are then assembled for the gradient. The BHHH estimator is used for the asymptotic covariance matrix.

E54.8 Binomial Logit Model

The binomial probit model with sample selection converts naturally into a form of bivariate probit model. The simple result follows from the bivariate normality of the heterogeneity in the main equation and the unobservables in the selection equation. Likewise, the tobit model with the sample selection correction in Section E54.7 becomes convenient through the bivariate normality of the two random components. If the main equation is a logit model, rather than a probit or some other structure based on the normal distribution, then an alternative method is needed. We use the following: The departure point is a logit model of interest,

$$\text{Prob}(y_i = 1|\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)},\ \text{Prob}(y_i = 0|\mathbf{x}_i) = 1 - \text{Prob}(y_i = 1|\mathbf{x}_i)$$

and the familiar selection mechanism,

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ u_i \sim N[0,1],\ z_i = 1(z_i^* > 0)$$

$$(y_i,\mathbf{x}_i) \text{ is observed only when } y_{i2} = 1.$$

There is no obvious connection between the logit model of interest and the selection mechanism. Since the force of sample selection is exerted through its influence on unobservables in the model, we modify the model as follows:

$$\text{Prob}(y_i = 1|\varepsilon_i, \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)},\ \varepsilon_i \sim N[0,1].$$

$$(u_i,\varepsilon_i) \sim \text{BVN}[(0,0),(1,\rho,1)]$$

The commands are

```
PROBIT    ; Lhs = zi ; Rhs = variables in w ; Hold $
LOGIT     ; Lhs = yi ; Rhs = variables in x  ; Selection $
```

Other features, such as marginal effects, predicted values, etc. are the same as with the binary logit model without selection. This model is estimated using maximum simulated likelihood. Full details are presented in Section E54.5. Note, marginal effects are computed using the final estimates of β, but without the indirect effects of the selection equation. Thus, as in the uncorrected case, the marginal effects are computed after estimation as

$$\boldsymbol{\delta} = \Lambda(\boldsymbol{\beta}'\mathbf{x})[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x})]\,\boldsymbol{\beta}.$$

To illustrate, we have reestimated the bivariate probit model of the preceding section. The commands are

```
CREATE     ; avgyrexp = avgyrexp/10000 $
CREATE     ; income = income/10000 $
CREATE     ; incper = incper/10000 $
NAMELIST   ; card = one,age,income,ownrent,selfempl,curntadd $
NAMELIST   ; dflt  = one,income,avgyrexp,depdnt,incper,credmajr,tradacct $
PROBIT     ; Lhs = cardhldr ; Rhs = card ; Hold $
LOGIT      ; Lhs = default ; Rhs = dflt ; Selection ; Halton ; Pts = 25 $
```

The second set of estimates are the uncorrected logit estimates from the restricted sample of cardholders. As might be expected given the small estimate of ρ, the results are very similar.

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable              CARDHLDR
Log likelihood function      -6866.36894
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
CARDHLDR|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .37122***      .04615     8.04  .0000       .28076    .46167
     AGE|    -.00323**       .00141    -2.29  .0218      -.00598   -.00047
  INCOME|     .16061***      .01019    15.76  .0000       .14064    .18059
 OWNRENT|     .16334***      .02753     5.93  .0000       .10938    .21731
SELFEMPL|    -.34597***      .05113    -6.77  .0000      -.44619   -.24575
CURNTADD| .33354D-04         .00021      .16  .8765 -.38739D-03  .45410D-03
--------+--------------------------------------------------------------------
Logit    Regression Start Values for DEFAUL
Dependent variable               DEFAULT
Log likelihood function      -3182.37907
Estimation based on N =  13444, K =   7
Inf.Cr.AIC  =    6378.8 AIC/N =     .474
Model estimated: Aug 09, 2011, 23:08:27
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 DEFAULT|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
Constant|    -.90918***      .12363    -7.35  .0000    -1.15150   -.66686
  INCOME|    -.33584***      .04154    -8.09  .0000     -.41724   -.25443
AVGYREXP|    -.20111*        .11529    -1.74  .0811     -.42707    .02486
  DEPDNT|     .09223**       .04173     2.21  .0271      .01045    .17402
  INCPER|    -.03839         .04690     -.82  .4131     -.13031    .05354
CREDMAJR|    -.17314**       .08547    -2.03  .0428     -.34065   -.00562
TRADACCT|    -.03129***      .00642    -4.88  .0000     -.04387   -.01872
--------+--------------------------------------------------------------------

Line search at iteration 18 does not improve fn. Exiting optimization.
-----------------------------------------------------------------------------
Selectivity Corrected Logit Model
Dependent variable               DEFAULT
Log likelihood function     -53323.21934
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 DEFAULT|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Coefficients in binary logit model
Constant|    -.90905***      .11501    -7.90  .0000    -1.13446   -.68364
  INCOME|    -.33587***      .04143    -8.11  .0000     -.41707   -.25467
AVGYREXP|    -.20075*        .10551    -1.90  .0571     -.40754    .00604
  DEPDNT|     .09232**       .04486     2.06  .0396      .00440    .18024
  INCPER|    -.03836         .04758     -.81  .4201     -.13161    .05489
CREDMAJR|    -.17357**       .08574    -2.02  .0429     -.34161   -.00553
TRADACCT|    -.03129***      .00640    -4.89  .0000     -.04384   -.01875
```

```
--------+-------------------------------------------------------------------
        |Coefficients in Probit = Alpha/sqr(1-rho^2)
Constant|     .37122***      .04582     8.10  .0000      .28141     .46103
     AGE|    -.00323**       .00140    -2.30  .0214     -.00598    -.00048
  INCOME|     .16061***      .00889    18.07  .0000      .14320     .17803
 OWNRENT|     .16334***      .02716     6.01  .0000      .11010     .21659
SELFEMPL|    -.34597***      .05042    -6.86  .0000     -.44479    -.24715
CURNTADD| .33363D-04         .00022      .15  .8780 -.39269D-03  .45941D-03
        |Standard deviation of e(i) in logit model
Sigma(e)|     .01473         .03354      .44  .6607     -.05102     .08047
        |Correlation of selection u(i) and logit e(i)
Rho(u,e)|    -.00012         .01215     -.01  .9919     -.02394     .02369
--------+-------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

E54.9 Grouped Data Model with Selection

The grouped data model is also extended to the sample selection treatment. (This model is developed in Bhat (1994).) (The model is described earlier in Section E47.3.) The structure is as follows:

$$
\begin{aligned}
y^* &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\ \varepsilon \sim N[0,\sigma^2],\\
y &= j \text{ if } A_{j-1} \leq y^* < A_j,\ j = 1,\ldots,J,\ A_0 = -\infty,\ A_J = +\infty,\\
d^* &= \boldsymbol{\alpha}'\mathbf{z} + u\\
d &= \text{if } d^* > 0 \text{ and } 0 \text{ otherwise,}\\
[\varepsilon,u] &\sim N[0,0,\sigma^2,1,\rho],\\
[y,\mathbf{x}] &\ \text{are observed only when } d = 1.
\end{aligned}
$$

The correlation between ε and u is ρ. The selection aspect of the model arises when ρ is not equal to zero. Note that this extension is the same as its counterpart discussed above for the tobit model. The command is

GROUPED DATA ; Lhs = y,d
; Rh1 = variables in x
; Rh2 = variables in z
; Limits = a1, a2,...,aJ-1 $

The **GROUPED DATA** command is exactly the same as in the nonselected case. As before, you give only the interior limit points. The difference is the specification of the probit equation by the second Lhs variable and the Rh2 list. (Since this model proceeds directly to the MLE, we do not begin with a separate **PROBIT** command, as we do with most other sample selection models.)

The usual options are available, including fitted values, residuals, optimization controls, etc., with two exceptions. First, the **; Partial Effects** option is not supported for this model. Second, the default algorithm is BFGS, and this cannot be changed. In addition, you may impose within equations restrictions with the **; Rst = list** option.

The retrievable results from this model are

Matrices: *b*, *varb*; use **; Par** to add (σ,ρ) to the parameter vector

Scalars: *s*, *rho*, *logl*, *kreg*, *nreg*, *ybar*, *sy*, *exitcode*

Last Model: *b_variables* = elements of $\boldsymbol{\beta}$
a_variables = elements of $\boldsymbol{\alpha}$, *sigma*, *r*12

Technical Details on the Grouped Data Regression Models

Optimization is the same as for **TOBIT**. All options, including **; Maxit**, **; Tlf**, **; Start**, **; Rst**, etc. operate the same. Olsen's transformation is used during the iterations. The log likelihood function for the grouped data model is

$$\log L = \Sigma_i\{\log[\Phi(\eta U - \boldsymbol{\gamma}'\mathbf{x}_i) - \Phi(\eta L - \boldsymbol{\gamma}'\mathbf{x}_i)]\}$$

where $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ and $\eta = 1/\sigma$.

For this case, U is the upper limit of the range in which y_i falls, and L is the lower limit. Gradients and Hessians for these can be derived using the results shown earlier for the tobit model, as the terms are identical. The second derivatives are used in estimating the asymptotic covariance matrix for the estimates.

$$\partial \text{logL}/\partial(\boldsymbol{\gamma},\eta) = \sum\nolimits_{i=1}^{n} \frac{1}{\Phi_U - \Phi_L}\left[\phi_U\begin{pmatrix}-\mathbf{x}_i\\ U\end{pmatrix} - \phi_L\begin{pmatrix}-\mathbf{x}_i\\ L\end{pmatrix}\right].$$

Let $\lambda_m = \phi_{\text{m}} / [\Phi_{\text{U}} - \Phi_{\text{L}}],\ m = L,U.$

and $w_m = [-\mathbf{x}, m]',\ m = L,U$

Then, $$\frac{\partial^2 \log L}{\partial(\boldsymbol{\gamma},\eta)\partial(\boldsymbol{\gamma},\eta)'} = \sum\nolimits_{i=1}^{n} \{\lambda_U\mathbf{w}_U\,[(-\alpha_U - \lambda_U)\mathbf{w}_U' + \lambda_L\mathbf{w}_L']\} - \{\lambda_L\mathbf{w}_L\,[(-\alpha_L + \lambda_L)\mathbf{w}_L' - \lambda_U\mathbf{w}_U']\}.$$

For the sample selection version, estimates of $[\boldsymbol{\beta},\boldsymbol{\alpha},\sigma,\rho]$ are obtained by full information maximum likelihood. The log likelihood is constructed from the simple probabilities of the events:

$$\log L = \Sigma_{d=0} \log[1 - \Phi(\boldsymbol{\gamma}'\mathbf{z})]$$
$$+\Sigma_{d=1,y=j} \log[\Phi_2(\eta A_j - \boldsymbol{\beta}'\mathbf{x},\boldsymbol{\alpha}'\mathbf{z},-\rho) - \Phi_2(\eta A_{j\text{-}1} - \boldsymbol{\beta}'\mathbf{x},\boldsymbol{\alpha}'\mathbf{z},-\rho)],$$

where Φ_2 = bivariate normal CDF.

In the two polar cases, if $j = 1$, $\Phi_2(\eta A_{j\text{-}1} - \boldsymbol{\beta}'\mathbf{x}, \boldsymbol{\alpha}'\mathbf{z}, -\rho) = 0$

and if $j = J$, $\Phi_2(\eta A_j - \boldsymbol{\beta}'\mathbf{x}, \boldsymbol{\alpha}'\mathbf{z}, -\rho) = \Phi(\boldsymbol{\alpha}'\mathbf{z})$.

Derivatives of the log likelihood may be constructed using the results given in Chapter E33 for the bivariate probit model. The BHHH estimator is used for the asymptotic covariance matrix of the estimates.

The fitted values for this model are computed using Bhat's results: Let

$$\delta = 1/(1 - \rho^2)^{1/2},$$

$$\eta = 1/\sigma,$$

$$q = \boldsymbol{\alpha}'\mathbf{z},$$

$$W_m = \eta A_m - \boldsymbol{\gamma}'\mathbf{x}, \quad m = L,U \text{ (limits for the range in which } y \text{ falls)}$$

$$V_m = \delta(q + \rho W_m), \quad m = L,U$$

$$T_m = \delta(W_m + \rho q), \quad m = L,U.$$

Then,

$$E[y^*|\mathbf{x},d=1] = \boldsymbol{\beta}'\mathbf{x} + \sigma\frac{\phi(W_L)\Phi(V_L) - \phi(W_U)\Phi(V_U) + \rho\phi(q)[\Phi(T_U) - \Phi(T_L)]}{\Phi_2(W_U,q,-\rho) - \Phi_2(W_L,q,-\rho)}.$$

E54.10 Parametric Survival Models with Sample Selection

The random parameters model opens the possibility of a sample selection model for parametric survival models. The structure would be the base case parametric model, using the Weibull model as the standard case,

$$h(t_i) = \lambda_i P (\lambda_i t_i)^{P-1}$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i).$$

We accommodate this case by treating the random component as a random constant term in the parametric model. The observation mechanism is now

$$z_i^* = \boldsymbol{\alpha}'\mathbf{z}_\mathrm{i} + u_i, \; z_i = 1(z_i^* > 0)$$

where the correlation between u_i and ε_i is ρ. We assume that the data for the duration model are only observed when $z_i = 1$. The model is fit by full information maximum likelihood. (This means that there is no 'lambda' = the inverse Mills ratio added to the duration model. That treatment is only appropriate for the linear model fit by two step least squares.)

This model is requested by the following command set:

```
PROBIT      ; Lhs = z
            ; Rhs = variables in w
            ; Hold $
SURVIVAL    ; Lhs = logt [, and possibly a censoring indicator]
            ; Rhs = variables in x
            ; Model = one of Weibull, Normal, Loglogistic
            ; Selection
            ; RPM
            ; Fcn = one(n) $
```

(Other controls for the RP models, such as the number of replications, Halton draws, and so on, operate as usual.)

E54.11 A General Approach to Incorporating Selectivity in a Model

Based on the wisdom obtained from Heckman's modification of the linear model, there seems to be a widespread tendency (temptation) to extend that model to other frameworks by mimicking the two step approach used there. Thus, for example, one might fit a Poisson model with sample selection (developed Section E54.4 above), with the following two steps:

Step 1. Fit the probit model for the sample selection equation.

Step 2. Using the selected sample, fit the second step Poisson model merely by adding the inverse Mills ratio from the first step to the Poisson model as an additional independent variable.

This approach is inappropriate for several reasons

- The impact on the conditional mean of the Poisson does not take the form of an inverse Mills ratio. That is specific to the linear model. (See Terza (1995, 1998, 2010).)

- The bivariate normality assumption needed to justify the inclusion of the inverse Mills ratio in the Poisson mean does not appear anywhere in the model.

- The dependent variable, conditioned on the sample selection, is unlikely to have the Poisson distribution needed to use this technique.

Counterparts to these three problems will show up in any nonlinear model. Thus, note, in all of the preceding applications, we have built the selection into the model, rather than attempt to deal with it by dropping the inverse Mills ratio into the model at a convenient point.

The preceding development for the Poisson model suggests a method of incorporating sample selection in a model. The model is based on the premise that the force of 'sample selectivity' is exerted through the behavior of the unobservables in the model. As such, the key to modeling the effect is to introduce the unobservables that might be affected into the model in a reasonable way that maintains the internal consistency of the model itself. In the Poisson model, the standard approach to introducing unobserved heterogeneity is through the conditional mean, specifically,

$$\lambda_i(\varepsilon_i) = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) \text{ where } \varepsilon \sim N[0,\sigma^2].$$

The negative binomial model arises, for example, if it is assumed that the unobserved heterogeneity, ε, has a log gamma distribution. 'Selectivity' would arise if the unobserved heterogeneity in this conditional mean is correlated with the unobservables in the selection mechanism, which is how it is modeled above. We propose a general approach to sample selection – one that we have used at several points above – by modifying the index function model along the lines of the Poisson model analyzed above. The following uses the Poisson model developed earlier as a template, and develops it more generally as an index function model that can be adapted to a specific model framework.

The generic model will take the form

$$
\begin{aligned}
z_i^* &= \boldsymbol{\alpha}'\mathbf{w}_i + u_i \text{ in which } u_i \sim N[0,1] \\
z_i &= \mathbf{1}(z_i^* > 0) && \text{(probit selection equation)} \\
\lambda_i |\, \varepsilon_i &= \boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i,\ \varepsilon_i \sim N[0,1] && \text{(index function with heterogeneity)} \\
y_i \,|\mathbf{x}_i, \varepsilon_i &\sim f(y_i \,|\, \mathbf{x}_i,\varepsilon_i) && \text{(index function model for outcome)} \\
[u_i,\varepsilon_i] &\sim N[(0,1),(1,\rho,1)] \\
y_i,\mathbf{x}_i & \quad \text{are observed only when } z_i = 1.
\end{aligned}
$$

The model given above is broad enough to include most of the models developed in the first 30 chapters of this manual, and most of them yet to follow. Again, the main equation of interest is taken to be an index function model, though, in fact, even that is merely a convenience, and the analysis will carry through if only we have $y_i \mid \varepsilon_i \sim f(\mathbf{x}_i,\sigma\varepsilon_i)$. But, this is more general than we will need.

The log likelihood function for the full model is the joint density for the observed data. When z_i equals one, $(y_i,\mathbf{x}_i,z_i,\mathbf{w}_i)$ are all observed. We seek $f(y_i, z_i=1|\mathbf{x}_i,\mathbf{w}_i)$. To obtain it, proceed as follows:

$$
\begin{aligned}
f(y_i, z_i=1|\mathbf{x}_i,\mathbf{w}_i) &= \int_{-\infty}^{\infty} f(y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i)\, f(\varepsilon_i)d\varepsilon_i \\
&= E_\varepsilon[f(y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\boldsymbol{\varepsilon}_i)].
\end{aligned}
$$

Conditioned on ε_i, z_i and y_i are independent. Therefore,

$$f(y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i,\boldsymbol{\varepsilon}_i) = f(y_i|\mathbf{x}_i,\varepsilon_i)\text{Prob}(z_i=1|\mathbf{w}_i,\varepsilon_i).$$

The first part, $f(y_i\,|\mathbf{x}_i,\varepsilon_i)$ is the conditional index function model, however specified. By joint normality, $f(u_i|\varepsilon_i) = N[\rho\varepsilon_i\,, (1-\rho^2)]$. Therefore, $\text{Prob}(z_i=1|\mathbf{w}_i,\varepsilon_i)$ is

$$\text{Prob}(z_i=1|\mathbf{w}_i,\varepsilon_i) = \Phi\left([\boldsymbol{\alpha}'\mathbf{w}_i + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right).$$

Combining terms and using the earlier approach, the unconditional joint density is

$$f[y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i] = \int_{-\infty}^{\infty} f(y_i|\mathbf{x}_i,\varepsilon_i)\ \Phi\left([\boldsymbol{\alpha}'\mathbf{w}_i + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right) \frac{\exp(-\varepsilon_i^2/2)}{\sqrt{2\pi}}]d\varepsilon_i\,.$$

Let $\quad v = \varepsilon/\sqrt{2}$, $\theta = \sigma\sqrt{2}$, $\tau = \sqrt{2}\,[\rho/\sqrt{1-\rho^2}\,]$, and $\boldsymbol{\gamma} = [1/\sqrt{1-\rho^2}\,]\boldsymbol{\alpha}$.

(Thus, the reverse transformations are $\rho^2 = [\tau^2/(2 + \tau^2)]$, Sgn($\rho$) = Sgn($\tau$), and $\sigma = \theta/\sqrt{2}$.) After making the change of variable and reparameterizing the probability as before, we obtain

$$f(y_i,z_i=1|\mathbf{x}_i,\mathbf{w}_i) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v^2)\, f(y_i|\mathbf{x}_i,v_i)\Phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_i)\, dv_i$$

where the index function model now involves $\lambda_i|\, v_i = \boldsymbol{\beta}'\mathbf{x}_i + \theta v_i$.

E54.11.1 Using Quadrature to Maximize the Log Likelihood

The function in the form above can be approximated with Hermite quadrature since no closed form exists. When z_i equals zero, only $(z_i,\mathbf{w}_i)$ are observed. The contribution to the likelihood function is

$$\text{Prob}(z_i = 0|\mathbf{w}_i) = E_\varepsilon[1 - \text{Prob}(u_i > -\boldsymbol{\alpha}'\mathbf{w}_i|\mathbf{w}_i,\varepsilon_i)] = E_\varepsilon[\text{Prob}(u_i \le -\boldsymbol{\alpha}'\mathbf{w}_i|\mathbf{w}_i,\varepsilon_i)].$$

This provides the probability needed to construct the likelihood function.

$$\text{Prob}(z_i = 0|\mathbf{w}_i,\varepsilon_i) = 1 - \Phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_i/\sqrt{2})$$

so

$$\text{Prob}(z_i=0|\mathbf{w}_i) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v^2)\,\Phi[-(\boldsymbol{\gamma}'\mathbf{w} + \tau v)]dv.$$

Maximum likelihood estimates of $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \tau]$ are obtained by maximizing

$$\log L = \Sigma_{z=0} \log\text{Prob}(z_i=0|\mathbf{w}) + \Sigma_{z=1} \log P(y_i,z_i=1|\mathbf{x},\mathbf{w}).$$

The approximating function is

$$\log L = \sum_{z_i=1} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h f(y_i\,|\,\mathbf{x}_i,v_h)\,\Phi\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h\right)\right]$$
$$+ \sum_{z_i=0} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h\Phi\left(-\boldsymbol{\gamma}'\text{w}_i - \tau v_h\right)\right]$$

where v_h and ω_h are the nodes and weights for the quadrature and

$$f(y_i|\,\mathbf{x}_i, v_h) = \text{the index function model, using } \boldsymbol{\beta}'\mathbf{x}_i + \theta v_h.$$

There are two useful further simplifications to employ. First, since z_i is binary,

$$(1\text{-}z_i) + z_i\, f(y_i|\,\mathbf{x}_i, v_h) = f(y_i|\,\mathbf{x}_i, v_h) \text{ when } z_i = 1 \text{ and } 1 \text{ when } z_i = 0.$$

Second, since the normal distribution is symmetric, the two appearances of the normal CDF above can be combined by using

$$\Phi[(2z_i - 1)(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h)] = \Phi[\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h] \text{ when } z_i = 1 \text{ and } \Phi[-(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h)] \text{ when } z_i = 0.$$

With these two devices, the approximating log likelihood function becomes

$$\log L = \sum_{i=1}^{N} \log\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h\left[(1-z_i) + z_i\, f(y_i\,|\,\mathbf{x}_i,v_h)\right]\Phi\left[(2z_i-1)\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h\right)\right]\right].$$

The BHHH estimator of the asymptotic covariance matrix for the parameter estimates is a natural choice given the complexity of the function. The first derivatives must be approximated as well. For convenience, let

$$P_{ih} = f(y_i|\,\mathbf{x}_i, v_h)$$

$$\Phi_{ih} = \Phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h) \qquad \text{(normal CDF)}$$

and

$$\phi_{ih} = \phi(\boldsymbol{\gamma}'\mathbf{w}_i + \tau v_h) \qquad \text{(normal density)}$$

and to save some notation, denote the individual terms summed in the log likelihood as log L_i. We also take advantage of the result that $\partial P(.,.)/\partial z = P\times\partial \log P(.,.)/\partial z$ for any argument z which appears in the function. Then,

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \Phi_{ih} P_{ih} \frac{\partial \log f(y_i\,|\,\mathbf{x}_i, v_h)}{\partial \lambda_i}\mathbf{x}_i$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \Phi_{ih} P_{ih} \frac{\partial \log f(y_i\,|\,\mathbf{x}_i, v_h)}{\partial \lambda_i} v_h$$

$$\frac{\partial \log L}{\partial \boldsymbol{\gamma}} = \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\phi_{ih}\,\mathbf{w}_i - \sum_{z_i=0} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \phi_{ih}\,\mathbf{w}_i$$

$$\frac{\partial \log L}{\partial \tau} = \sum_{z_i=1} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h P_{ih}\phi_{ih}\,v_h - \sum_{z_i=0} \frac{1}{L_i}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H}\omega_h \phi_{ih}\,v_h$$

Estimates of the structural parameters, $(\boldsymbol{\alpha},\rho,\sigma)$ and their standard errors are computed using the delta method. The main parameter vector, $\boldsymbol{\beta}$, has been estimated explicitly.

E54.11.2 Using Simulation to Maximize the Log Likelihood

Simulation is another effective approach to maximizing the log likelihood function. To set this up, we return to the problem in its untransformed form. Using the simplifications suggested above, the log likelihood function to be maximized is

$$\log L = \sum_{i=1}^{N}\log \int_{-\infty}^{\infty} [(1\text{-}z_i) + z_i\, f(y_i|\mathbf{x}_i,\sigma\varepsilon_i)]\ \Phi[(2z_i-1)\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_i\right)]\ \ \phi(\varepsilon_i)d\varepsilon_i$$

where $\boldsymbol{\gamma} = [1/\sqrt{1-\rho^2}\,]\boldsymbol{\alpha}$, and $\tau = \rho/\sqrt{1-\rho^2}$. (The $\sqrt{2}$ has fallen out of the expression because we are not setting this up for Hermite quadrature.) The log likelihood in this form is an expectation that is amenable to estimation by simulation. The simulated log likelihood would be

$$\log L_S = \sum_{i=1}^{N}\log \frac{1}{R}\sum_{r=1}^{R} [(1\text{-}z_i) + z_i\, f(y_i|\mathbf{x}_i,\sigma\varepsilon_{ir})]\ \Phi[(2z_i-1)\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir}\right)]$$

where ε_{ir} is a set of R random draws from the standard normal population. (We would propose to improve this part of the estimation by using Halton draws instead. See Section R24.7 for details.)

Derivatives of the simulated log likelihood are straightforward. For the *i*th observation,

$$\frac{\partial \log L_{S,i}}{\partial \boldsymbol{\beta}} = \frac{1}{\log L_{S,i}} \frac{1}{R} \sum_{r=1}^{R} z_i \left\{ f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir}) \left[\frac{\partial \log f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir})}{\partial \lambda_\iota} \right] \Phi[(2z_i - 1)(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir})] \right\} \mathbf{x}_i$$

$$\frac{\partial \log L_{S,i}}{\partial \sigma} = \frac{1}{\log L_{S,i}} \frac{1}{R} \sum_{r=1}^{R} z_i \left\{ f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir}) \left[\frac{\partial \log f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir})}{\partial \lambda_i} \right] \Phi[(2z_i - 1)(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir})] \right\} \varepsilon_{ir}$$

$$\frac{\partial \log L_{S,i}}{\partial \boldsymbol{\gamma}} = \frac{1}{\log L_{S,i}} \frac{1}{R} \sum_{r=1}^{R} [(1 - z_i) + z_i f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir})] \phi[(2z_i - 1)(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir})] \mathbf{w}_i$$

$$\frac{\partial \log L_{S,i}}{\partial \tau} = \frac{1}{\log L_{S,i}} \frac{1}{R} \sum_{r=1}^{R} [(1 - z_i) + z_i f(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{ir})] \phi[(2z_i - 1)(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir})] \varepsilon_{ir}$$

To illustrate the technique, we consider constructing a binary logit model subject to sample selection. The immediate obstacle is the lack of a functional form for the joint distribution of a normally distributed ε and the logistically distributed variable that underlies the logit model. We use the template described here, instead.

z_i^* = $\boldsymbol{\alpha}'\mathbf{w}_i + u_i$ in which $u_i \sim N[0,1]$,

z_i = $\mathbf{1}(z_i^* > 0)$ (probit selection equation)

$$\text{Prob}(y_i = 1 \mid \mathbf{x}_i . \varepsilon_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}, \ \varepsilon_i \sim N[0,1]$$

$[u_i, \varepsilon_i] \sim N[(0,1),(1,\rho,1)]$

$y_i, \mathbf{x}_i$ are observed only when $z_i = 1$.

The simulated log likelihood function is

$$\log L_S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left[(1 - z_i) + z_i \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)} \right] \Phi[(2z_i - 1)\left(\boldsymbol{\gamma}'\mathbf{w}_i + \tau\varepsilon_{ir}\right)].$$

The Hermite quadrature method was used to obtain the estimates in the Poisson model applications in Section E54.4.4. The simulation technique was used for the binomial logit model in Section E54.8 and the stochastic frontier model in Section E54.6.

E55: Alternative Sample Selection Equations

E55.1 Introduction

Several of the forms of the selection model which can be estimated with *LIMDEP* depart from Heckman's now canonical form, a linear regression with a binary probit selection criterion model:

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

$$z^* = \boldsymbol{\alpha}'\mathbf{w} + u,$$

$$\varepsilon,u \sim N[0,0,\sigma_\varepsilon^2,\sigma_u^2,\rho].$$

A bivariate classical normal (seemingly unrelated) regressions model applies to the structural equations. The standard deviations are σ_ε and σ_u, and the covariance is $\rho\sigma_\varepsilon\sigma_u$. If the data were randomly sampled from this bivariate population, the parameters could be estimated by least squares, or GLS combining the two equations. However, z^* is not observed. Its observed counterpart is z, which is determined by

$$z = 1 \text{ if } z^* > 0$$

and

$$z = 0 \text{ if } z^* \leq 0.$$

Moreover, values of y and $\mathbf{x}$ are only observed when z equals one. Thus, the model is two steps removed from the two equations seemingly unrelated regressions which would be simple to estimate. The essential feature of the model is that under the sampling rule, $E[y|\mathbf{x},z = 1]$ is not a linear regression.

This chapter will describe a set of alternative specifications for the selection mechanism. The regression part of the model is assumed to be linear, as in the first specification shown in Chapter E53. Some of the models are 'hardwired' as procedures in *LIMDEP*, for example the bivariate probit model. Others require fairly lengthy sets of commands in order to program the computations. We will include these in full in order to present the model and techniques and to demonstrate the range of calculations that can be added to the preprogrammed procedures.

WARNING: Users of these procedures should watch very closely for conflicts between their own variable, namelist, matrix, and scalar names and those which we are using in these programs. Sometimes these can cause subtle errors which will not be picked up by the program. For example, if you use one of our matrix or variable names for your own matrices, then use the program below as is, you may find the wrong calculations being done, for reasons that will not be obvious.

E55.2 The Univariate Probit Model

The standard selection rule in most of the existing programs is the single equation probit model, which you set up with

```
PROBIT      ; Lhs = dependent variable
            ; Rhs = independent variables
            ; Hold $
```

If you have already estimated the probit equation you want to use in a selection model, you can bypass the estimation stage with the following:

```
PROBIT      ; Lhs = z
            ; Rhs = v
            ; Hold
            ; Load ; Start = parameters $
```

Just set up the **PROBIT** command as if the model were to be estimated. When you provide a set of parameters with **; Load**, the estimation stage is skipped. The parameter vector you provide must correspond exactly to the list you provide in the **; Rhs** specification. This method would be useful primarily in a very large data set, in which multiple passes through the data could take a significant amount of time compared to the three needed for the selection model.

E55.3 Bivariate Probit Selection Rule

An extension of the sample selection model which allows a bivariate probit selection equation can be handled just by using **; Hold** with a bivariate probit model. Nothing else needs to be changed. The model is as follows:

$$
\begin{aligned}
y &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon,\\
z_a^* &= \boldsymbol{\alpha}_a'\mathbf{w}_a + u_a,\\
z_b^* &= \boldsymbol{\alpha}_b'\mathbf{w}_b + u_b,\\
z_j &= 1 \text{ if } z > 0 \text{ and } 0 \text{ otherwise for } j = a,b.
\end{aligned}
$$

The random components, ε, u_a, and u_b have a trivariate normal distribution with variances σ^2, 1, and 1, respectively and correlations $\rho_{a\varepsilon}$, $\rho_{b\varepsilon}$ and ρ_{ab}. Estimates are obtained for all parameters of the model. We assume that y and $\mathbf{x}$ are observed only if $z_a = j_a$ and $z_b = j_b$ where j_a and j_b are either 0 or 1. I.e., you can select on any of the four possible combinations of z_a and z_b. The assumed combination is 1,1, but you can change this, as shown below.

If u_a and u_b are correlated, a bivariate probit model applies. The corresponding counterparts to the inverse Mills ratios, the λs, are complicated, but this is all taken care of internally. If u_a and u_b are not correlated, the model is one of two independent selection criteria, which is also easily handled. The estimates are obtained using a method analogous to the single equation selectivity model. In this case, the truncated bivariate normal distribution is needed to compute the estimates. As in the single equation case, negative estimated standard errors can arise in a finite sample. In this instance, OLS standard errors are used. The output contains the standard output for a least squares regression plus a listing of the corrected standard errors and the estimates of the two correlation coefficients.

All of the usual options for single equation models are available, including **; Test:** for restrictions, lists of predicted values, and so on. The **; Fill** option can be used to fill in missing data if the values of the regressors are provided. To estimate this model, you must provide the estimated probit equations through the **; Hold** option. We consider the bivariate probit case first. The set of instructions might look as follows: (You would, of course, substitute your own appropriate variables.)

```
BIVARIATE   ; Lhs = za,zb
            ; Rh1 = one,wa1,wa2
            ; Rh2 = one,wb1,wb2
            ; Hold for SELECT $
SELECT      ; Lhs = hours
            ; Rhs = one,x1,x2,x3,x29 $
```

The estimator assumes that you want to select those observations which have $za = zb = 1$. To use some other criterion, you add to the above

```
; Selection = ja,jb
```

where **ja** applies to the first probit equation in your model and **jb** the second. For example, to select observations with $za = 0$ and $zb = 1$, we would use **; Selection = 0,1**.

Once you estimate and **; Hold** a bivariate probit model, you can use it again without reestimating it. For example, in the preceding specification, we might reestimate the model for observations with $za = zb = 0$ by adding another line to the procedure:

```
SELECT      ; Lhs = hours
            ; Rhs = one,x1,x2,x3,x29
            ; Selection = 0,0 $
```

Remember that the bivariate probit criterion which you **; Hold** is replaced by another **; Hold** command. This means that if you estimate a single equation probit model and use **; Hold**, you will lose the bivariate probit you estimated earlier. Since the bivariate probit model can be time consuming to estimate, the option in the next section may be useful.

E55.3.1 Independent Probit Equations

The method of the previous section provides an easy way to estimate the selection model with two independent selection equations. Just estimate the two probit equations separately by maximum likelihood and pass a zero for the starting value for ρ_{ab}. I.e.,

```
PROBIT      ; ... first  equation $
MATRIX      ; ba = b $
PROBIT      ; ... second equation $
MATRIX      ; bb = b $
BIVARIATE   ; [specifications] ; Start = ba,bb,0 ; Load $
```

This sets up the regression (**SELECT**) command so it can be used as above.

E55.3.2 Loading a Probit Equation

If you have the parameter values from the bivariate probit model stored somewhere, you can use the **BIVARIATE PROBIT** (or just **BIVARIATE**) command just to load these known values, and bypass the estimation step. To do so, there are two ways to proceed. In each, you provide the entire command, exactly as if the model were to be estimated. If you just have the two slope vectors, you can use

```
BIVARIATE   ; Lhs = ... ; Rh1 = ... ; Rh2 = ... ; Start = ba,bb ; Load $
```

Given in this fashion, the command requests the procedure to compute an internal starting value for ρ. This will be attached to your slope vectors and passed on with the **; Load** specification which is now equivalent to **; Hold**. However, if you have your own estimate of the correlation coefficient, just add it to the list of starting values. That is,

```
BIVARIATE   ; Lhs = ... ; Rh1 = ... ; Rh2 = ... ; Start = ba,bb,rhoab ; Load $
```

Now, no computation is done at all. The equation is merely loaded and passed on to **SELECT**.

E55.3.3 Computing Lambda for the Sample Selection Model

The underlying regression model is

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$$

Corr(u_a,ε) = $\rho_{a\varepsilon}$, Corr(u,ε) = $\rho_{b\varepsilon}$. But, (y,$\mathbf{x}$) are only observed when (z_1 = 1, z_2 = 1). Estimation of this model is done by a two step extension of Heckman's method for a single probit selection model. The linear regression is computed using the observed data, with regression of y on $\mathbf{x}$, λ_a and λ_b where the two 'lambda' variables are, in fact, g_a/Φ_2 and g_b/Φ_2 as defined in the next section.

These variables are computed internally during estimation, but not retained anywhere accessible. We are often asked how these can be computed and, moreover, can they be computed for the 'nonselected' observations. Using what is already done above, the computation is actually simple. The full set of computations would look as follows: (This is generic. Only the first two commands that set up the data would be specific to any application.)

```
NAMELIST    ; xa = equation a variables ; xb = equation b variables $
CREATE      ; ya = Lhs variable in equation a
            ; yb = Lhs variable in equation b $
BIVARIATE   ; Lhs = ya, yb ; Rh1 = xa ; Rh2 = xb $
CREATE      ; qa = 2*ya - 1 ; qb = 2*yb - 1 $
CALC        ; ka = Col(xa) ; kba = ka + 1 ; kvar = Row(b) $
MATRIX      ; ba = b(1:ka) ; bb = b(kba:kvar) $
CREATE      ; va = qa*xa'ba ; vb = qb * xb'bb ; rs = qa*qb*rho $
NAMELIST    ; v = va,vb $
CREATE      ; lambdaa = qa*Bv1(v,rs) / Bvn(v,rs)
            ; lambdab = qb*Bv2(v,rs) / Bvn(v,rs) $
```

E55.3.4 Technical Details

For the selectivity model with bivariate probit selection equation, the augmented regression is

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \theta_a \lambda_{ai} + \theta_b \lambda_{bi} + \eta_i.$$

There are three correlation coefficients in the model,

$$\rho_{ab} = \text{corr}(u_a, u_b), \ \rho_{a\varepsilon} = \text{corr}(u_a, \varepsilon), \ \rho_{b\varepsilon} = \text{corr}(u_b, \varepsilon).$$

The bivariate probit model estimates ρ_{ab} in isolation. In the regression model, the parameters are

$$\theta_a = \rho_{a\varepsilon}\sigma_{a\varepsilon}, \ \theta_\text{b} = \rho_{b\varepsilon}\sigma_{b\varepsilon}.$$

The ‘λ‘ variables in the regression are

$$\lambda_a = \phi(w_a)\Phi[(w_b - \rho_{ab}z_a)/(1 - \rho_{ab}^2)^{1/2}] \ / \ \Phi_2$$

where $w_a = -\boldsymbol{\alpha}_a'\mathbf{w}_a$, and likewise for ‘$b$,’ and Φ_2 = bivariate normal CDF, $\Phi(w_a, w_b, \rho_{ab})$. The coefficients are computed by least squares regression of y on $\mathbf{x}$, λ_a, and λ_b. The estimator of the asymptotic covariance matrix is

$$\mathbf{V} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}[\mathbf{X}^{*\prime}(\sigma^2\mathbf{I} - \boldsymbol{\Pi})\mathbf{X}^* + \theta_a^2\mathbf{X}^{*\prime}\mathbf{G}_a\boldsymbol{\Sigma}\mathbf{G}_a'\mathbf{X}^* + \theta_b^2\mathbf{X}^{*\prime}\mathbf{G}_b\boldsymbol{\Sigma}\mathbf{G}_b'\mathbf{X}^*](\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1},$$

where $\mathbf{X}^* = [\ \mathbf{X} : \boldsymbol{\lambda}_a : \boldsymbol{\lambda}_b\ ]$,

$\boldsymbol{\Pi} = \text{diag}(\pi_1, \ldots, \pi_N)$,

$\pi_i = \theta_a^2 w_a \lambda_a + \theta_b^2 w_b \lambda_b + (\theta_a \lambda_a + \theta_b \lambda_b)^2 - [2\theta_a \theta_b - \rho_{ab}(\theta_a^2 + \theta_b^2)]\phi_2/\Phi_2$,

$\boldsymbol{\Sigma}$ = asymptotic covariance matrix for estimates of $[\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_b, \rho_{ab}]$,

and $\mathbf{G}_j = \partial\boldsymbol{\lambda}_j/\partial[\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_b, \rho_{ab}]$, $j = a,b$.

The expressions for the derivatives are exceedingly cumbersome. The estimate of σ^2 is

$$\hat{\sigma}^2 = (1/N)\mathbf{e}'\mathbf{e} - (1/N)\Sigma_i \pi_i.$$

E55.4 A Binary Logit Selection Model

Lee (1983) describes a reformulation of the selection model which allows more general specifications of the criterion equations. The most common application of the techniques in this paper would be the use of a logit instead of a probit equation for the selection criterion equation. This results in a minor modification of the estimation procedure. In the probit case, λ is computed using $\boldsymbol{\alpha}'\mathbf{v}$. For the logit model, we use the transformed variable

$$q = \Phi^{-1}\left[\frac{\exp(\boldsymbol{\alpha}'\mathbf{v})}{1+\exp(\boldsymbol{\alpha}'\mathbf{v})}\right]$$

$$= \Phi^{-1}[\,P_{logit}\,].$$

Then, $\lambda = \phi(q) \,/\, \Phi(q)$ or $-\phi(-q)/\Phi(-q)$ if selection is on z = 0.

Other computations are now the same as before. To use this estimator in *LIMDEP*, simply use

```
LOGIT     ; Lhs = z ; Rhs = list of v ; Hold $
SELECT    ; ... exactly the same as before ... $
```

All necessary modifications – there are very few – are already set up internally. The model specification

; Hold (IMR = name)

is usually used for the probit model to set up a sample selection model. The same parameter may now be used with **LOGIT**, for the same purpose. This variable, λ can be computed as follows:

```
NAMELIST  ; v = ... $
LOGIT     ; Lhs = y ; Rhs = v $
CREATE    ; alphax = v'b           ? computes index α'v
          ; j1 = Inp(alphax)       ? computes Lee's j1
          ; lambda = Lmd(j1,z)     $ computes lambda
```

To illustrate the estimator, we analyze a data set on credit applications. The file (credit.lpj) contains 13,444 observations on applications for a major credit cards. Variables used in the model below include:

cardhldr = whether the application was accepted (0/1),
income = average monthly income,
depdnt = number of dependents in the household,
incper = income / (1 + depdnt),
credmajr = number of major credit cards held,
tradacct = number of merchant credit accounts,
age = age in months,
ownrent = dummy variable for home ownership,
selfempl = dummy variable for self employed,
curntadd = number of months living at current address,
spending = average monthly expenditure.

The regression model analyzes average yearly expenditure. The selection mechanism is, as before, cardholder status.

NAMELIST ; card = one,age,income,ownrent,selfempl,curntadd $
NAMELIST ; spending = one,income,depdnt,incper,credmajr,tradacct $
LOGIT ; Lhs = cardhldr ; Rhs = card ; Hold $
SELECT ; Lhs = avgyrexp ; Rhs = spending $

Results are shown below. Based on the estimate of ρ, cardholder status does not have much impact on average yearly spending.

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable             CARDHLDR
Log likelihood function     -6861.79654
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
CARDHLDR|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[CARDHL=1]
Constant|     .52749***      .08067     6.54  .0000       .36939     .68560
     AGE|    -.00581**       .00241    -2.42  .0157     -.01053    -.00110
  INCOME|     .30124***      .01952    15.43  .0000       .26298     .33949
 OWNRENT|     .27087***      .04803     5.64  .0000       .17672     .36501
SELFEMPL|    -.58068***      .08662    -6.70  .0000     -.75046    -.41090
CURNTADD| .96096D-04         .00037      .26  .7956 -.63079D-03   .82298D-03
--------+--------------------------------------------------------------------
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Logit  selection equation based on CARDHLDR                |
| Selection rule is: Observations with CARDHLDR =  1         |
| Results of selection:                                      |
|                     Data points     Sum of weights         |
| Data set               13444           13444.0             |
| Selected sample        10499           10499.0             |
+------------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model..........................
Two step     least squares regression ...........
LHS=AVGYREXP Mean                 =          .30156
             Standard deviation   =          .39858
             Number of observs.   =           10499
Model size   Parameters           =               7
             Degrees of freedom   =           10492
Residuals    Sum of squares       =         1543.79
             Standard error of e  =          .38359
Fit          R-squared            =          .07373
             Adjusted R-squared   =          .07320
Model test   F[  6, 10492] (prob) =   139.2(.0000)
Standard error corrected for selection       .38372
Correlation of disturbance in regression
and Selection Criterion (Rho)     =          .05371
```

AVGYREXP	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	.03352	.04517	.74	.4581	-.05502	.12206
INCOME	.05541***	.00565	9.80	.0000	.04433	.06649
DEPDNT	.01885***	.00482	3.91	.0001	.00941	.02830
INCPER	.02812***	.00454	6.19	.0000	.01922	.03703
CREDMAJR	.02498**	.01042	2.40	.0165	.00457	.04540
TRADACCT	-.00225***	.00063	-3.59	.0003	-.00349	-.00102
LAMBDA	.02061	.07645	.27	.7875	-.12923	.17044

E55.5 Multinomial Logit Selection Model

Lee (1983) also describes the computation of an estimator for the sample selection model when selection is based on the multinomial logit or discrete choice model of Chapter E37. The following will show how to compute the first model. (Another relevant study on this model is Hay (1980).) We first present some background, since Lee's paper stops at the point of presenting the estimating regression equation. Following Lee, we suppose, then, that z is the selection variable which takes values 0, 1, ..., J for J + 1 outcomes. The model for determination of z is

$$\text{Prob}[z_i=j] = \exp(\boldsymbol{\alpha}_j'\mathbf{w}_i) \,/\, \left[1 + \sum_{j=1}^{J} \exp(\boldsymbol{\alpha}_j'\mathbf{w}_i)\right],$$

where 'i' indexes the observation and 'j' indexes the choice or outcome. This embodies a number of assumptions about the joint and marginal distributions of disturbances in the model, for which reference can be made to Lee's paper. Selection is based on $z_i = j$. For convenience below, we drop the observation subscript. The implied regression equation for estimation derived by Lee is

$$\begin{aligned} y_j &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho_j\sigma_j)\phi[H_j(\boldsymbol{\alpha}_j'\mathbf{w}_i)]/\Phi[H_j(\boldsymbol{\alpha}_j'\mathbf{w}_i)] + \eta_j \\ &= \boldsymbol{\beta}'\mathbf{x}_i + (\rho_j\sigma_j)\lambda_j + \eta_j \\ &= \boldsymbol{\beta}'\mathbf{x}_i + \theta_j\lambda_j + \eta_j. \end{aligned}$$

Our notation differs slightly from Lee's. We use 'H' for his 'J' function (the inverse of the standard normal CDF evaluated at Prob[$z = j$]) to avoid a conflict with our formulation of the logit model. We have also reversed the sign of the second term in the regression to be consistent with our notation elsewhere. This is merely a matter of interpreting ρ. As in Lee's paper, the functions $\phi(t)$ and $\Phi(t)$ are the PDF and CDF of the standard normal distribution. Finally, although the denominator in λ_j is just Prob[$z = j$], it is convenient to have it in the form above when we derive the appropriate standard errors. This full set of computations is fully automated. To fit this model, just use

MLOGIT ; Lhs = z ; Rhs = w; Hold $
SELECT ; Lhs = y ; Rhs = x ; Choice = j $

The **MLOGIT** command may fit a binomial or multinomial logit model. The **; Choice = j** may be omitted in the **SELECT** command if you are selecting observations with choice = 1. For any other choice, you must provide this specification. All of the other features of the selection model available in the standard case, including marginal effects, are supported for this selection mechanism as well.

E55.5.1 Application

To illustrate the estimator, we have contrived an application based on credit data used above. We suppose that cardholders (i.e., those with cardhldr = 1) are further divided into three card types, cardtype = 0,1,2. (The partition is artificial, for the purpose of our simulation.)

CREATE ; cardtype = cardhldr*(Rnd(3)-1) $
REJECT ; cardhldr = 0 $
MLOGIT ; Lhs = cardtype ; Rhs = card ; Hold $
SELECT ; Lhs = avgyrexp ; Rhs = spending ; Choice = 1 $
SELECT ; Lhs = avgyrexp ; Rhs = spending ; Choice = 2 $

Results are shown below for the first selection model. The second model result is similar, but is based on the different subset of the observations. Note that the method of moments based estimate of the correlation coefficient is larger than one, so one is used in the subsequent computations.

```
-----------------------------------------------------------------------------
Multinomial Logit Model
Dependent variable             CARDTYPE
Log likelihood function    -11523.41187
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
CARDTYPE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[CARDTY=1]
Constant|    -.12458         .08836    -1.41  .1586     -.29775    .04860
     AGE|     .00449         .00281     1.60  .1101     -.00102    .01000
  INCOME|     .02549         .01678     1.52  .1287     -.00739    .05838
 OWNRENT|    -.12473**       .05367    -2.32  .0201     -.22991   -.01954
SELFEMPL|    -.26083**       .10933    -2.39  .0170     -.47511   -.04655
CURNTADD|    -.00058         .00041    -1.41  .1594     -.00139    .00023
        |Characteristics in numerator of Prob[CARDTY=2]
Constant|    -.12395         .08877    -1.40  .1626     -.29793    .05003
     AGE|     .00435         .00281     1.55  .1219     -.00116    .00987
  INCOME|     .00069         .01704      .04  .9677     -.03270    .03409
 OWNRENT|    -.11334**       .05400    -2.10  .0358     -.21917   -.00751
SELFEMPL|    -.02944         .10462     -.28  .7784     -.23449    .17561
CURNTADD| .94742D-04         .00041      .23  .8159 -.70298D-03  .89246D-03
--------+--------------------------------------------------------------------
Estimated correlation is outside the range -1 < r < 1. Using 1.0
+--------------------------------------------------------+
| Sample Selection Model                                 |
| MLogit selection equation based on CARDTYPE            |
| Selection rule is: Observations with CARDTYPE =  2     |
| Results of selection:                                  |
|                  Data points     Sum of weights        |
| Data set            10499           10499.0            |
| Selected sample      3438            3438.0            |
+--------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step      least squares regression ............
LHS=AVGYREXP Mean                  =          .30009
             Standard deviation    =          .37459
             Number of observs.    =            3438
Model size   Parameters            =               7
             Degrees of freedom    =            3431
Residuals    Sum of squares        =         442.841
             Standard error of e   =          .35926
Fit          R-squared             =          .07988
             Adjusted R-squared    =          .07827
Standard error corrected for selection        .66895
Correlation of disturbance in regression
and Selection Criterion (Rho)      =         1.00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
AVGYREXP|  Coefficient     Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|    -.88837*       .48195    -1.84  .0653    -1.83298     .05624
  INCOME|     .04312***     .00843     5.11  .0000      .02660     .05964
  DEPDNT|     .02154**      .01088     1.98  .0478      .00021     .04288
  INCPER|     .03321***     .01007     3.30  .0010      .01347     .05295
CREDMAJR|     .02536        .02420     1.05  .2947     -.02207     .07278
TRADACCT|    -.00012        .00142     -.08  .9336     -.00290     .00267
  LAMBDA|     .84996*       .44253     1.92  .0548     -.01739    1.71731
--------+--------------------------------------------------------------------
```

E55.5.2 Technical Details

The two step estimation technique is as follows: (The reader is referred to Lee's paper for some of the relevant background for this.) The first step is to estimate the multinomial logit model by maximum likelihood, retaining the coefficients, estimated asymptotic covariance matrix of these estimates, and the full set of predicted probabilities. Select those observations for which z takes the value in question. (This depends on the application.) For these observations, compute λ_j by obtaining, first, the predicted probability, P_i (Lee denotes this F_j), then

$$\mathrm{H}_j = \Phi^{-1}(P_j),\ \ \lambda_j = \phi(H_j) / \Phi(H_j).$$

In the following, 'i' is the observation, 'j' is the selection choice, $j = 0,...,J$, and K_2 is the number of variables in $\mathbf{w}$.

The second step is to obtain consistent estimates of $\boldsymbol{\beta}$ and θ_j by least squares regression of y_j on $\mathbf{x}$ and λ_j. Denote by $\mathbf{X}_j$ the $N_j{\times}(K_1{+}1)$ matrix of regressors used in this regression including λ_j. Then, the appropriate asymptotic covariance matrix is:

$$\mathbf{C} = (\mathbf{X}_j'\mathbf{X}_j)^{-1}[\sigma_j^2\mathbf{X}_j'(\mathbf{I}\text{-}\rho_j^2\boldsymbol{\Delta}_j)\mathbf{X}_j + \theta_j^2\ \mathbf{F}_j\boldsymbol{\Sigma}\mathbf{F}_j'](\mathbf{X}_j'\mathbf{X}_j)^{-1},$$

where $\delta_{ij} = \lambda_{ij}^2 + H_{ij}\lambda_{ij},\ \ \boldsymbol{\Delta}_j = \mathrm{diag}(\delta_{1j},\delta_{2j},...,\delta_{Nj,j}),$

$$\boldsymbol{\Sigma} = \text{asymptotic covariance matrix of estimated } \boldsymbol{\alpha} = [\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...\boldsymbol{\alpha}_J],$$

$$\mathbf{F}_j = \mathbf{X}_j'\mathbf{G}_\mathrm{j},$$

and $\mathbf{G}_j = N_j{\times}(JK_2) = \partial[N_j{\times}1 \text{ vector of } \lambda\text{s}]/\partial\boldsymbol{\alpha}'.$

This is a matrix of derivatives of the lambdas with respect to the logit parameters. We construct

$$\mathbf{G}_j = [\mathbf{G}_{1j}, \mathbf{G}_{2j}, ..., \mathbf{G}_{Jj}].$$

The ith row of the $N_j \times K_2$ matrix $\mathbf{G}_{sj}$ is

$$\mathbf{g}_{isj}' = (\delta_{ij} / F_{ij}) q_{isj} \mathbf{w}'.$$

The scalar q_{isj} depends on the choice. If selection is on $z = 0$,

$$q_{isj} = - P_{0i} P_{si},\ s=1,...,J.$$

If selection is on z = some other value, say, k, then, for the kth item,

$$q_{ikj} = P_{ik}(1 - P_{ik}),$$

while for all other items,

$$q_{isj} = -P_{ik} P_{sk},\ s = 1,...,J \text{ but not equal to } k.$$

E55.6 Tobit Selection Rule

It is possible to use a tobit or classical regression model as the selection criterion. (See Lee, Maddala, and Trost (1980).) Consider the tobit model first, where we take as the example the standard specification. One might select, for instance on 'hours worked' in a labor supply setting. This variable would be either zero for nonworkers, or positive for the remaining observations. One simple way to estimate this revised model is to reformulate the tobit model as a probit model. We would recode the dependent variable as zero as before and one for the existing nonzero observations. It is straightforward to show that doing so amounts to estimating by probit analysis the parameter vector $\boldsymbol{\alpha}/\sigma$, instead of obtaining the separate estimates of the two parameters with the tobit estimator. This is a practical solution which may sacrifice some efficiency by discarding information on the dependent variable. This would show up in the standard errors of the selection equation's parameter estimates, as the $\boldsymbol{\Sigma}$ in the second part of the bracketed term in the expressions given earlier is the asymptotic covariance matrix for the estimator of the parameters in the selection equation. It is not necessarily true though, that there would be any efficiency loss in a finite sample. The reason is that although the tobit estimator does, asymptotically, estimate $\boldsymbol{\alpha}/\sigma_u$ more efficiently than the probit model, in a finite sample, this nonlinear function of the parameters of the tobit model may be as variable as the estimates from the probit equation.

In a case in which the limit values in the tobit model are not zero and $+\infty$, the preceding is moot. When the censoring point in the tobit model is nonzero, the parameter σ (or its reciprocal) enters λ independently from $\boldsymbol{\beta}$. As such, in computing the asymptotic covariance matrix for the selectivity corrected parameter estimates, it is necessary to account for the variation in the estimate of σ.

We depart from the doubly censored regression model with lower censoring limit L and upper limit U. We use Olsen's reformulation of the tobit model,

$$\boldsymbol{\gamma} = \boldsymbol{\alpha}/\sigma_u,\ \eta = 1/\sigma_u.$$

Then, the auxiliary variable in the selectivity corrected regression will be

$$\lambda = \frac{\phi(\eta L - \boldsymbol{\gamma}'\mathbf{x}) - \phi(\eta U - \boldsymbol{\gamma}'\mathbf{x})}{\Phi(\eta U - \boldsymbol{\gamma}'\mathbf{x}) - \Phi(\eta L - \boldsymbol{\gamma}'\mathbf{x})}.$$

The asymptotic covariance matrix for the slopes in the regression of **y** on **X** and λ would take a familiar form,

$$\mathbf{C} = (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}[\sigma_\varepsilon^2(\mathbf{X}^{*\prime}(\mathbf{I} - \rho^2\boldsymbol{\Delta})\mathbf{X}^{*} + \mathbf{X}^{*\prime}\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}'\mathbf{X}^{*}](\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1},$$

where

$$z_j = \eta j - \boldsymbol{\gamma}'\mathbf{x},\ j = L,U,$$

$$\phi_j = \phi(z_j),\ j = L,U,\ \ \Phi_j = \Phi(z_j),\ j = L,U,$$

$$\delta = \frac{\phi_L z_L - \phi_U z_U}{\boldsymbol{\Phi}_U - \boldsymbol{\Phi}_L},\ \ \boldsymbol{\Delta} = \text{diag}(\delta_1, \delta_2, \ldots),$$

$$\mathbf{X}^{*} = (\mathbf{X} : \boldsymbol{\lambda}),\ \ \mathbf{G} = \text{matrix of derivatives, rows} = \mathbf{g}_i,$$

$$\mathbf{g}_i = (\partial\lambda_i/\partial\boldsymbol{\gamma}' : \partial\lambda_i/\partial\eta),$$

$$\partial\lambda/\partial\boldsymbol{\gamma}' = \delta\mathbf{v}',\ \ \partial\lambda/\partial\eta = [U\phi_U(z_U - \lambda) - L\phi_L(z_L - \lambda)] / (\Phi_U - \Phi_L).$$

In the computation of **C**, $\boldsymbol{\Sigma}$ is the asymptotic covariance matrix for estimates of $[\boldsymbol{\gamma},\eta]$. The tobit program estimates

$$\mathbf{Q} = \text{asymptotic covariance matrix for estimates of } [\boldsymbol{\beta},\sigma],$$

not $\boldsymbol{\Sigma}$. To estimate $\boldsymbol{\Sigma}$ from **Q**, we reverse the usual application of the delta method. Let

$$\mathbf{K} = \begin{bmatrix} \partial\boldsymbol{\gamma}/\partial\boldsymbol{\beta}' & \partial\boldsymbol{\gamma}/\partial\sigma \\ \partial\eta/\partial\boldsymbol{\beta}' & \partial\eta/\partial\sigma \end{bmatrix} = \begin{bmatrix} (1/\sigma)\mathbf{I} & -\left(1/\sigma^2\right)\boldsymbol{\beta} \\ \mathbf{0} & -\left(1/\sigma^2\right) \end{bmatrix}.$$

Then,

$$\boldsymbol{\Sigma} = \mathbf{KQK}'.$$

The following is a program which will do all of the computations for the selectivity model for tobit equations with either lower, upper, or both tails censored. You control this aspect of the computation by specifying the limit values. E.g., for upper censoring at 6.25, you would use $-\infty$ (use -10000) as the lower limit and 6.25 as the upper, etc. The first few lines set up the data for the specific application.

```
CREATE     ; y  = dependent variable in regression
           ; z  = dependent variable in tobit model $
NAMELIST   ; x  = the Rhs in selectivity corrected regression
           ; w = the Rhs in tobit model $
```

Lower and *upper* must now be set to define the lower and upper limits of the selection region for *y*. Either or both may be infinite. For minus or plus infinity, use 10000 or -10000. These may be different from those used in the tobit model. For example, if the limit values used in estimation were zero and one and you want to select those at the upper limit value of one, you would set *lower* = 1 and *upper* = 10000 in the command below. Do note, in all cases, both values must be provided below, even if one of them is plus infinity or minus infinity. To allow these to vary by observations, we use variables, even if the limits do not vary by observation.

```
CREATE      ; Lower = the specification ; Upper = the specification $
```

None of the commands to follow require any modification. First, pick up tobit estimates of σ and β, and transform them to η and γ. Then, get the appropriate asymptotic covariance matrix for the transformed estimates.

```
TOBIT       ; Lhs = z ; Rhs = w ; Par ; Limits = lower,upper $
CALC        ; kw = Col(w) ; eta = 1/s ; etasq = -eta^2 $
MATRIX      ; beta = b(1 : kw)
            ; k11 = eta * Iden(kw) ; k12 = etasq * beta
            ; k21 = Init(1,kw,0) ; k22 = etasq
            ; k = [k11,k12 / k21,k22]
            ; sg = k * varb * k’ ; gamma = eta * beta $
```

Isolate the selected sample.

```
REJECT      ; y <= lower | y >= upper $
```

Create the variables needed to compute the estimates and asymptotic covariance matrix.

```
CREATE      ; wgamma = w’gamma
            ; zu = eta * upper - wgamma ; fu = N01(zu) ; pu = Phi(zu)
            ; zl  = eta * lower  - wgamma ; fl  = N01(zl) ; pl = Phi(lz)
            ; pul = pu - pl
            ; lambda = (fl - fu) / pul
            ; delta = (zl*fl - zu*fu) / pul - lambda ^ 2
            ; dldeta = (upper*fu*(zu - lambda) - lower*fl *(zl - lambda)) / pul $
```

Compute the slopes in the selectivity corrected regression. Pick up three parts: *c*1 = slope on *lambda*, *sigma*(1) = OLS disturbance standard deviation, *b*1 = full set of slopes including *c*1. Then, compute consistent estimates of σ^2 and ρ^2.

```
NAMELIST    ; xl = x,lambda $
REGRESS     ; Lhs = y ; Rhs = xl $
MATRIX      ; b1 = b $
CALC        ; l1 = b(kreg) ; csq = c1 ^ 2
            ; sigma1 = Sqr(s ^ 2 - csq * Xbr(delta))
            ; rho12 = c1 / sigma1 $
```

Compute the components of the variance matrix and $\theta = \sigma^2(1 - \rho^{2*})$.

```
CREATE        ; theta = sigma1 - rho12 ^ 2 * delta ; delta = Abs(delta) $
```

Compute the matrix components. We must deal with an extra column because of σ.

```
MATRIX        ; psi1 = xl' [theta]v
              ; f1 = xl' [delta]v ; f2   = xt'dldeta ; f = [f1,f2]
              ; c = xl'[theta]xl + csq * f * sg * f'
              ; f = <xt'xt> * c * <xt'xt>
              ; Stat(b1,f,xl)  $
```

E55.7 Ordered Probit Selection Rule

This program computes the regression coefficient estimates and the appropriate asymptotic covariance matrix for a sample selection model based on the ordered probit model. The ordered probit model is:

$$
\begin{aligned}
z^* &= \boldsymbol{\alpha}'\mathbf{w} + u,\\
z &= 0 \ \text{ if } -\infty \quad z^* \le 0,\\
& \quad 1 \ \text{ if } 0 < z^* \le \mu_1,\\
& \quad 2 \ \text{ if } \mu_1 < z^* \le \mu_2,\\
& \quad \text{and so on,}\\
& \quad J \ \text{ if } \mu_{J-1} < z^* \le +\infty.
\end{aligned}
$$

z^* is not observed; z is its observed counterpart. The disturbance, u is assumed to be distributed as standard normal. (See Greene and Hensher (2010) for details on the ordered probit model.) The equation of interest is

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

where ε is normally distributed with mean zero, standard deviation σ and correlation ρ with u. Data on y are only observed when z takes a particular value, so the selection mechanism is

y is only observed when $z = j$ for some j in $(0,1,...,J)$.

The estimation of this model by a two step procedure follows exactly the steps in Heckman (1979) and Greene (1981), which provide the standard results for the case in which $J = 1$ (simple probit model). The steps are

Step 1. Estimate the ordered probit by MLE using all observations.

Step 2. Select the observations for the regression.

Step 3. Estimate the primary equation by OLS including the correction term $E[\varepsilon \mid z = j]$.

Step 4. Correct the estimated asymptotic covariance matrix of the estimates.

The procedure follows, with annotation provided within the commands. The following must be set by the user prior to using this routine to set up the data:

```
NAMELIST  ; w = the Rhs variables in the ordered probit $
CREATE    ; z  = the Lhs variable in the ordered probit $
NAMELIST  ; x  = the Rhs variables in the regression model $
CREATE    ; y  = the Lhs variable in the regression model $
CALC      ; j  = the value on which sample selection is based $
```

Estimate the ordered probit model and collect results. The number of values taken by the dependent variable in the ordered probit model is $JP = J+1$. Retrieve the dimensions and estimates from ordered probit. KP is the number of variables in the ordered probit, $KP1 = KP+1$, M = number of threshold parameters, L = number of parameters estimated. Retrieve the slope vector as *alpha*.

```
ORDERED   ; Lhs = z ; Rhs = w ; Par $
CALC      ; Nolist ; jp = Max(z) + 1 ; jp1 = jp + 1
          ; kp = Col(w) ; kp1 = kp + 1 ; m = jp - 2 ; l = kp + m $
MATRIX    ; alpha = b(1 : kp) $
```

The full threshold vector is $-\infty,0,\mu_1,\mu_2,...,\mu_{J-1},+\infty$. The name *mu* is used by the ordered probit program, so we use *mua*.

```
MATRIX    ; u1 = [-10000 / 0 ] ; u2 = b(kp1:l ) ; u3 = [10000] ; mua = [u1/u2/u3] $
```

The covariance matrix for this full parameter vector, including thresholds with $(-\infty,0)$ embedded and ∞ at the end is as follows: (the four parts are KP, 2, M and 1 parameter, respectively)

$$Var\begin{pmatrix}\hat{\alpha}\\ -\infty,0\\ \hat{\mu}\\ \infty\end{pmatrix}=\begin{bmatrix}\Sigma_{\alpha\alpha} & \mathbf{0} & \Sigma_{\alpha\mu} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \Sigma_{\mu\alpha} & \mathbf{0} & \Sigma_{\mu\mu} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0\end{bmatrix}$$

```
MATRIX    ; z11 = varb(1:kp, 1:kp)
          ; z21 = Init(2,kp,0) ; z22 = [0,0/0,0]
          ; z31 = varb(kp1:l, 1:kp) ; z32 = Init(m,2,0) ; z33 = varb(kp1:l, kp1,l)
          ; z41 = Init(1,kp,0) ; z42 = [0,0] ; z43 = Init(1,M,0) ; z44 = [0]
          ; v = [z11 / z21,z22 / z31,z32,z33 / z41,z42,z43,z44] $
```

Select the sample.

```
INCLUDE   ; New ; z  = j $
```

Construct some variables needed for the regression. For selection on z = j, $E[\varepsilon|z=j] = \rho\ \sigma\ \lambda$, $Var[\varepsilon|z=j] = \sigma^2(1 - \rho^2\ \delta)$.

```
CALC      ; j1 = j + 1 ; j2 = j + 2 $
CREATE    ; aj1 = mua(j1) - w'alpha ; aj = mua(j2) - w'alpha
          ; dj1 = N01(aj1) ; dj = N01(aj) ; fj1 = Phi(aj1) ; fj = Phi(aj)
          ; lambda = (dj1 - dj) / (fj - fj1)
          ; delta = (aj1*dj1 - aj*dj) / (fj - fj1) - lambda ^ 2 $
```

The regression is computed by regressing y on $\mathbf{x}$ and λ. Let c be the coefficient on λ. We estimate the residual variance with $s^2 = e'e/N(j) - c^2\ \bar{\delta}$ (the same as in the simpler case). Then the correlation between the regression disturbance and the structural disturbance in the ordered probit is estimated with $\rho^2 = c^2 / s^2$.

```
NAMELIST    ; xl = x,lambda $
REGRESS     ; Lhs  = y ; Rhs = xl $
CALC        ; p = Col(xl) ; c  = b(p)
            ; s2 = sumsqdev / nreg - c^2 * Xbr(delta)
            ; rhosqd = c^2 / s2 $
```

The asymptotic covariance matrix for the estimates is

$$\mathbf{C} \ = \ s^2\ (\mathbf{X'X})^{-1}[\mathbf{X'}(\mathrm{I}\text{-}\rho^2\Delta)\mathbf{X} + \rho^2(\mathbf{X'G})\boldsymbol{\Sigma}(\mathbf{G'X})]\ (\mathbf{X'X})^{-1}$$

where $\mathbf{X}$ includes lambda, $\boldsymbol{\Delta}$ is a diagonal matrix of δs, and $\mathbf{G}$ is a matrix whose columns are the derivatives of the lambda variable with respect to the parameters $[\boldsymbol{\alpha},\boldsymbol{\mu}]$. $\mathbf{G}$ is more complicated here than in the standard probit case. The first **KP** columns of $\mathbf{G}$ are $\mathbf{-\Delta W}$ where $\mathbf{W}$ is the $n\times KP$ matrix of regressors in the ordered probit model. We have augmented the slope parameter vector with M+3 values, three of which are zeros and $M = J$-1 are elements in the estimated threshold vector.

```
CREATE      ; pj1 = (j > 1) * dj1 / (fj - fj1) * (lambda - aj1)
            ; pj = (j > 0) * (j < (jp-1)) * ( - dj/(fj - fj1) * (lambda - aj)) $
CREATE      ; h = 1 - rhosqd * delta $
```

Note that if j = 0, both of these values are zero, while if j = 1, the first is zero, and if $j = J$, the second is zero. These are then inserted as the (KP+2)+(j-1) and (KP+2)+j columns of $\mathbf{G}$. The remaining columns of $\mathbf{G}$ are columns of zeros. To assemble this, we work with $\mathbf{X'G}$ rather than $\mathbf{G}$ itself and use some tricks from matrix algebra. Let $\mathbf{X'G}$ consist of two parts, $\mathbf{X'G1}$ and $\mathbf{X'G2}$. $\mathbf{X'G1}$ is just -$\mathbf{X'\Delta W}$, which is easy to get. $\mathbf{X'G2}$ is the $P\times(J+2)$ matrix of moments defined using the variables above. Suppose the two columns are $\mathbf{X'p1}$ and $\mathbf{X'p}$. We have to place these two columns into $\mathbf{G}$ in the right place. The following will do so without requiring the user to modify the program. Let $\mathbf{H}$ be a $P\times3$ column matrix consisting of $\mathbf{X'p1}$, $\mathbf{X'p}$, and a third column filled with zeros. Then, the full $\mathbf{X'G}$ matrix that we need is just $\mathbf{HR}$ where $\mathbf{R}$ is the matrix of zeros. The first row is JP+1 zeros save for a one in the column corresponding to where we want $\mathbf{X'p1}$ to be in $\mathbf{X'G}$. The second row is defined likewise for $\mathbf{X'p}$.

```
MATRIX      ; xp1 = xl'pj1 ; xp = xl'pj ; zero = Init(p,1,0)
            ; r = Init(3,jp1,0) ; r(1,j1) = 1 ; r(2,j2) = 1
            ; xpp = [xp1,xp,zero]
            ; xg1 = xl'[delta]w ; xg2 = xpp * r ; xg = [xg1,xg2] $
```

Obtain the corrected covariance matrix and display the results.

```
MATRIX      ; vc = xl'[h]xl + rhosqd * xg * v * xg'
            ; vc = s2 * <xl'xl> * vc * <xl'xl>
            ; Stat(b,varb,xl)      ? (uncorrected results)
            ; Stat(b,vc,xl)        $ (corrected results)
```

E56: Treatment Effects and Switching Regressions

E56.1 Introduction

The essential form of the treatment effects 'model' is a measure of an outcome, y, and an input, say z, which takes the form of some treatment – income equations with a college attendance dummy variable, or wage equations with training program participation dummy variables are common examples. Some of the more widely used methods are programmed in *LIMDEP*. This chapter will detail several models that are related to this type of analysis.

A variety of procedures are presented here. This chapter examines some formal regression approaches that specify the 'treatment effect' as an endogenous dummy variable in a model. The switching regressions and mover stayer models describe the effect of 'treatment' as a change in regime – that is, as a change in the applicable regression model. These are more general than the endogenous dummy variable models, but they are, as well, fully parameterized.

E56.2 The Mover Stayer Model

A fully parameterized model for treatment effects is the structural model

$$y_i = Y_1 = \boldsymbol{\beta}_1'\mathbf{x}_i + \varepsilon_{i1} \text{ when } z_i = 1$$

$$y_i = Y_0 = \boldsymbol{\beta}_0'\mathbf{x}_i + \varepsilon_{i0} \text{ when } z_i = 0$$

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + \mathbf{u}_i, \; z_i = 1(z_i^* > 0).$$

In this model, z_i represents the presence ($z_i = 1$) or absence ($z_i = 0$) of the treatment. In this instance, a different model applies in the two 'states.' This has been labeled a 'mover stayer' model in studies of income of migrants (movers) and nonmigrants (stayers). The model can be estimated for the purpose of examining the various coefficients. However, in recent treatments (such as Heckman, Tobias and Vytlacil (2003)) the interesting feature of the model is what it can reveal about 'treatment effects,' such as $E[y|\mathbf{x},\mathbf{w},z=1]$ - $E[y|\mathbf{x},\mathbf{w},z=0]$.

This is an application of the mover stayer model. (See Willis and Rosen (1978), Lee (1978), Robinson and Tomes (1982), and Nakosteen and Zimmer (1980).) The structural equations of the model are:

$$y_1 = \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1 \text{ (may be a tobit model),}$$

$$y_0 = \boldsymbol{\beta}_0'\mathbf{x}_0 + \varepsilon_0 \text{ (may be a tobit model),}$$

$$c = \boldsymbol{\alpha}'\mathbf{w} + u_c,$$

$$z^* = y_1 - y_0,$$

$$z = 1 \text{ if } z^* > c \text{ and } z = 0 \text{ if } z^* \leq c,$$

$$y = y_1 \text{ if } I = 1 \text{ and } y = y_0 \text{ if } I = 0.$$

For an example (the setting of the Robinson and Tomes study), suppose y_j is the market wage obtainable at location '*j*' while '*c*' is the cost of moving from initial state 1 to alternative state 0. Observed wage is *y*, which will be in state 0 if the premium I^* exceeds the cost of the move, *c*. Otherwise, *y* is observed in state 1 and no transition of state takes place. Variants of this model can be applied in a variety of situations, as suggested by the sampling of the literature noted above.

Observed data consist of *y*, $\mathbf{x}_1$, $\mathbf{x}_0$, *z*, and **w**. The disturbances are assumed to be joint normally distributed. *I* is an indicator of whether the individual 'moves' ($I = 1$) or 'stays' ($I = 0$). The sample selectivity model described earlier applies here with only minor variation, and, as noted, could be applied separately to each structural equation. (The two step estimation techniques are presented in Lee (1976).) This section presents a FIML estimator for the full model. We will denote the model with endogenous switching, whether or not it fully conforms to the structural model shown above, as the mover stayer model.

NOTE: The two regression equations in the mover stayer model may be tobit models.

E56.2.1 Sample Selection Models

With an assumption of trivariate normality,

$$\begin{pmatrix}\varepsilon_{i1}\\ \varepsilon_{i0}\\ u_i\end{pmatrix} \sim N_3\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & 0 & \rho_1\sigma_1\\ 0 & \sigma_0^2 & \rho_0\sigma_0\\ \rho_1\sigma_1 & \rho_0\sigma_0 & 1\end{pmatrix}\right].$$

This is actually precisely the sample selection model developed in Chapter E52 in each of the two regimes. (Note that in the trivariate normal distribution, it is assumed that ε_{i0} and ε_{i1} are uncorrelated. This is not a restriction. The sample will never contain individuals who exist in both states, so a nonzero correlation could never be estimated with any sample data. The zero 'assumption' is merely a convenient notation. This unidentified correlation will not play a role in any estimation.

Each regression equation can be treated separately, with the probit model, as a sample selection model. The conditional mean functions are

$$E[y_i \mid \mathbf{x}_i, z_i = 1] \;=\; \boldsymbol{\beta}_1'\mathbf{x}_1 \;+\; (\rho_1\sigma_1)\left(\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}_1)}{\Phi(\boldsymbol{\alpha}'\mathbf{w}_1)}\right) \quad \text{if } z = 1$$

$$E[y_i \mid \mathbf{x}_i, z_i = 0] \;=\; \boldsymbol{\beta}_0'\mathbf{x}_0 \;+\; (\rho_0\sigma_0)\left(\frac{-\phi(\boldsymbol{\alpha}'\mathbf{w}_0)}{\Phi(-\boldsymbol{\alpha}'\mathbf{w}_0)}\right) \text{ if } z = 0,$$

so, the familiar two step estimator can be used for each equation. The commands would be

```
PROBIT      ; Lhs = z
            ; Rhs = variables in w
            ; Hold $
SELECT      ; Lhs = y
            ; Rhs = variables in x $ for the first equation
SELECT      ; Lhs = y
            ; Rhs = variables in x ? for the second equation
            ; Limits = 1 $
```

E56.2.2 Commands for the Mover Stayer Model

All relevant results for the estimated selection models above are those described in Section E52.2. However, one can fit the entire equation system at once, including the probit equation, by full information maximum likelihood. The commands will be as follows: The estimator requires several steps to set it up, but these are routine. The essential commands are as follows:

```
PROBIT     ; Lhs = z ; Rhs = w ; Hold $
SELECT     ; Lhs = y ; Rhs = x1 $
MATRIX     ; beta1 = bsr1 $
SELECT     ; Lhs = y ; Rhs = x0 ; Limits = 1 $
MATRIX     ; beta0 = bsr0 $
SELECT     ; Lhs = y ; Rh1 = x1 ; Rh2 = x0
           ; MLE ; All ; Start = beta1, beta0 $
```

The right hand sides of the two regressions can be different if desired. However, the formal model has the same regressors in both equations. The last select command may also include

; Tobit

if the model is a pair of tobit equations. The first two sample selection models are estimated just to get the complete set of starting values. Since this model requires several steps, there is no command builder for the mover stayer model.

NOTE: You may use **; Rst = list** to impose restrictions anywhere in the mover stayer model.

WARNING: *Do not fit the first two selection equations as tobit equations, even if the model actually is a tobit model (as specified by* ***; Tobit*** *in the final command).*

The two step linear regressions (**SELECT**) are needed to get the proper starting values put in the right place for *LIMDEP* to find them when the next command is carried out. Also, note that after each linear selection model is estimated, we pick up the matrix *bsrj* where *j* is 1 or 0, depending on what the selection rule was with respect to *z*. After estimating the selection model, *LIMDEP* automatically creates either *bsr*1 or *bsr*0 but the *bsr* matrix replaces the previous one. So, after the first **SELECT** command, *bsr*1 will be in your matrix work area, but *bsr*0 will not. After the **SELECT** command which selects on $z = 0$ (i.e., **; Limits = 1**), *bsr*0 will exist, but *bsr*1 will not. To make sure that the values are saved, we follow each **SELECT** command with a **MATRIX** command to make a copy of the coefficient vector in a place where it will not be overwritten. These matrices are constructed as follows:

$$bsrj = [\text{Estimate of } \boldsymbol{\beta} \text{ without the coefficient on } \lambda, \text{ or } \sigma_j, \rho_{uj}].$$

They are, thus, constructed precisely with the configuration needed for the mover stayer model. If you want to provide a different set of starting values, the full set you need is

$$\boldsymbol{\theta} = [\boldsymbol{\beta}_1, \sigma_1, \rho_{u1}, \boldsymbol{\beta}_0, \sigma_0, \rho_{u0}].$$

Note in the command template above how *beta*1, *beta*0 is used to supply exactly these values to the **SELECT** command.

For imposing restrictions with **; Rst = list**, it is necessary to rearrange the parameter vector. *(This is the only instance in LIMDEP's estimators that the parameter vector used to impose restrictions differs from the one that is set up with your starting values.)* For imposing restrictions, use the modified parameter vector, $\boldsymbol{\theta}^* = [\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \sigma_0, \rho_{u0}, \sigma_1, \rho_{u1}]$. Other options for the mover stayer model are the same as in the list in Section E56.8 for the switching regressions model.

E56.2.3 Results for the Mover Stayer Model

Output from the mover stayer model is the same as the sample selection model estimated by maximum likelihood. Since the single equation, linear regressions have (presumably) preceded this command, the program proceeds directly to the maximum likelihood procedure, without a first round least squares estimation. Final output includes the log likelihood and a guide to the partitioning of the parameter vector.

Predictions for the mover stayer model are computed exactly the same as for the sample selection model In this case, we compute

$$\hat{y} = \boldsymbol{\beta}_1'\mathbf{x}_1 + (\rho_1\sigma_1)[\phi/\Phi] \qquad \text{if } z = 1$$

and

$$\hat{y} = \boldsymbol{\beta}_0'\mathbf{x}_0 + (\rho_0\sigma_0)[-\phi/(1-\Phi)] \quad \text{if } z = 0.$$

Recall that the coefficient on λ in the linear regression is an estimate of $\rho_j\sigma_j$, j = 0,1. In the mover stayer model, these are estimated separately, and both pairs of estimates will appear in the results. Note, however, since this is the maximum likelihood estimator, the parameters ρ_j and σ_j are estimated separately. There is no separate estimate of the product produced or displayed.

The following results are saved by this estimator:

Scalars: *ybar*, *sy*, *logl*, $s = \sigma_1$, $rho = \rho_{1u}$, *sigma*0, *rho*0*u*.

Matrices: *b* contains the entire vector of parameters estimated. In order, this is:

$\boldsymbol{\alpha}$ = parameters of the probit equation,
$\boldsymbol{\beta}_1$ = parameters in first regression,
$\boldsymbol{\beta}_2$ = parameters in second regression,
$[\sigma_0, \rho_{0u}, \sigma_1, \rho_{1u}]$ = ancillary parameters.

varb is the full asymptotic covariance matrix.

To obtain specific parts of the parameter vector, use the matrix **; Part** function to extract them.

E56.2.4 Application

The following will simulate the conditions of the switching regressions and mover stayer models in order to demonstrate the output that results. The commands were executed all at once from the editor.

```
SAMPLE    ; 1-500 $
CALC      ; Ran(12345) $
CREATE    ; x1 = Rnn(0,1)              ? regressor for equation 1
          ; x0 = Rnn(0,1)              ? regressor for equation 0
          ; e1 = Rnn(0,1)              ? disturbance for equation 1
          ; e0 = .5*e1+.5*Rnn(0,1)     ? e for equation 0, correlated
          ; u = Rnn(0,1)+.5*(e1+e0)    ? u for endogenous selection
          ; w = Rnn(0,1)               ? regressor for selection equation
          ; z = w+u                    ? underlying regression for probit
          ; z = z > 0                  ? binary variable for probit
          ; y1 = x1+e1                 ? structural variable, y1*
          ; y0 = x0+e0                 ? structural variable, y0*
          ; If(y1 < y0) ys = y1        ? choose minimum of y1*, y0*
          ; (Else) ys = y0
          ; yms = z*y1+(1-z)*y0        $ Lhs for mover stayer model
PROBIT    ; Lhs = z ; Rhs = one,w ; Hold $
SELECT    ; Lhs = yms ; Rhs = one,x1,x0 $
MATRIX    ; b1 = bsr1 $
SELECT    ; Lhs = yms ; Rhs = one,x1,x0 ; Limits = 1 $
MATRIX    ; b0 = bsr0 $
SELECT    ; Lhs = yms
          ; Rh1= one,x1,x0 ; Rh2= one,x1,x0
          ; Start = b1,b0 ; MLE ; All $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Z
Log likelihood function      -269.37019
Results retained for SELECTION model.
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Z|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .00342      .06254      .05  .9563     -.11915     .12600
       W|     .87139***   .08062    10.81  .0000      .71337    1.02941
--------+--------------------------------------------------------------------
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on Z                       |
| Selection rule is: Observations with Z        =  1         |
| Results of selection:                                      |
|                  Data points     Sum of weights            |
| Data set             500               500.0               |
| Selected sample      239               239.0               |
+------------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step      least squares regression ............
LHS=YMS       Mean                     =        .32104
              Standard deviation       =       1.28792
              Number of observs.       =           239
Model size    Parameters               =             4
              Degrees of freedom       =           235
Residuals     Sum of squares           =       164.802
              Standard error of e      =        .83743
Fit           R-squared                =        .57544
              Adjusted R-squared       =        .57002
Standard error corrected for selection          .89237
Correlation of disturbance in regression
and Selection Criterion (Rho)          =        .47568
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YMS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .11713         .10888     1.08  .2820     -.09627    .33052
      X1|    1.01122***      .05769    17.53  .0000      .89815   1.12429
      X0|    -.05795         .05237    -1.11  .2685     -.16060    .04470
  LAMBDA|     .42448***      .14287     2.97  .0030      .14446    .70450
--------+--------------------------------------------------------------------
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on Z                       |
| Selection rule is: Observations with Z        =  0         |
| Results of selection:                                      |
|                  Data points     Sum of weights            |
| Data set              500              500.0               |
| Selected sample       261              261.0               |
+------------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model............................
Two step      least squares regression ............
LHS=YMS       Mean                     =       -.23974
              Standard deviation       =       1.15506
              Number of observs.       =           261
Model size    Parameters               =             4
              Degrees of freedom       =           257
Residuals     Sum of squares           =       99.8946
              Standard error of e      =        .62345
Fit           R-squared                =        .70754
              Adjusted R-squared       =        .70412
Standard error corrected for selection          .68682
Correlation of disturbance in regression
and Selection Criterion (Rho)          =        .59235
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YMS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    -.00384         .07759     -.05  .9606     -.15592    .14824
      X1|     .00844         .03943      .21  .8305     -.06884    .08573
      X0|     .98204***      .03959    24.81  .0000      .90444   1.05963
  LAMBDA|     .40684***      .10918     3.73  .0002      .19285    .62083
--------+--------------------------------------------------------------------
```

```
+--------------------------------------------+
| Sample Selection Model                     |
| Probit selection equation based on Z       |
| MOVER/STAYER model (MLE). LHS= YMS         |
+--------------------------------------------+
Normal exit:  16 iterations. Status=0, F=    812.0407

-----------------------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                    YMS
Log likelihood function      -812.04069
MOVER/STAYER model (MLE). LHS= YMS
FIRST  2 estimates are probit equation.
Next   3 slopes are for the Y=1 equation.
Next   3 slopes are for the Y=0 equation.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YMS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Selection (probit) equation for Z
Constant|     .01933         .06220      .31  .7559     -.10257    .14123
       W|     .87198***      .08241    10.58  .0000      .71045   1.03350
        |Corrected regression, Regime 1
Constant|     .01926         .09946      .19  .8465     -.17568    .21420
      X1|    1.00866***      .05500    18.34  .0000      .90086   1.11647
      X0|    -.05293         .05800     -.91  .3615     -.16662    .06076
        |Corrected regression, Regime 0
Constant|    -.03587         .07531     -.48  .6338     -.18348    .11173
      X1|     .00835         .04375      .19  .8485     -.07739    .09410
      X0|     .98324***      .03983    24.68  .0000      .90517   1.06132
        |Variance parameters
SIGMA(0)|     .67448***      .03828    17.62  .0000      .59945    .74952
RHO(0,u)|     .51879***      .12893     4.02  .0001      .26609    .77148
SIGMA(1)|     .92669***      .05794    15.99  .0000      .81313   1.04025
RHO(1,u)|     .61988***      .11638     5.33  .0000      .39178    .84799
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E56.2.5 Technical Details

The log likelihood function for the mover stayer model is that of the sample selection model applied to each regime separately. Thus, we maximize

$$\log L \;=\; \sum\nolimits_{z=1} \log\left[\frac{\exp\left(-(1/2)\varepsilon_{i1}^2/\sigma_1^2\right)}{\sigma_1\sqrt{2\pi}}\Phi\left(\frac{\rho_1\varepsilon_{i1}/\sigma_1 \;+\; \boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho_1^2}}\right)\right] \;+$$

$$\sum\nolimits_{z=0} \log\left[\frac{\exp\left(-(1/2)\varepsilon_{i0}^2/\sigma_0^2\right)}{\sigma_0\sqrt{2\pi}}\Phi\left(-\left(\frac{\rho_0\varepsilon_{i0}/\sigma_0 \;+\; \boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho_0^2}}\right)\right)\right].$$

E56.2.6 Treatment Effects

There are several treatment effects that one can identify in the model. (See Heckman, Tobias and Vytlacil (2003) for example.) The obvious candidate is the 'treatment effect,'

$$TE = E[Y_1|z=1] - E[Y_0|z=0]$$

$$= \boldsymbol{\beta}_1'\mathbf{x}_1 + (\rho_1\sigma_1)\left(\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}_1)}{\Phi(\boldsymbol{\alpha}'\mathbf{w}_1)}\right) - \boldsymbol{\beta}_0'\mathbf{x}_0 - (\rho_0\sigma_0)\left(\frac{-\phi(\boldsymbol{\alpha}'\mathbf{w}_0)}{\Phi(-\boldsymbol{\alpha}'\mathbf{w}_0)}\right).$$

The problem with this measure is that it refers to different people – no one can be in both states. Nonetheless, it could be averaged over all individuals under the assumption that the treatment assignment is random at least with respect to **x** (that is the point of the previous section). For an individual selected at random from the entire population, the 'average treatment effect' is

$$ATE = \boldsymbol{\beta}_1'\mathbf{x} - \boldsymbol{\beta}_0'\mathbf{x} = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x}.$$

On the other hand, perhaps more interesting is the 'treatment effect on the treated,' which is

$$ATT = \mathrm{E}[\mathrm{Y}_1|z=1] - \mathrm{E}[\mathrm{Y}_0|z=1]$$

$$= \boldsymbol{\beta}_1'\mathbf{x} + (\rho_1\sigma_1)\left(\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}_1)}{\Phi(\boldsymbol{\alpha}'\mathbf{w}_1)}\right) - \boldsymbol{\beta}_0'\mathbf{x} + (\rho_0\sigma_0)\left(\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}_1)}{\Phi(\boldsymbol{\alpha}'\mathbf{w}_1)}\right)$$

$$= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x} + [(\rho_1\sigma_1) - (\rho_0\sigma_0)]\left(\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}_1)}{\Phi(\boldsymbol{\alpha}'\mathbf{w}_1)}\right).$$

Heckman et al. define as well, the 'local average treatment effect' which is the expected outcome gain for those induced to receive treatment through a change in the instrument from w_k to w_k^+. The variable w_k is assumed to change the treatment decision but not to directly affect the outcomes. Define, then, **w** to be the original vector and $\mathbf{w}^+$ to be the changed vector such that the one element has changed and moreover, $\boldsymbol{\alpha}'\mathbf{w}^+ > \boldsymbol{\alpha}'\mathbf{w}$, so that the margin increases the probability of choosing the treatment. Then, the local average treatment effect is

$$LATE = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x} + [(\rho_1\sigma_1) - (\rho_0\sigma_0)]\left[\frac{\phi(\boldsymbol{\alpha}'\mathbf{w}^+) - \phi(\boldsymbol{\alpha}'\mathbf{w})}{\Phi(\boldsymbol{\alpha}'\mathbf{w}^+) - \Phi(\boldsymbol{\alpha}'\mathbf{w})}\right].$$

Unfortunately, both sign and magnitude of these quantities are completely ambiguous – only the actual computation with the data can reveal either. Signs of the coefficients are uninformative.

These results can be computed easily from the regression results. The following program does the computation. Its length is due to the need to collect quite a few specific inputs to the functions. We begin with the same data setup used previously. Your own application would replace the indicated parts.

Compute the raw data. Your own application would provide the variables for the namelists *wi* and *x* and the two dependent variables *yms* for the regression and *z* for the probit model.

```
SAMPLE     ; 1-500 $
CALC       ; Ran(12345) $
CREATE     ; x1 = Rnn(0,1)                ? regressor for equation 1
           ; x0 = Rnn(0,1)                ? regressor for equation 0
           ; e1 = Rnn(0,1)                ? disturbance for equation 1
           ; e0 = .5*e1+.5*Rnn(0,1)       ? e for equation 0, correlated
           ; u = Rnn(0,1)+.5*(e1+e0)      ? u for endogenous selection
           ; w = Rnn(0,1)                 ? regressor for selection equation
           ; z  = w+u                     ? underlying regression for probit
           ; z  = z > 0                   ? binary variable for probit
           ; y1 = x1+e1                   ? structural variable, y1*
           ; y0 = x0+e0                   ? structural variable, y0*
           ; If(y1 < y0) ys = y1          ? choose minimum of y1*, y0*
           ; (Else) ys = y0
           ; yms = z*y1+(1-z)*y0          $ Lhs for mover stayer model
CALC       ; n0 = n - Sum(z) ; n1 = Sum(z)$
NAMELIST   ; wi = one,w $
NAMELIST   ; x = one,x1,x0 $
```

The remainder of the program is generic and need not be changed for a particular application. This block computes the means for the whole sample and the two subsamples.

```
CALC       ; k = Col(x) ; m = Col(wi) $
PROBIT     ; Lhs = z ; Rhs = wi ; Hold $
MATRIX     ; alpha = b $
MATRIX     ; wb = Mean(wi)
           ; wb1 = 1/n1 *wi'z ; wb0 = 1/n0 * Mdif(wi'1,wi'z) $
MATRIX     ; xb = Mean(x)
           ; xb1 = 1/n1 *x'z ; xb0 = 1/n0 * Mdif(x'1,x'z) $
```

The two regressions produce coefficient vectors and estimates of ρ_j and σ_j.

```
SELECT     ; Lhs = yms ; Rhs = x $
MATRIX     ; beta1 = bsr1 ; b1  = bsr1(1:k) $
CALC       ; r1 = rho ; s1 = s $
SELECT     ; Lhs = yms ; Rhs = x ; Limits = 1 $
MATRIX     ; beta0 = bsr0 ; b0 = bsr0(1:k) $
CALC       ; r0 = rho ; s0 = s $
```

Compute the treatment effects.

```
CALC       ; List ; TE = b1'xb1 + r1*s1*  N01(alpha'wb1)/Phi(alpha'wb1)
                        -b0'xb0 - r0*s0*(-N01(alpha'wb0)/Phi(-alpha'wb0)) $
CALC       ; List ; ATE = b1'xb - b0'xb $
CALC       ; List ; ATT = ATE + (r1*s1 - r0*s0)*N01(alpha'wb)/Phi(alpha'wb) $
```

It would be useful to have standard errors for the computed average treatment effects. In principle, this can be done using the delta method. However, there are two obstacles in the preceding. The complexity of the computations does suggest it will be tedious. However, the single equation estimates do not provide the necessary asymptotic variances for the estimates of ρ_j and σ_j, so as it stands, the computation cannot be done (at least not without treating these as constants.) However, the MLE for the mover stayer model and the **WALD** command solve both problems easily.

```
SELECT    ; Lhs = yms
          ; Rh1 = x
          ; Rh2 = x
          ; Start = beta1, beta0 ; MLE ; All $
WALD      ; Start = b
          ; Var = varb
          ; Labels = m_alpha, k_bs1, k_bs0, ss0, rs0, ss1, rs1
          ; Fn1 = bs11'xb1 + rs1*ss1*N01(alpha1'wb1)/Phi(alpha'wb1)
                  -bs01'xb0 + rs0*ss0*N01(alpha1'wb0)/Phi(-alpha'wb0)
          ; Fn2 = bs11'xb - bs01'xb
          ; Fn3 = Fn2 + (rs1*ss1-rs0*ss0)*N01(alpha1'wb)/Phi(alpha1'wb) $
```

We applied these computations to the data in the preceding example. The estimates of the parameters are the same. The computations of the treatment effects are shown below. (The estimate of the variance of the first treatment effect was too close to zero; evidently with the rounding error of the computation, that diagonal element of the matrix became negative. The third estimate, the effect of treatment on the treated is the usual object of estimation.)

```
+-------------------------------------+
| Listed Calculator Results           |
+-------------------------------------+
 TE      =       .505201
 ATE     =       .107003
 ATT     =       .121913
+------------------------------------------------+
| WALD procedure. Estimates and standard errors  |
| for nonlinear functions and joint test of      |
| nonlinear restrictions.                        |
| VC matrix for the functions is singular.       |
| Standard errors are reported, but the          |
| Wald statistic cannot be computed.             |
+------------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 Fncn(1) |     .65997144    ......(Fixed Parameter).......
 Fncn(2) |     .04118714       .12689739      .325   .7455
 Fncn(3) |     .22861000       .13082471     1.747   .0806
```

E56.3 Alternative Distribution for Selection and Treatment Effects

The treatment effects model developed above has structure

$$y_i = Y_1 = \boldsymbol{\beta}_1'\mathbf{x}_i + \varepsilon_{i1} \text{ when } z_i = 1$$
$$y_i = Y_0 = \boldsymbol{\beta}_0'\mathbf{x}_i + \varepsilon_{i0} \text{ when } z_i = 0$$
$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ z_i = 1(z_i^* > 0).$$

$$\begin{pmatrix}\varepsilon_{i1}\\ \varepsilon_{i0}\\ u_i\end{pmatrix} \sim N_3\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & 0 & \rho_1\sigma_1\\ 0 & \sigma_0^2 & \rho_0\sigma_0\\ \rho_1\sigma_1 & \rho_0\sigma_0 & 1\end{pmatrix}\right].$$

The trivariate normality of the random components is an important feature of the specification. Heckman, Tobias and Vytlacil (2001) (see also Heckman and Vytlacil (2000) and references cited) suggest that the model can be improved by allowing a different distribution with thicker tails to govern the selection mechanism. The logit model, instead of the probit is a natural candidate. In order to maintain the flexible structure of the model, they propose to link the selection model to the regressions through an inverse transformation that reproduces the joint normality of the system. Formally, using their notation, we maintain that

$u_i \sim F(u_i)$ with CDF $F(u_i)$ where $F(u_i)$ defines a symmetric distribution,

Let $J(t)$ be a strictly increasing function, such that

$$z_i = 1 \text{ when } u_i > -\boldsymbol{\alpha}'\mathbf{w}_i$$

or

$$z_i = 1 \text{ when } J(u_i) > J(-\boldsymbol{\alpha}'\mathbf{w}_i).$$

To map the model with nonnormal selection rule into the model where the normal distribution applies, they propose the mapping $\tilde{u}_i = J_\Phi(u_i) = \Phi^{-1}[F(u_i)]$. Then, $\tilde{u}_i$ has a standard normal distribution. The revised system is

$$y_i = Y_1 = \boldsymbol{\beta}_1'\mathbf{x}_i + \varepsilon_{i1} \text{ when } z_i = 1$$
$$y_i = Y_0 = \boldsymbol{\beta}_0'\mathbf{x}_i + \varepsilon_{i0} \text{ when } z_i = 0$$
$$z_i^{**} = J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_i) + \tilde{u}_i,\ z_i = 1(z_i^* > 0).$$

$$\begin{pmatrix}\varepsilon_{i1}\\ \varepsilon_{i0}\\ \tilde{u}_i\end{pmatrix} \sim N_3\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & 0 & \rho_1\sigma_1\\ 0 & \sigma_0^2 & \rho_0\sigma_0\\ \rho_1\sigma_1 & \rho_0\sigma_0 & 1\end{pmatrix}\right].$$

The implications for the conditional mean functions and treatment effects are simple;

$$E[y_i \mid \mathbf{x}_i, z_i = 1] \;=\; \boldsymbol{\beta}_1'\mathbf{x}_1 \;+\; (\rho_1\sigma_1)\left(\frac{\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_1))}{F(\boldsymbol{\alpha}'\mathbf{w}_1)}\right) \text{ if } z = 1$$

$$E[y_i \mid \mathbf{x}_i, z_i = 0] \;=\; \boldsymbol{\beta}_0'\mathbf{x}_0 \;+\; (\rho_0\sigma_0)\left(\frac{-\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_0))}{F(-\boldsymbol{\alpha}'\mathbf{w}_0)}\right) \text{ if } z = 0.$$

$$ATE \;=\; \boldsymbol{\beta}_1'\mathbf{x} - \boldsymbol{\beta}_0'\mathbf{x} \;=\; (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x}$$

$$ATT \;=\; \text{E}[\text{Y}_1|\text{z}=1] \;-\; \text{E}[\text{Y}_0|\text{z}=1]$$

$$= \boldsymbol{\beta}_1'\mathbf{x} \;+\; (\rho_1\sigma_1)\left(\frac{\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_1))}{F(\boldsymbol{\alpha}'\mathbf{w}_1)}\right) - \boldsymbol{\beta}_0'\mathbf{x} \;+\; (\rho_0\sigma_0)\left(\frac{\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_1))}{F(\boldsymbol{\alpha}'\mathbf{w}_1)}\right)$$

$$= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x} + [(\rho_1\sigma_1) \;-\; (\rho_0\sigma_0)]\left(\frac{\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}_1))}{F(\boldsymbol{\alpha}'\mathbf{w}_1)}\right)$$

$$LATE \;=\; (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{x} + [(\rho_1\sigma_1) \;-\; (\rho_0\sigma_0)]\left[\frac{\phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}^+)) - \phi(J_\Phi(\boldsymbol{\alpha}'\mathbf{w}))}{F(\boldsymbol{\alpha}'\mathbf{w}^+) - F(\boldsymbol{\alpha}'\mathbf{w})}\right]$$

The denominators of the results are all obtained from $\Phi\{\Phi^{-1}[F(u_i)]\} = F(u_i)$.

Though advocated as a ‘modern’ approach, this is precisely Lee’s (1982, 1983) model (the authors do note this) developed in Section E55.4. Thus, estimating the parameters and computing the treatment effects with this form is a small modification of what we have already done. The commands would be as follows. We have computed the treatment effects, but not the asymptotic standard errors. The Lee model is fit by two step least squares, so we do not have the asymptotic covariance matrix needed for the computation. One could possibly use bootstrapping to add this computation to the results.

The results below display only a trivial change due to the modification of the model. Of course, these are simulated data. The authors construct a Monte Carlo study which produces more pronounced differences. They also suggest that one might relax the joint normality assumption. Their proposal, a multivariate t distribution with small degrees is a bit ad hoc, but their results suggest that differences in the sizes of the tails of the distributions does induce changes in the results.

The commands are:

```
NAMELIST    ; wi = one,w $
NAMELIST    ; x = one,x1,x0 $
CALC        ; k = Col(x) ; m = Col(wi) $
LOGIT       ; Lhs = z ; Rhs = wi ; Hold $
MATRIX      ; alpha = b $
MATRIX      ; wb = Mean(wi)
            ; wb1 = 1/n1 *wi'z ; wb0 = 1/n0 * Mdif(wi'1,wi'z) $
MATRIX      ; xb = Mean(x)
            ; xb1 = 1/n1 *x'z ; xb0 = 1/n0 * Mdif(x'1,x'z) $
SELECT      ; Lhs = yms ; Rhs = x $
MATRIX      ; beta1 = bsr1 ; b1 = bsr1(1:k) $
CALC        ; r1 = rho ; s1 = s $
SELECT      ; Lhs = yms ; Rhs = x ; Limits = 1 $
MATRIX      ; beta0 = bsr0 ; b0 = bsr0(1:k) $
CALC        ; r0 = rho ; s0 = s $
CALC        ; jphi1 = Ntb(Lgp(alpha'wb1)); f1 = Lgp(alpha'wb1) $
CALC        ; jphi0 = Ntb(Lgp(alpha'wb0)); f0 = Lgp(-alpha'wb0) $
CALC        ; jphi = Ntb(Lgp(alpha'wb) ; f = Lgp(alpha'wb) $
CALC        ; List ; TE = b1'xb1 + r1*s1*N01(jphi1)/f1
                         -b0'xb0 - r0*s0*(-N01(jphi0)/f0) $
CALC        ; List ; ATE = b1'xb - b0'xb $
CALC        ; List ; ATT = ATE + (r1*s1 - r0*s0)*N01(jphi)/f $
```

These are the results from estimation with the logit selection rule. The results obtained earlier with the probit model follow.

```
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 TE      =       .497570
 ATE     =       .107562
 ATT     =       .127761

+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 TE      =       .505201
 ATE     =       .107003
 ATT     =       .121913
```

E56.4 Treatment Effects Regression – Endogenous Dummy Variable Models

This section will narrow the analysis of the preceding estimator by formally embedding the treatment effect in a single regression model. The basic structure is

$$y_i = \gamma z_i + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

where z_i is a dummy variable that once again indicates the presence ($z_i = 1$) or absence ($z_i = 0$) of some treatment. For example, y_i might be lifetime income and z_i might record attendance at an elite college. As long as z_i is exogenous, this is merely a classical regression with a dummy variable in it. The problem is the likely endogeneity of the treatment. This is formalized in this model with the familiar auxiliary probit equation

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ z_i = 1(z_i^* > 0).$$

This is an ordinary probit equation. The problem for estimation of $(\gamma,\boldsymbol{\beta})$ is the possible endogeneity of the dummy variable. This is the 'treatment effects' sample selection model examined in Section E52.3. It is also a restricted version of the mover stayer model in the previous section, in which the two regimes, rather than having separate regressions, now have the same regression simply with different constant terms. (The models are nested, so this is a testable restriction.)

The modification of the earlier sample selection model is as follows:

$$E[y_i\,|\mathbf{x}_i, z_i = 1] = \boldsymbol{\beta}'\mathbf{x}_i + \gamma + (\rho\sigma)[\phi(\boldsymbol{\alpha}'\mathbf{w}_i)/\Phi(\boldsymbol{\alpha}'\mathbf{w}_i)]$$

while

$$E[y_i\,|\mathbf{x}_i, z_i = 0] = \boldsymbol{\beta}'\mathbf{x}_i + (\rho\sigma)[-\phi(\boldsymbol{\alpha}'\mathbf{w}_i)/\Phi(-\boldsymbol{\alpha}'\mathbf{w}_i)].$$

Once again, we are interested in estimation of the 'treatment effect' in the model. Contrary to intuition, this is not γ, which is what motivates the sample selection model approach to this model (and, more generally, much of the literature.)

E56.4.1 Estimation

There are three estimators available for this model, two step, maximum likelihood and two stage least squares. (There are others, including a nonlinear least squares approach, not considered here.)

Two Step Estimation

Heckman's two step, or 'Heckit' estimation method is consistent, but not efficient:

Step 1. Use a probit model for z_i to estimate $\boldsymbol{\alpha}$. For each observation, compute

$$\lambda_i = \phi(\boldsymbol{\alpha}'\mathbf{w}_i) / \Phi(\boldsymbol{\alpha}'\mathbf{w}_i) \quad \text{when } z_i = 1 \text{ and}$$

$$\lambda_i = -\phi(\boldsymbol{\alpha}'\mathbf{w}_i) / \Phi(-\boldsymbol{\alpha}'\mathbf{w}_i) \quad \text{when } z_i = 0$$

using the probit coefficients.

Step 2. Linearly regress y_i on $\mathbf{x}_i$, z_i and λ_i to estimate $\boldsymbol{\beta}$, δ and $\theta = \rho\sigma$.

After estimation, it is necessary to adjust the standard errors and the usual least squares estimate of σ^2, which is inconsistent. This uses the same prescription used in Chapter E52 for the simpler model. The corrected asymptotic covariance matrix for the two step estimator, (**b**,*c*), is

$$\text{Asy.Var}[\mathbf{b},c] = \sigma_\varepsilon^2(\mathbf{X^{*\prime}X^*})^{-1}[\mathbf{X^{*\prime}}(\mathbf{I} - \rho^2\boldsymbol{\Delta})\mathbf{X^*} + \rho^2(\mathbf{X^{*\prime}}\boldsymbol{\Delta}\mathbf{W})\boldsymbol{\Sigma}(\mathbf{W}'\boldsymbol{\Delta}\mathbf{X^*})](\mathbf{X^{*\prime}X^*})^{-1}$$

where $\mathbf{X}_* = [\mathbf{X},\mathbf{z} : \boldsymbol{\lambda}]$,

$\delta_i = -\lambda_i(\boldsymbol{\alpha}'\mathbf{w}_i + \lambda_i)$ $(-1 \le \delta_i \le 0)$,

$\boldsymbol{\Delta} = \text{diag}[\boldsymbol{\delta}]$,

and $\boldsymbol{\Sigma}$ = asymptotic covariance matrix for the estimator of $\boldsymbol{\alpha}$.

A consistent estimator of σ^2 is $\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/n - \hat{\theta}^2\overline{\hat{\delta}}$. The remaining parameters are estimated using the least squares coefficients. The computations used in the estimation procedure are those discussed in Heckman (1979) and in Greene (1981).

To estimate this model with *LIMDEP*, it is necessary first to estimate the probit model, then request the selection model. The pair of commands is

```
PROBIT      ; Lhs = name of z
            ; Rhs = list for w
            ; Hold results $
SELECT      ; Lhs = name of y
            ; Rhs = list for x,z
            ; All $
```

For this simplest case, **; Hold ...** may be abbreviated to **; Hold**. All of the earlier discussion for the probit model applies. (See Chapter E27.) This application differs only in the fact the **; Hold** specification requests that the model definition and results be saved to be used later. Otherwise, they disappear with the next model command. The **PROBIT** command is exactly as described in Chapter E18. The selection model is completely self contained. You do not need to compute or save λ_i. Here, we must use entire sample, that is, not select out any observations. Use the specification **; All** in the **SELECTION** command, and otherwise, set it up in the usual manner. In this instance, all computations are exactly as described earlier, save those in the calculations. This is precisely the same as the application in Chapter E52 for the basic selection model, save for the addition of the dummy variable to the right hand side of the regression, and **; All** to the command.

Maximum Likelihood Estimation

The log likelihood function for this treatment effects model is the same as that for the mover stayer model, with the various equality restrictions imposed. It is also a minor modification of the log likelihood for the basic sample selection model. Thus, we maximize

$$\log L = \sum\nolimits_{z=1}\log\left[\frac{\exp\left(-(1/2)(y_i - \delta - \boldsymbol{\beta}'\mathbf{x}_i)^2/\sigma^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(\frac{\rho(y_i - \delta - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma + \boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho^2}}\right)\right] +$$

$$\sum\nolimits_{z=0}\log\left[\frac{\exp\left(-(1/2)(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2/\sigma^2\right)}{\sigma\sqrt{2\pi}}\Phi\left(-\left(\frac{\rho(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2/\sigma + \boldsymbol{\alpha}'\mathbf{w}_i}{\sqrt{1-\rho^2}}\right)\right)\right].$$

To fit the treatment effects model, just add **; MLE** to the two step estimator. The commands are

```
PROBIT    ; Lhs = variable z
          ; Lhs = variables in w
          ; Hold $
SELECT    ; Lhs = variable y
          ; Rhs = variables in x, variable z
          ; All
          ; MLE $
```

Two Stage Least Squares – Instrumental Variable Estimation

A second means of estimating the model is with two stage least squares. The problem with ordinary least squares estimates of the model based on the observed data is the correlation between z and ε. A solution to the inconsistency of OLS is to use 2SLS, using as the instrumental variable for z the predicted probabilities from the probit equation. It is not necessary to **; Hold** the results of the probit model in this case. The set of commands would be

```
NAMELIST  ; w = ... ; x = ... $
PROBIT    ; Lhs = z ; Rhs = w ; Prob = zfit $
2SLS      ; Lhs = y ; Rhs = x, z ; Inst = x,  zfit $
```

We note, there is a tendency in the literature to equate the simple replacement of z_i in the regression with the fitted probability as an 'instrumental variable' estimator. Ordinary least squares is then used to estimate the parameters. We emphasize, this is not 2SLS for this model and the replacement variable is not an instrument, it is a proxy. Whether the estimator so constructed is even consistent is debatable.

E56.4.2 Treatment Effects

Under the assumptions of the model, the 'treatment effect' would be

$$E[y \mid z = 1,x,w] - E[y \mid z=0,x,w] = \gamma + (\rho\sigma)\left[\frac{\phi(\boldsymbol{\alpha}'\mathbf{w})}{\Phi(\boldsymbol{\alpha}'\mathbf{w})} + \frac{\phi(\boldsymbol{\alpha}'\mathbf{w})}{\Phi(-\boldsymbol{\alpha}'\mathbf{w})}\right]$$

$$= \gamma + (\rho\sigma)\left[\frac{\phi(\boldsymbol{\alpha}'\mathbf{w})}{\Phi(\boldsymbol{\alpha}'\mathbf{w})[1-\Phi(\boldsymbol{\alpha}'\mathbf{w})]}\right].$$

As suggested earlier, the notable aspect is that this is not equal to γ unless ρ equals zero. The result is straightforward to compute using **CALC**, and the results of any of the estimators suggested below. If a standard error is desired, then the FIML estimator, and **WALD** would be the preferred approach. For the application that is developed below, the following commands would compute the effect and estimate its standard error.

The following does this computation for the model estimated below. The calculation is based on the maximum likelihood estimator. This is the third set of estimates given below. In the estimated model, the coefficient on *lfp*, the endogenous variable, is -153.226961. But, when the full model is accounted for in the **WALD** command below, the impact of the 'treatment' goes up substantially, to -250.110003.

```
NAMELIST   ; x = one,ha,he,hw,faminc $
CREATE     ; age = wa ; agesq = age*age $
NAMELIST   ; w = one,we,age,agesq,kl6,k618 $
PROBIT     ; Lhs = lfp ; Rhs = w ; Hold ; Prob = pfit $
SELECT     ; Lhs = hhrs ; Rhs = x,lfp ; All ; MLE $
CALC       ; kw = Col(w) ; kx = Col(x) $
MATRIX     ; wbar = Mean(w) $
WALD       ; Start = b ; Var = varb
           ; Labels = kw_alpha,kx_beta,gamma,sgma,ro
           ; Fn1 = gamma + sgma*ro*N01(alpha1'wbar) /
                   (Phi(alpha1'wbar)*Phi(-alpha1'wbar)) $
```

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic              =     40.98666
Prob. from Chi-squared[ 1] =       .00000
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
WaldFcns|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
 Fncn(1)|   -250.110***    39.06698    -6.40  .0000    -326.680  -173.540
--------+--------------------------------------------------------------------
```

E56.4.3 Application

In the following, we fit a 'treatment model' for the husband's hours, where the endogenous dummy variable is the wife's labor force participation. The following uses all three estimators. The probit estimates are the same as those obtained earlier.

```
NAMELIST   ; x = one,ha,he,hw,faminc $
NAMELIST   ; w = one,we,age,agesq,kl6,k618 $
PROBIT     ; Lhs = lfp   ; Rhs = w ; Hold ; Prob = pfit $
SELECT     ; Lhs = hhrs ; Rhs = x,lfp ; All $
2SLS       ; Lhs = hhrs ; Rhs = x,lfp ; Inst = x,pfit $
SELECT     ; Lhs = hhrs ; Rhs = x,lfp ; All ; MLE $
```

These are the two step estimators using Heckman's method.

```
+----------------------------------------------------------+
| Sample Selection Model                                   |
| Probit selection equation based on LFP                   |
| Sample is all observations.                              |
| Results of selection:                                    |
|                   Data points     Sum of weights         |
| Data set               753              753.0            |
| Selected sample        753              753.0            |
+----------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model...........................
Two step      least squares regression ............
LHS=HHRS      Mean                    =     2267.27092
              Standard deviation      =      595.56665
              Number of observs.      =            753
Model size    Parameters              =              7
              Degrees of freedom      =            746
Residuals     Sum of squares          =    .181436E+09
              Standard error of e     =      493.16533
Fit           R-squared               =         .31340
              Adjusted R-squared      =         .30788
Standard error corrected for selection   495.44230
Correlation of disturbance in regression
and Selection Criterion (Rho)     =         -.12587
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HHRS|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2350.53***    150.0000    15.67  .0000     2056.53   2644.52
      HA|   -6.42709***     2.32011    -2.77  .0056   -10.97442  -1.87976
      HE|    30.4775***     6.94286     4.39  .0000     16.8697   44.0852
      HW|   -112.030***     6.43336   -17.41  .0000    -124.639   -99.421
  FAMINC|     .03248***      .00223    14.58  .0000      .02812    .03685
     LFP|   -150.410       109.3608    -1.38  .1690    -364.753    63.934
  LAMBDA|   -62.3601       70.67071     -.88  .3776   -200.8722   76.1519
--------+--------------------------------------------------------------------
Two stage     least squares regression ............
LHS=HHRS      Mean                    =     2267.27092
              Standard deviation      =      595.56665
              Number of observs.      =            753
Model size    Parameters              =              6
              Degrees of freedom      =            747
Residuals     Sum of squares          =    .183027E+09
              Standard error of e     =      494.99127
Fit           R-squared               =         .30831
              Adjusted R-squared      =         .30368
Instrumental Variables:
ONE       HA        HE        HW        FAMINC    PFIT
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    HHRS|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2334.61***    159.8080    14.61  .0000     2021.39   2647.83
      HA|   -6.26441***     2.35599    -2.66  .0078   -10.88206  -1.64675
      HE|    31.4818***     6.78263     4.64  .0000     18.1881   44.7755
      HW|   -109.163***     7.48834   -14.58  .0000    -123.840   -94.486
  FAMINC|     .03159***      .00260    12.17  .0000      .02651    .03668
     LFP|   -159.015       111.4911    -1.43  .1538    -377.533    59.504
```

```
+-----------------------------------------------------------+
| Sample Selection Model                                    |
| Probit selection equation based on LFP                    |
| Sample is all observations.                               |
| Results of selection:                                     |
|                   Data points     Sum of weights          |
| Data set               753             753.0              |
| Selected sample        753             753.0              |
+-----------------------------------------------------------+
-----------------------------------------------------------------------------
Sample Selection Model...........................
Two step     least squares regression ...........
LHS=HHRS     Mean                    =     2267.27092
             Standard deviation      =      595.56665
             Number of observs.      =            753
Model size   Parameters              =              7
             Degrees of freedom      =            746
Residuals    Sum of squares          =    .181436E+09
             Standard error of e     =      493.16533
Fit          R-squared               =         .31340
             Adjusted R-squared      =         .30788
Standard error corrected for selection   495.44230
Correlation of disturbance in regression
and Selection Criterion (Rho)        =        -.12587
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
    HHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    2350.53***    150.0000    15.67  .0000     2056.53   2644.52
      HA|   -6.42709***     2.32011    -2.77  .0056   -10.97442  -1.87976
      HE|    30.4775***     6.94286     4.39  .0000     16.8697   44.0852
      HW|   -112.030***     6.43336   -17.41  .0000    -124.639   -99.421
  FAMINC|     .03248***      .00223    14.58  .0000      .02812    .03685
     LFP|   -150.410       109.3608    -1.38  .1690    -364.753    63.934
  LAMBDA|   -62.3601       70.67071     -.88  .3776   -200.8722   76.1519
--------+--------------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                   HHRS
Log likelihood function     -6202.52230
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
    HHRS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Selection (probit) equation for LFP
Constant|    -.23352        1.54515     -.15  .8799    -3.26195   2.79491
      WE|     .11944***      .02223     5.37  .0000      .07588    .16300
     AGE|     .00276         .07099      .04  .9690     -.13638    .14190
   AGESQ|    -.00047         .00081     -.58  .5625     -.00207    .00112
     KL6|    -.87593***      .11397    -7.69  .0000    -1.09932   -.65255
    K618|    -.05539         .04028    -1.38  .1691     -.13434    .02355
        | Corrected regression, Regime 1
Constant|    2351.32***    140.8639    16.69  .0000     2075.23   2627.41
      HA|   -6.43033***     2.19962    -2.92  .0035   -10.74150  -2.11916
      HE|    30.5281***     6.68052     4.57  .0000     17.4345   43.6217
      HW|   -112.027***     4.13167   -27.11  .0000    -120.125  -103.929
  FAMINC|     .03250***      .00153    21.25  .0000      .02950    .03549
     LFP|   -153.227       115.7198    -1.32  .1855    -380.034    73.580
   SIGMA|    495.319***    11.30461    43.82  .0000     473.162   517.475
     RHO|    -.12200         .14665     -.83  .4055     -.40944    .16543
```

E56.5 Sample Selection with Two Treatments

Consider evaluating the impact of two treatments (e.g., programs). The basic regression is

$$y_i \;=\; \boldsymbol{\beta}'\mathbf{x}_i + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \varepsilon_i.$$

Binary variables D_{1i} and D_{2i} indicate presence (=1) or absence (=0) of the two treatments. To this point, save for D_1 and D_2 in the equation (which need not be the case), this is the selection model with bivariate probit sample selection equations. To the preceding, however, we add an 'eligibility requirement,'

$$E_{1i} \;=\; 1 \text{ if individual is eligible for program 1, 0 otherwise,}$$

$$E_{2i} \;=\; 1 \text{ if individual is eligible for program 2, 0 otherwise.}$$

Then, $D_{1i} = 0$ if $E_{1i} = 0$ by definition and likewise for D_{2i}. Thus, the sample contains data on $[y_i,\mathbf{x}_i,D_{1i},D_{2i},E_{1i},E_{2i}]$ for observations $i = 1,...,N$. We suppose that $m_1 \le N$ individuals are eligible for program 1 and $m_2 \le N$ are eligible for program 2.

The selection equations defined over those eligible for the programs are:

$$D_j \;=\; \mathbf{z}_{1j}'\boldsymbol{\delta}_1 + v_{1j},\; j = 1,...,m_1,$$

$$D_k \;=\; \mathbf{z}_{2k}'\boldsymbol{\delta}_2 + v_{2k},\; k = 1,...,m_2.$$

The Ds (for participation *given eligibility* are determined by probit models in the usual way, save for the complications introduced by the eligibility requirements. The final element of the specification is a trivariate normal distribution assumed for $[\varepsilon_i,v_{1i},v_{2i}]$. The marginal distributions of v_1 and v_2 are standard normal as usual for the probit model.

Estimation proceeds along lines similar to those for the basic model, with the following changes:

Step 1. The two step estimation procedure is based on all observations in the sample, not just those for which $D_{1i} = D_{2i} = 1$. The model is estimated by using least squares in an augmented regression containing two 'λ' variables. These are computed differently, however, depending on eligibility.

Step 2. For observations with $E_{1i} = E_{2i} = 1$, the usual bivariate probit model applies and the procedure shown in Section E55.3 applies. This group will still contain observations in all four D_1/D_2 cells.

Step 3. If $E_{1i} = 1$ but $E_{2i} = 0$, λ_1 is computed based on the univariate probit model for D_{1i}, but λ_2 is taken to be 0. This group may have observations with D_{1i} equal to 1 or 0, but D_{2i} must be 0.

Step 4. For $E_{2i} = 1$ and $E_{1i} = 0$, the reverse of the procedure in Step 3 applies.

Step 5. If $E_{1i} = E_{2i} = 0$, then both D_{1i} and D_{2i} must be 0, so both λs are taken to be 0.

The necessary adjustments to the way the asymptotic covariance matrix is computed are all made internally. Note that it is possible to miscode the data. For example, the pair of values $D_{1i} = 1$, $E_{1i} = 0$, is invalid. Individuals cannot participate in programs for which they are ineligible. *LIMDEP* checks for miscoded data.

The procedure for estimation of this model must be as follows: With the exception noted, you can use any names you like.

```
CREATE     ; d1 = dependent variable in first probit
           ; d2 = dependent variable in second probit
           ; y = dependent variable in main regression $
NAMELIST   ; x = Rhs in primary equation
           ; z1 = Rhs for probit for d1
           ; z2 = Rhs for probit for d2 $
INCLUDE    ; New ; e1 = 1 | e2 = 0 $
PROBIT     ; Lhs = d1
           ; Rhs = z1 $      (No need to ; Hold)
MATRIX     ; delta1 = b $    (You must use name delta1)
INCLUDE    ; New ; e1 = 0 & e2 = 1 $
PROBIT     ; Lhs = d2
           ; Rhs = z2 $
MATRIX     ; delta2 = b $
INCLUDE    ; New ; e1 = 1 & e2 = 1 $
BIVARIATE  ; Lhs = d1, d2
           ; Rh1 = z1; Rh2 = z2
           ; Hold
           ; Start = delta1, delta2, 0 $
MATRIX     ; vdelta = varb $
INCLUDE    ; New ; e1 = 1 | e1 = 0 | e2 = 1 | e2 = 0 $ (all properly coded data)
SELECT     ; Lhs = y
           ; Rhs = x ; Rh2 = e1,e2 $
```

Output from this procedure is essentially the same as that for the selection model with bivariate probit selection. However, a complete tabulation of the numbers of observations in the various cells is given at the beginning of the results for the **SELECT** command.

E56.6 ML Estimation of a Tobit Model with an Endogenous Dummy Variable

Our interest is in a model which includes censoring of the dependent variable in the second equation. A fully operational estimator for the simultaneous equations model

$$z_i^* = \gamma_1 y_{i2}^* + \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ z_i = \mathbf{1}(z_{i1}^* > 0)$$

$$y_i^* = \gamma_2 z_i + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ y_i = \text{Max}(L_i, y_i^*) \text{ (note, } z_i\text{, not } z_{i1}^*)$$

remains to be derived. Maddala (1983) shows that the model cannot be internally consistent if γ_1 is nonzero.) This restricts us to recursive models. Consider, then, the model with $\gamma_1 = 0$

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\ z_i = \mathbf{1}(z_{i1}^* > 0)$$

$$y_i^* = \gamma_2 z_i + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ y_i = \text{Max}(L_i, y_i^*).$$

The two step estimator is problematic in this case. The partial reduced form for y_{i2}* is not available – the precise prediction that should be inserted for z_i is unclear. But, a full information maximum likelihood estimator is feasible.

We continue to assume that u and ε are bivariate normally distributed with zero means, variances one and σ^2, and correlation ρ. Consider, first, the cases in which y_i* is censored. The probabilities associated with these outcomes are the probabilities of the joint events,

$$\text{Prob}[y_i = L_i\,,\, z_i = 0] \text{ and } \text{Prob}[y_i = L_i\,,\, z_i = 1].$$

These are simply the bivariate standard normal integrals,

$$\begin{aligned}\text{Prob}[y_i = L_i, z_i = 0] &= \text{Prob}[(\varepsilon_i/\sigma) \le (L_i - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma,\ u_i \le -\boldsymbol{\alpha}'\mathbf{w}_i, \mid \rho] \\ &= \Phi_2[(L_i - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma,\ -\boldsymbol{\alpha}'\mathbf{w}_i, \rho]\end{aligned}$$

and

$$\begin{aligned}\text{Prob}[y_i = L_i, z_i = 1] &= \text{Prob}[(\varepsilon_i/\sigma) \le (L_i - \gamma - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma,\ u_i > -\boldsymbol{\alpha}'\mathbf{w}_i \mid \rho] \\ &= \text{Prob}[(\varepsilon_i/\sigma) \le (L_i - \gamma - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma,\ u_i < \boldsymbol{\alpha}'\mathbf{w}_i \mid -\rho] \\ &= \Phi_2[(L_i - \gamma - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma,\ \boldsymbol{\alpha}'\mathbf{w}_i, -\rho]\end{aligned}$$

where Φ_2 denotes the bivariate standard normal CDF. These are the terms in the likelihood function for the fully censored data. It might seem odd that this has ignored the simultaneity. However, the result can be obtained trivially by writing

$$\text{Prob}[y_i = L_i, z_i = 1] = \text{Prob}[y_i = L_i \mid z_i = 1]\text{Prob}[z_i = 1].$$

The former probability is just the joint probability divided by the marginal, which then cancels out of the product, and, of course, conditioned on z_i, we are free to treat z_i as a constant. The 'simultaneity' only becomes an issue in regression because of the use of covariances and moments. In this instance, we are using the probabilities directly.

For the uncensored observations, we require the two mixed distributions, $f(y_i, z_i = 0)$ and $f(y_i, z_i = 1)$. The first of these, $f(y_i, z_i{=}0)$, is derived from

$$\begin{aligned}f(\varepsilon_i, u_i \le -\boldsymbol{\alpha}'\mathbf{w}_i) &= \int_{-\infty}^{-\boldsymbol{\alpha}'\mathbf{w}_i} f(\varepsilon_i, u_i)\,du_i \\ &= \int_{-\infty}^{-\boldsymbol{\alpha}'\mathbf{w}_i} f(\varepsilon_i) f(u_i \mid \varepsilon_i)\,du_i \\ &= f(\varepsilon_i)\int_{-\infty}^{-\boldsymbol{\alpha}'\mathbf{w}_i} f(u_i \mid \varepsilon_i)\,du_i \\ &= \frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{\beta}'\mathbf{x}}{\sigma}\right)\int_{-\infty}^{-\boldsymbol{\alpha}'\mathbf{w}_i} \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u_i - (\rho/\sigma)\varepsilon_i}{\sqrt{1-\rho^2}}\right)du_i \\ &= \frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{\beta}'\mathbf{x}}{\sigma}\right)\int_{-\infty}^{-\boldsymbol{\alpha}'\mathbf{w}_i} \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u_i - (\rho/\sigma)(y_i - \boldsymbol{\beta}'\mathbf{x}_i)}{\sqrt{1-\rho^2}}\right)du_i \\ &= \frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{\beta}'\mathbf{x}}{\sigma}\right)\Phi\left(\frac{-\boldsymbol{\alpha}'\mathbf{w}_i - (\rho/\sigma)(y_i - \boldsymbol{\beta}'\mathbf{x}_i)}{\sqrt{1-\rho^2}}\right).\end{aligned}$$

For $f(y_i, z_i = 1)$, the numerator inside the CDF is changed to $\boldsymbol{\alpha}'\mathbf{w}_i + (\rho/\sigma)(y_i - \gamma - \boldsymbol{\beta}'\mathbf{x}_i)$ and all other parts remain the same. These four results, then, give the parts of the likelihood function, which can then be maximized to estimate the parameters.

The following program will estimate the parameters of the model by maximum likelihood. This is not built into *LIMDEP*, so we use **MAXIMIZE** instead. The left and right hand sides of the two equations are defined for the specific problem. The censoring limit for the second equation will typically be zero, but can be nonzero. That is defined here as well. The rest of the command set is generic, and can be used without modification.

```
NAMELIST    ; w = the Rhs of the probit model $
NAMELIST    ; x  = exogenous variables in the censored regression $
CREATE      ; z  = binary dependent variable $
CREATE      ; y  = censored dependent variable $
CREATE      ; li  = censoring limit $
```

Obtain the dimensions of the problem and pointers to partition the parameter vector.

```
CALC        ; k = Col(x) ; k1 = k+1 ; m = Col(w) $
```

Get the starting values for the probit model. These are consistent, but LIML so inefficient.

```
PROBIT      ; Lhs = z ; Rhs = w $
MATRIX      ; alpha 0 = b $
```

Obtain the starting values for censored regression. These are inconsistent, but better than zero.

```
TOBIT       ; Lhs = y ; Rhs = z,x ; Limit = li $
CALC        ; gamma0 = b(1) $
MATRIX      ; beta0 = b(2:k1) $
```

Compute a starting value for σ in the tobit equation, then use the Olsen transformation. The starting value for ρ is zero.

```
CALC        ; s0 = s  ; r0 = 0 ; h0 = 1/s $
CREATE      ; d = y > li ; q = 2*z - 1 $
```

Finally, compute the FIML estimation of all model parameters using maximum likelihood.

```
MAXIMIZE    ; Labels = c,k_beta,m_alpha,h,r
            ; Start = gamma0,beta0,alpha0,h0,r0
            ; Fcn = alphaw = alpha1'w                  |
                    betax = beta1'x                    |
                    a0 = (li – c*z - betax)* h         |
                    a1 = (y  - c*z - betax)* h         |
                    dr = 1/Sqr(1 - r*r)                |
                    uz = q*alphaw                      |
                    ur = -q*r                          |
              Log((1-d) * BVN(a0,uz,ur) + d * h*N01(a1) * Phi(dr*(q*alphaw + r*a1))) $
```

The following application reestimates the husband's hours equation that was estimated earlier. To force the data into the present framework, we have censored the hours variable at 1,800 hours.

```
NAMELIST   ; x = one,ha,he,hw,faminc $
NAMELIST   ; w = one,we,age,agesq,kl6,k618 $
CREATE     ; hours = hhrs ; If(hours < 1800)hours = 1800 $
CREATE     ; z = lfp ; li = 1800 ; y = hours $
CALC       ; k = Col(x) ; k1 = k+1 ; m = Col(w) $
PROBIT     ; Quietly ; Lhs = z ; Rhs = w $
MATRIX     ; alpha 0 = b $
TOBIT      ; Quietly ; Lhs = y ; Rhs = z,x ; Limit = li $
CALC       ; gamma0 = b(1) $
MATRIX     ; beta0 = b(2:k1) $
CALC       ; s0 = s ; r0 = 0 ; h0 = 1/s $
CREATE     ; d = y > li ; q = 2*z - 1 $
MAXIMIZE   (exactly as shown above)
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function      -5395.66339
Estimation based on N =     753, K =  14
Inf.Cr.AIC  =  10819.3 AIC/N =   14.368
Model estimated: Aug 10, 2011, 22:14:06
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
       C|   -97.7351*      52.14551    -1.87  .0609   -199.9384    4.4682
   BETA1|    2039.00***    116.1124    17.56  .0000     1811.42   2266.57
   BETA2|   -7.28367***     1.94937    -3.74  .0002   -11.10436  -3.46299
   BETA3|    25.1337***     5.15560     4.88  .0000     15.0289   35.2385
   BETA4|   -73.8593***     6.20867   -11.90  .0000    -86.0281  -61.6905
   BETA5|     .01853***      .00160    11.60  .0000      .01540    .02166
  ALPHA1|    -.48662        1.00183     -.49  .6272    -2.45017   1.47693
  ALPHA2|     .02912**       .01145     2.54  .0110      .00668    .05156
  ALPHA3|     .03019         .04755      .63  .5254     -.06300    .12339
  ALPHA4|    -.00047         .00056     -.85  .3966     -.00156    .00062
  ALPHA5|    -.20567***      .05693    -3.61  .0003     -.31724   -.09409
  ALPHA6|    -.03359         .02517    -1.33  .1820     -.08291    .01574
       H|     .00149***    .4428D-04   33.70  .0000      .00141    .00158
       R|     .96643***      .00982    98.42  .0000      .94718    .98568
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E56.7 Endogenous Dummy Variable in a Probit Model

A natural extension of the model examined in the previous section is one in which both variables are binary. This would be a probit model with an endogenous variable on the right hand side,

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,\; z_i = \mathbf{1}(z_{i1}^* > 0)$$

$$y_i^* = \gamma z_i + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\; y_i = 1(y_i^* > 0).$$

Estimation of this model turns out to be considerably simpler than the models that we have considered thus far. Consider the model for the probabilities of the event $y = 0/1$ and $z = 0/1$. For the (1,1) case,

$$\text{Prob}[y = 1, z = 1 \mid \mathbf{x}, \mathbf{w}] \quad \text{Prob}[y = 1 \mid z = 1, \mathbf{x}, \mathbf{w}] \times \text{Prob}[z = 1 \mid \mathbf{w}]$$

$$\Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x} + \gamma + \rho(\boldsymbol{\alpha}'\mathbf{w})}{\sqrt{1-\rho^2}}\right]\Phi\left[\boldsymbol{\alpha}'\mathbf{w}\right]$$

$$\Phi_2(\boldsymbol{\beta}'\mathbf{x} + \gamma, \boldsymbol{\alpha}'\mathbf{w}, \rho).$$

This is simply the joint probability from the bivariate probit model. The other three cells would be constructed likewise, giving

$$\text{Prob}[y_1 = 1, y_2 = 0 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1, \quad -\boldsymbol{\beta}_2'\mathbf{x}_2, -\rho)$$

$$\text{Prob}[y_1 = 0, y_2 = 1 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(-\boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_1, \boldsymbol{\beta}_2'\mathbf{x}_2, -\rho)$$

$$\text{Prob}[y_1 = 0, y_2 = 0 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi_2(-\boldsymbol{\beta}_1'\mathbf{x}_1, \quad -\boldsymbol{\beta}_2'\mathbf{x}_2, \rho)$$

This is a recursive simultaneous equations model. Surprisingly enough, it can be estimated by full information maximum likelihood *ignoring the simultaneity* in the system;

```
BIVARIATE   ; Lhs  = y,z
            ; Rh1 = x,z ; Rh2 = w $
```

An application of the result to the gender economics study is given in Greene (1998), redone below. Some extensions are presented in Greene (2011).

This model presents the same ambiguity in the conditional mean function and marginal effects that were noted earlier in Chapter E33 in the bivariate probit model. The conditional mean for *y* is

$$E[y \mid z = 1, \mathbf{x}, \mathbf{w}] = \Phi_2(\boldsymbol{\beta}'\mathbf{x} + \gamma_1, \boldsymbol{\alpha}'\mathbf{w}, \rho) / \Phi(\boldsymbol{\alpha}'\mathbf{w})$$

for which derivatives were given earlier. Given the form of this result, we can identify direct and indirect effects in the conditional mean:

$$\frac{\partial E[y \mid z = 1, \mathbf{x}, \mathbf{w}]}{\partial \mathbf{x}} = \frac{g_1}{\Phi(\boldsymbol{\alpha}'\mathbf{w})}\boldsymbol{\beta} \;=\text{ direct effects}$$

$$\frac{\partial E[y \mid z = 1, \mathbf{x}, \mathbf{w}]}{\partial \mathbf{w}} = \left[\frac{g_2}{\Phi(\boldsymbol{\alpha}'\mathbf{w})} - \frac{\Phi_2(\boldsymbol{\beta}'\mathbf{x}, \boldsymbol{\alpha}'\mathbf{w}, \rho)\phi(z_2)}{\left[\Phi(\boldsymbol{\alpha}'\mathbf{w})\right]^2}\right]\boldsymbol{\alpha} \;=\text{ indirect effects}$$

The total effect for any variable which appears in both **x** and **w** would be the sum of the two effects above. The unconditional mean function is

$$
\begin{aligned}
E[y \mid \mathbf{x}, \mathbf{w}] &= \Phi(\boldsymbol{\alpha}'\mathbf{w})\, E[y \mid z = 1, \mathbf{x}, \mathbf{w}] + [1-\Phi(\boldsymbol{\alpha}'\mathbf{w})]\, E[y \mid z = 0, \mathbf{x}, \mathbf{w}] \\
&= \Phi_2\, (\boldsymbol{\beta}'\mathbf{x} + \gamma_1, \boldsymbol{\alpha}'\mathbf{w}, \rho) + \Phi_2\, (\boldsymbol{\beta}'\mathbf{x}, -\boldsymbol{\alpha}'\mathbf{w}, -\rho).
\end{aligned}
$$

Derivatives for partial effects can be derived using the results given earlier. Analysis appears in Greene (1998).

To illustrate the estimator, we examine the model estimated in Burnett (1997) and revisited in Greene (1998). The study examines the likelihood that an economics department at a liberal arts college will offer a gender economics course ($y = 1$). The endogenous dummy variable is whether there is a women's studies program offered on the campus ($z = 1$). There are 132 observations in the data set. The variables in the data set are

gndrecon	=	$y = 1$ if a gender economics course is offered, 0 if not
womstud	=	$z = 1$ if there is a women's studies program, 0 if not
acrep	=	a measure of the academic reputation of the school, a ranking
econfac	=	size of the economics faculty
pctwecn	=	percentage of the economics faculty that are women
pctwfac	=	percentage of the faculty that are women
relig	=	1 if the school has a religious affiliation, 0 if not
sou	=	1 if the school is located in the south, 0 if not
nor	=	1 if the school is located in the north, 0 if not
mid	=	1 if the school is located in the middle of the country, 0 if not
west	=	1 if the school is located in the west.

The bivariate probit model described above is estimated in Greene (1998) and examined further in Greene (2011). The lists of variables are

x	=	*constant*, *acrep*, *econfac*, *pctwecn*, *relig*
w	=	*acrep*, *pctwfac*, *relig*, *sou*, *west*, *nor*, *mid*.

The commands are as follows. The second estimator constrains ρ to equal zero.

```
NAMELIST    ; gendrecn = one,acrep,womstud,econfac,pctwecn,relig $
NAMELIST    ; womnstud = acrep,pctwfac,relig,sou,west,nor,mid $
BIVARIATE   ; Lhs = gndrecon,womstud
            ; Rh1 = gendrecn ; Rh2 = womnstud ; Partial Effects $
CALC        ; kg = Col(gendrecn) ; kw = Col(womnstud) $
BIVARIATE   ; Lhs = gndrecon,womstud ; Rh1 = gendrecn ; Rh2 = womnstud
            ; Rst  = kg_bg, kw_bw, 0 $
```

```
-----------------------------------------------------------------------------
FIML - Recursive Bivariate Probit Model
Dependent variable               WOMGND
Log likelihood function       -85.63172
Estimation based on N =    132, K =  14
Inf.Cr.AIC  =    199.3 AIC/N =    1.510
--------+--------------------------------------------------------------------
 WOMSTUD|                 Standard            Prob.      95% Confidence
GNDRECON|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index    equation for WOMSTUD
   ACREP|    -.01939***      .00570    -3.40  .0007     -.03057   -.00821
 PCTWFAC|    1.89144**       .87140     2.17  .0300      .18354   3.59935
   RELIG|    -.45838         .34033    -1.35  .1780    -1.12541    .20864
     SOU|    1.34706*        .68968     1.95  .0508     -.00469   2.69881
    WEST|    2.33757***      .86108     2.71  .0066      .64989   4.02525
     NOR|    1.90088**       .84946     2.24  .0252      .23597   3.56579
     MID|    1.80703**       .89525     2.02  .0435      .05237   3.56169
        |Index    equation for GNDRECON
Constant|   -1.19114        2.21546     -.54  .5908    -5.53336   3.15109
   ACREP|    -.01233         .00794    -1.55  .1203     -.02789    .00323
 ECONFAC|     .06769         .06952      .97  .3303     -.06858    .20395
 PCTWECN|    2.56355**      1.01441     2.53  .0115      .57536   4.55175
   RELIG|    -.37410         .52644     -.71  .4773    -1.40591    .65771
 WOMSTUD|     .88349        2.26034      .39  .6959    -3.54668   5.31367
        |Disturbance correlation
RHO(1,2)|     .13594        1.25392      .11  .9137    -2.32170   2.59358
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

---------------------------------------------------------------
Decomposition of Partial Effects for Recursive Bivariate Probit
Model is  WOMSTUD = F(x1b1), GNDRECON = F(x2b2+c*WOMSTUD )
Conditional mean function is E[GNDRECON|x1,x2] =
            Phi2(x1b1,x2b2+gamma,rho) + Phi2(-x1b1,x2b2,-rho)
Partial effects for continuous variables are derivatives.
Partial effects for dummy variables (*) are first differences.
Direct effect is wrt x2, indirect is wrt x1, total is the sum.
---------------------------------------------------------------
Variable   Direct Effect   Indirect Effect    Total Effect
---------+---------------+-----------------+-------------------
   ACREP |  -.0017329       -.0005207          -.0022536
 ECONFAC |   .0095116        .0000000           .0095116
 PCTWECN |   .3602429        .0000000           .3602429
   RELIG*|  -.0716051       -.0716051          -.1432101
 PCTWFAC |   .0000000        .0508014           .0508014
     SOU*|   .0000000        .0266914           .0266914
    WEST*|   .0000000        .0420631           .0420631
     NOR*|   .0000000        .0580612           .0580612
     MID*|   .0000000        .0382104           .0382104
---------+-----------------------------------------------------

-----------------------------------------------------------------------------
FIML - Recursive Bivariate Probit Model
Dependent variable               WOMGND
Log likelihood function       -85.64578
Estimation based on N =    132, K =  13
Inf.Cr.AIC  =    197.3 AIC/N =    1.495
Model estimated: Aug 10, 2011, 22:21:01
```

```
--------+-------------------------------------------------------------
 WOMSTUD|                    Standard            Prob.       95% Confidence
GNDRECON|   Coefficient        Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Index    equation for WOMSTUD
   ACREP|    -.01957***       .00552    -3.54  .0004     -.03039   -.00874
  PCTWFAC|    1.94293**        .84350     2.30  .0213      .28971   3.59615
   RELIG|    -.44937          .33313    -1.35  .1774    -1.10230    .20355
     SOU|    1.35969**        .65941     2.06  .0392      .06727   2.65211
    WEST|    2.33865***       .81044     2.89  .0039      .75021   3.92708
     NOR|    1.88670**        .82040     2.30  .0215      .27874   3.49465
     MID|    1.82481**        .87231     2.09  .0364      .11510   3.53451
        |Index    equation for GNDRECON
Constant|   -1.41763*         .80692    -1.76  .0789    -2.99917    .16391
   ACREP|    -.01143***       .00408    -2.80  .0051     -.01943   -.00344
 ECONFAC|     .06730          .06874      .98  .3275     -.06742    .20202
 PCTWECN|    2.53916**        .98691     2.57  .0101      .60486   4.47347
   RELIG|    -.34825          .49842     -.70  .4847    -1.32513    .62864
 WOMSTUD|    1.10951*         .56742     1.96  .0505     -.00260   2.22163
        |Disturbance correlation
RHO(1,2)|        0.0     .....(Fixed Parameter).....
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
----------------------------------------------------------------------
```

E56.8 Switching Regressions

We consider variants of the following model:

Latent structure, two regimes: $y_{1i} = \boldsymbol{\beta}_1'\mathbf{x}_{1i} + \varepsilon_{1i}, \varepsilon_{1i} \sim N[0,\sigma_{11}]$,

$y_{0i} = \boldsymbol{\beta}_0'\mathbf{x}_{0i} + \varepsilon_{0i}, \varepsilon_{0i} \sim N[0,\sigma_{00}]$,

$\text{Corr}[\varepsilon_{1i}, \varepsilon_{0i}] = \rho_{10}$ (may be assumed to equal zero).

Observation mechanism: $y_i = \min(y_{0i}, y_{1i})$ or $y_i = \max(y_{0i}, y_{1i})$.

Observed data: y_i, $\mathbf{x}_{1i}$, and $\mathbf{x}_{0i}$, $i = 1,...,N$.

The observed quantity traded in a disequilibrium model of supply and demand is an example. Maddala (1983) contains extensive discussion. The model with 'exogenous switching' (Maddala's terminology) holds that the separation into one regime or the other is determined outside the structure of the model. Two cases are possible:

Observed separation indicator: There exists an observed indicator variable, z_i, which equals 1 if regime 1 applies and 0 if regime 0 applies. Continuing the earlier example, one suggestion for an indicator has been Sgn(P_t - $P_{t\text{-}1}$). I.e., whether price is rising or falling indicates whether the market is in shortage (Q on supply equation) or surplus (Q on demand equation).

No separation indicator: It is not known which regime applies for a given observation. This case produces a variant of the latent class model.

All combinations of the preceding are available for the basic model, i.e.:

- Minimum or maximum observation mechanism,
- Correlated or uncorrelated disturbances,
- Observed sample separation or none observed.

A model with ‘endogenous switching’ would have an auxiliary equation for the separation indicator,

$$z_i^* = \boldsymbol{\alpha}'\mathbf{w}_i + u_i,$$

$$z_i = 1 \text{ if } z_i^* > 0, \text{ and } 0 \text{ otherwise},$$

$$\text{Corr}[u_i, \varepsilon_{1i}] = \rho_{u1},$$

$$\text{Corr}[u_i, \varepsilon_{0i}] = \rho_{u0}.$$

This is the mover stayer model presented above in Section E56.2. This section is concerned with the model with exogenous, observed switching, or with an unobserved switching indicator.

E56.8.1 Model Commands

Commands for the switching regressions models are as follows:

```
SWITCH    ; Lhs  = y
          ; Rh1 = x1
          ; Rh2 = x2 $
```

This requests the basic model with

- no sample separation indicator,
- uncorrelated disturbances, and
- $y = \text{Min}(y_1, y_0)$.

To request the alternative specifications, use

; Sep = name of z,	sample separation binary variable
; Cor	to request the model with ρ_{10} not fixed at 0
; Max	to use the alternative observation mechanism
; Wts = weighting variable	

The parameter vector for **SWITCH** is

$$\boldsymbol{\theta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \sigma_1, \sigma_0].$$

The restrictions in **; Rst = list** may be used for within and across equations. If the model is estimated with correlation, ρ_{10} will precede σ_1 in the parameter vector.

Standard Model Specifications for the Switching Regressions Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps ancillary parameters such as a correlation in main results vector *b*.
; Partial Effects	displays marginal effects, same as **; Marginal Effects**.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Cluster = spec	requests computation of the cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm. The only fitting algorithm available is Newton’s method.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level ‘**n**’ is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E56.8.2 Results for Switching Regressions Models

Initial output consists of least squares regressions for both equations. If there is no sample separation indicator, the full sample is used in both regressions. If a sample separation indicator is available, the relevant subsample is used in each regression. Either way, the OLS estimates are inconsistent, and are used only as starting values. The starting value for ρ, when the model with correlation is requested, is 0.0.

The iterations are followed by the maximum likelihood estimates. Output includes the log likelihood, an indication of whether the observation mechanism is '**; Minimum**' or '**; Maximum**,' and, if one is present, the identification of the sample separation indicator. The final estimates include, in order, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_0$, σ_1, and σ_0.

Results saved by this procedure are

Matrices: $b = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_0]$ and *varb*. **; Par** adds (σ_1, σ_0) to the parameter vector.

Scalars: *sy*, *ybar*, *kreg* = $k_1 + k_0$, *nreg* = N, *logl*, *sigma*1, *sigma*0.
If the correlation model is specified, *rho* contains the estimate of D.

Last Model: *b*1_*variables*, *b*0_*variables*, *r*10, *sigma*1, *sigma*0.

Last Function: None

Predicted values are computed as follows: If there is a separation indicator, z, available, then

$$\hat{y} = \boldsymbol{\beta}_1'\mathbf{x}_1 \text{ if } z = 1, \text{ and } \hat{y} = \boldsymbol{\beta}_0'\mathbf{x}_0 \text{ if } z = 0.$$

If there is no separation indicator, the prediction is

$$\hat{y} = \text{Prob}[y = y_1^*]\boldsymbol{\beta}_1'\mathbf{x}_1 + \text{Prob}[y = y_0^*]\boldsymbol{\beta}_0'\mathbf{x}_0.$$

The probability is computed using

$$\text{Prob}[y = y_1^*] = \text{Prob}[\boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1 < \boldsymbol{\beta}_0'\mathbf{x}_0 + \varepsilon_0]$$

$$= \text{Prob}[\varepsilon_1 - \varepsilon_0 < \boldsymbol{\beta}_0'\mathbf{x}_0 - \boldsymbol{\beta}_1'\mathbf{x}_1].$$

Let $$\sigma = [\sigma_1^2 + \sigma_2^2 - 2\rho_{10}\sigma_1\sigma_0]^{1/2}$$

Then, $$\text{Prob}[y = y_1^*] = \Phi[(\boldsymbol{\beta}_0'\mathbf{x}_0 - \boldsymbol{\beta}_1'\mathbf{x}_1)/\sigma].$$

If the observation is the maximum instead of the minimum, the sign of the argument in the normal CDF is reversed. The other two variables shown when **; List** is requested are $\boldsymbol{\beta}_1'\mathbf{x}_1$ and $\boldsymbol{\beta}_0'\mathbf{x}_0$.

E56.8.3 Application

The following will simulate the conditions of the switching regressions model in order to demonstrate the output that results. The commands were executed all at once from the editor.

```
SAMPLE    ; 1-500 $
CALC      ; Ran(12345) $
CREATE    ; x1 = Rnn(0,1)                 ? regressor for equation 1
          ; x0 = Rnn(0,1)                 ? regressor for equation 0
          ; e1 = Rnn(0,1)                 ? disturbance for equation 1
          ; e0 = .5*e1+.5*Rnn(0,1)        ? u for equation 0, correlated
          ; u = Rnn(0,1)+.5*(e1+e0)       ? u for endogenous selection
          ; w = Rnn(0,1)                  ? regressor for selection equation
          ; z = w+u                       ? underlying regression for probit
          ; z = z > 0                     ? binary variable for probit
          ; y1 = x1+e1                    ? structural variable, y1*
          ; y0 = x0+e0                    ? structural variable, y0*
          ; If(y1<y0) ys = y1             ? choose minimum of y1*, y0*
          ; (Else) ys = y0 $
```

We fit four variants of switching regressions model:

1. Uncorrelated disturbances, no separation indicator. Least squares are displayed.

   SWITCH ; Lhs = ys ; Rh1 = one,x1 ; Rh2 = one,x0 ; OLS $

2. Correlated equation, no separation indicator.

   SWITCH ; Lhs = ys ; Rh1 = one,x1 ; Rh2 = one,x0 ; Cor $

3. Uncorrelated disturbances, observed separation indicator.

   SWITCH ; Lhs = ys ; Rh1 = one,x1 ; Rh2 = one,x0 ; Sep = z $

4. Correlated disturbances, observed separation indicator.

   SWITCH ; Lhs = ys ; Rh1 = one,x1 ; Rh2 = one,x0 ; Sep = z ; Cor $

Results are shown for the fourth model.

```
-----------------------------------------------------------------------------
Switching Regressions
Dependent variable                     YS
Log likelihood function     -1037.65076
Estimation based on N =    500, K =   7
Inf.Cr.AIC  =   2089.3 AIC/N =    4.179
Sample separation variable is Z
YS       = the minimum of y*_1 and y*_0
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
      YS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS for Regime 1
Constant|      .26815***      .10196     2.63  .0085       .06832    .46798
      X1|      .58947***      .08506     6.93  .0000       .42275    .75619
        |RHS for Regime 2
Constant|      .50201***      .12090     4.15  .0000       .26505    .73897
      X0|      .68764***      .07816     8.80  .0000       .53445    .84084
     Rho|     -.61372***      .11099    -5.53  .0000      -.83125   -.39618
Sigma(1)|     1.58227***      .11322    13.98  .0000      1.36036   1.80417
Sigma(0)|     1.26835***      .07055    17.98  .0000      1.13008   1.40662
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E56.8.4 Technical Details

Technical results for the switching regressions model with exogenous switching appear in Maddala (1983), though most of his results pertain to the uncorrelated case. For the general case here, we have:

$$\text{parameters} = \sigma_1, \sigma_0, \rho,$$

$$\delta = 1 / (1 - \rho^2)^{1/2},$$

$$v_1 = \varepsilon_1 / \sigma_1, \qquad v_0 = \varepsilon_0 / \sigma_0,$$

$$f_j = (1/\sigma_j)\phi(v_j), \quad j = 0,1,$$

$$u_1 = \rho v_1 - v_0, \qquad u_0 = \rho v_0 - v_1,$$

$$P_1 = \Phi[\delta v_0], \qquad P_0 = \Phi[\delta v_1],$$

$$P_{01} = P_0 f_1, \qquad P_{10} = P_1 f_0.$$

With no sample separation,

$$\log L = \Sigma_i \log(P_{10} + P_{01}).$$

With sample separation,

$$\log L = \Sigma_i z_i \log P_{10} + (1 - z_i)\log P_{01}.$$

In both cases, the BHHH estimator is used to estimate the asymptotic covariance matrix of the parameter estimates. The BFGS algorithm is used by default.

NOTE: The nonconvergence problems and possible unboundedness of the log likelihood function in the case in which the sample separation is unknown have been widely documented.

E56.8.5 MLE for the Endogenous Switching Model

As noted earlier, this is the mover stayer model in Section E56.2. We provide some further details and a command language estimator for this model. The log likelihood can be reparameterized to equal

$$L_i = \theta_0\phi(\varepsilon_i)\Phi\left[\frac{-\mathbf{z}_i'\boldsymbol{\alpha} - \rho_0\varepsilon_{0i}}{\sqrt{1-\rho_0^2}}\right] + \theta_1\phi(\varepsilon_{1i})\Phi\left[\frac{\mathbf{z}_i'\boldsymbol{\alpha} + \rho_1\varepsilon_{1i}}{\sqrt{1-\rho_1^2}}\right]$$

where

$$\begin{aligned}
&\theta_0 = 1/\sigma_0,\ \rho_0 = \sigma_{u0}/\sigma_0,\\
&\theta_1 = 1/\sigma_1,\ \rho_1 = \sigma_{u1}/\sigma_1,\\
&\varepsilon_{0i} = \theta_0 y_i - \mathbf{x}_{i0}'\boldsymbol{\lambda}_0,\ \ \boldsymbol{\lambda}_0 = (1/\sigma_0)\boldsymbol{\beta}_0\\
&\varepsilon_{1i} = \theta_1 y_i - \mathbf{x}_{i1}'\boldsymbol{\lambda}_1,\ \ \boldsymbol{\lambda}_1 = (1/\sigma_1)\boldsymbol{\beta}_1
\end{aligned}$$

The log likelihood function is Σ_i log(L_i). The formulation suggested here uses the Olsen transformation of the model parameters. A command file that estimates the model parameters is

```
? This part is specific to the application
NAMELIST    ; x0 = ... $
NAMELIST    ; x1 = ... $
NAMELIST    ; z = ... $
CREATE      ; y = the dependent variable $
? Commands from here on are generic.
REGRESS     ; Lhs = y ; Rhs = x0 $
CALC        ; theta00 = 1/s ; k0 = Col(x0) $
MATRIX      ; lambda00 = theta00 * b $
REGRESS     ; Lhs = y ; Rhs = x1 $
CALC        ; theta10 = 1/s ; k1 = Col(x1) $
MATRIX      ; lambda10 = theta10 * b $
CALC        ; kz = Col(z) $
MATRIX      ; alpha0 = Init(kz,1,0) $
MAXIMIZE    ; Labels = k0_lmda0,theta0,k1_lmda1,theta1,q0,q1,kz_a
            ; Start = lambda00,theta00,lambda10,theta10,0,0,alpha0
            ; Fcn = r0   = -(Exp(q0)-1)/(Exp(q0)+1)              |
                    dr0 = 1/Sqr(1-r0*r0)                         |
                    r1   = -(Exp(q1)-1)/(Exp(q1)+1)              |
                    dr0 = 1/Sqr(1-r2*r2)                         |
                    e0   = theta0*y - lmda01'x0                  |
                    e1   = theta1*y - lmda11'x1                  |
                    za   = a1'z                                  |
                    li   = theta0*N01(e0)*Phi(-dr0*(za+r0*e0)) +
                           theta1*N01(e1)*Phi( dr1*(za+r1*e1))   |
                    Log(li) $
```

E57: Propensity Score Matching

E57.1 Introduction

Propensity score matching is the least parametric approach provided for examining treatment effects. This procedure is targeted essentially at measuring the change in the average value of y pre- and post- treatment. This program is used for estimating average treatment effects by matching observations based on propensity scores. Let O denote the outcome variable and T denote the treatment dummy variable, such that for an observation which has experienced the 'treatment,' $T = 1$, and $T = 0$ if not. We are interested in the effect of treatment *on the treated*. In principle, this means observing the treated individual before and after treatment. The problem, of course, is that ex post, we don't observe the counterfactual outcome variable, O, for the treated, in the absence of the treatment. If assignment to the treatment is nonrandom, estimation of treatment effects is biased by the effect of the variables that effect the treatment assignment. The strategy is to locate an untreated individual who looks like the treated one in every respect except the treatment, then compare the outcomes. We then average this across individual pairs to estimate the 'average treatment effect on the treated.'

E57.2 Methodology

Let $\mathbf{x}$ denote the vector of characteristics of the individual, before the treatment. Let the probability of treatment be denoted $P(T{=}1|\mathbf{x}) = \mathrm{P}(\mathbf{x})$. Since T is binary, $P(\mathbf{x}) = E[T|\mathbf{x}]$. If treatment is random *given* $\mathbf{x}$, then treatment is random given $P(\mathbf{x})$, which in this context is called the *propensity score*. It will generally not be possible to match individuals based on all the characteristics individually – with continuously measured characteristics, such as income, there are too many cells. The matching is done via the propensity score. Individuals with similar propensity scores are expected (on average) to be individuals with similar characteristics. Overall, the strategy is, for a 'treated' individual with propensity $P(\mathbf{x}_i)$ and outcome O_i, we locate a control observation with similar propensity $P(\mathbf{x}_c)$ and with outcome O_c. The effect of treatment on the treated for this individual is estimated by O_i - O_c. This is averaged across individuals to estimate the average treatment effect on the treated. The underlying theory asserts that the estimates of treatment effects across treated and controls are unbiased if the treatment assignment is random among individuals with the same propensity score – the propensity score, itself, captures the drivers of the treatment assignment. (Relevant papers that establish this methodology are too numerous to list here. Useful references are three canonical papers, Heckman et al. (1997, 1998a, 1998b) and a study by Becker and Ichino (2002).)

The steps in the propensity matching analysis consist of the following: (Steps 2 and 3 are tests of the 'balancing hypothesis.')

Step 1. Estimate the propensity score function, $P(\mathbf{x})$, for each individual by fitting a probit or logit model, and using the fitted probabilities.

Step 2. Establish that the average propensity scores of treatment and controls are the same within particular ranges of the propensity scores.

Step 3. Establish that the averages of the characteristics for treatment and controls are the same for observations in specific ranges of the propensity score.

Step 4. For each treated observation in the sample, locate similar control observation(s) based on the propensity scores. Compute the treatment effect, O_i - O_c. Average this across observations to get the average treatment effect.

Step 5. In order to estimate a standard error for this estimate, Step 4 is repeated with a set of bootstrapped samples.

E57.3 Commands for Matching

The commands used for propensity score matching are similar to those for the sample selection estimator. First, you must set the sample to include observations to be used to estimate the propensity score function. These use any set of observations you wish – they need not be the same ones subject to the propensity score analysis.

The commands to request the propensity score matching program are

PROBIT ; Lhs = treatment dummy variable
or LOGIT ; Rhs = covariates ; Hold $
MATCH ; Lhs = outcome variable $

The sample may be changed in any way desired after the **PROBIT/LOGIT** command. The sample is set to be those observations containing the treated individuals and control individuals to be used in the analysis. No other specifications are mandatory in the **MATCH** command. The current sample and the previously fit propensity score function are used. Some optional specifications are:

; Nbt = number of bootstraps (default = 25)

NOTE: To replicate an earlier set of bootstrap results, use **CALC ; Ran(seed) $** to set the seed for the random number generator to a specific value immediately before the **MATCH** command.

; List to request detailed output during analysis

; Common Support to use observations in common support of propensity scores for treated and controls. (See Section E57.6.)

Matching on the single nearest neighbor is the default. The other methods are specified with either a kernel weighting function or a ‘caliper’ to define a range of neighbors. Use

; Kernel to use kernel weights to create the neighbor. Epanechnikov is the default, with bandwidth of 0.06

or **; Range = value** to use a caliper approach. Values .001 to .50 are allowed. If the value given is positive, then the range is +/- value. If the value given is negative, then the range is +/- a proportion of the propensity scores. For example, -5.0 means +/- 5% of P_{max} - $P_{min.}$

Options for the kernel estimator are

; Normal (with **; Kernel**) to specify the standard normal kernel function
; Logit (with **; Kernel**) to specify the logistic kernel function
; Smooth = bandwidth (default = .06, .001 to .25 allowed)

E57.4 Retained Results

In addition to the numerical displays shown in the examples below, this routine keeps the following results in your project. (All are defaults; there are no options.)

Scalars: *nused* = number of observations analyzed.

Beginning from the original current sample, *nused* is the number of observations that remain after observations with missing values for any *x* variables, *O*, or *T* are eliminated and, if the common support option is requested, after observations which fall outside the common support are eliminated from the sample.

ntreated = number of observations among *nused* with $T = 1$.
ncontrol = number of controls = *nused* – *ntreated.*
trt_efct = estimated average treatment effect.
sd_trtmt = estimated standard error for estimated effect.

The standard deviation is the square root of the mean squared deviation of the bootstrap estimates around the estimated treatment effect (not around the mean of the bootstrap estimates).

Variable: *ps_range* = number of the interval in the partitioned range of propensity scores that each observation falls in.

This is the identity of the interval in the mesh [P*] determined in Section E57.6 in the mathematical details below. For example, if the mesh is [.1, .2, .3, .4, .5, .6] and a score is 0.34, this variable would take value 3 for this observation.

Matrix: *psranges* = partitioning of the range of propensity scores used to analyze the balancing hypothesis.

In the example immediately above, this matrix would be the column vector,

psranges = [.1, .2, .3, .4, .5, .6]′

E57.5 Applications

In the following, we redo the example reported in Becker and Ichino (2002) (BI). This application and data are derived from Dehejia and Wahba (1999) (DW), whose study, in turn was based on LaLonde (1986). The data set consists of observed samples of treatments and controls from the *National Supported Work* demonstration. Some of the institutional features of the data set are given by Becker and Ichino. The data were downloaded from Rajeev Dehejia's data page http://www.nber.org/~rdehejia/nswdata.html. Becker and Ichino report that they were unable to replicate DW's results, though they did obtain similar results. (They indicate that they did not have the original authors' specifications of the number of blocks used in the partitioning of the range of propensity scores, significance levels, or exact procedures for testing the balancing property.) In turn, we could not replicate BI's results – we can identify the reason, as discussed below. Likewise, however, we obtain qualitatively similar results.

There are 2,675 observations in the data set. The variables in the data set are

t	=	treatment dummy variable
age	=	age in years
educ	=	education in years
marr	=	dummy variable for married
black	=	dummy variable for black
hisp	=	dummy variable for Hispanic
nodegree	=	dummy for no degree (not used)
*re*74	=	real earnings in 1974
*re*75	=	real earnings in 1975
*re*78	=	real earnings in 1978

Transformed variables added to the equation are

*age*2	=	*age* squared
*educ*2	=	*educ* squared
*re*742	=	*re*74 squared
*re*752	=	*re*75 squared
*blacku*74	=	*black* times 1(*re*74 = 0)

In order to improve the readability of some of the reported results, we have divided the income variables by 10,000. The outcome variable is *re*78. The sample contains, in total, 2490 controls and 185 treated observations.

The data set is setup and described first.

```
CREATE   ; age2 = age^2 ; educ2 = educ^2 $
CREATE   ; re74 = re74/10000 ; re75 = re75/10000 ; re78 = re78/10000 $
CREATE   ; re742 = re74^2 ; re752 = re75^2 $
CREATE   ; blacku74 = black * (re74 = 0) $
DSTAT    ; Rhs = * $
```

```
Descriptive Statistics
--------+-----------------------------------------------------------------
Variable|       Mean       Std.Dev.      Minimum      Maximum     Cases Missing
--------+-----------------------------------------------------------------
       T|     .069159      .253772          0.0          1.0      2675       0
     AGE|    34.22579     10.49984         17.0         55.0      2675       0
    EDUC|    11.99439     3.053556          0.0         17.0      2675       0
    MARR|     .819439      .384726          0.0          1.0      2675       0
   BLACK|     .291589      .454579          0.0          1.0      2675       0
    HISP|     .034393      .182269          0.0          1.0      2675       0
NODEGREE|     .333084      .471404          0.0          1.0      2675       0
    RE74|    1.823000     1.372225          0.0     13.71490      2675       0
    RE75|    1.785089     1.387778          0.0     15.66530      2675       0
    RE78|    2.050238     1.563252          0.0     12.11740      2675       0
    AGE2|    1281.610     766.8415        289.0       3025.0      2675       0
   EDUC2|    153.1862     70.62231          0.0        289.0      2675       0
   RE742|    5.205628     8.465891          0.0     188.0980      2675       0
   RE752|    4.883459     8.250059          0.0     214.8480      2675       0
  BLKU74|     .055327      .228660          0.0          1.0      2675       0
--------+-----------------------------------------------------------------
```

We next fit the logit model for the propensity scores. An immediate problem arises with the data set as used by Becker and Ichino. The income data are in raw dollar terms – the mean of *re*74, for example is $13,714.86. The square of it, which is on the order of 200,000,000, as well as the square of *re*75 which is similar, is included in the logit equation with a dummy variable for Hispanic which is zero for 96.5% of the observations and the *blacku*74 dummy variable which is similar. This data set is numerically unstable, and estimation of the logit model in this form is next to impossible. It was not possible to replicate the (*Stata* generated) coefficients without scaling the data. Thus, we have divided the income variables by 10,000 before beginning the analysis. From this point forward, none of their reported results can be reproduced. However, as noted at various points below, our results are quite similar to theirs in spite of this. (Comparable values appear in parentheses at some points below.)

```
NAMELIST    ; x = age,age2,educ,educ2,marr,black,hisp,
                  re74,re75,re742,re752,blacku74,one $
LOGIT       ; Lhs = t ; Rhs = x ; Hold ; Summary $
```

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                        T
Log likelihood function      -205.12591
Restricted log likelihood    -672.64954
Chi squared [  12 d.f.]       935.04727
Significance level               .00000
McFadden Pseudo R-squared      .6950479
Estimation based on N =   2675, K =  13
Inf.Cr.AIC  =    436.3 AIC/N =     .163
Model estimated: Aug 10, 2011, 23:12:24
Hosmer-Lemeshow chi-squared =  12.90811
P-value=  .11505 with deg.fr. =       8
-----------------------------------------------------------------------------

--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       T|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[T=1     ]
     AGE|     .32965***     .12043     2.74  .0062      .09361     .56569
    AGE2|    -.00633***     .00186    -3.41  .0007    -.00997    -.00269
    EDUC|     .88403***     .34150     2.59  .0096      .21471    1.55335
   EDUC2|    -.05215***     .01702    -3.06  .0022    -.08552    -.01878
    MARR|   -1.89160***     .29919    -6.32  .0000   -2.47801   -1.30519
   BLACK|    1.13696***     .35195     3.23  .0012      .44715    1.82677
    HISP|    1.96830***     .56695     3.47  .0005      .85709    3.07951
    RE74|   -1.04742***     .35896    -2.92  .0035   -1.75097    -.34386
    RE75|   -2.18585***     .41827    -5.23  .0000   -3.00564   -1.36607
   RE742|     .23048***     .08231     2.80  .0051      .06916     .39180
   RE752|     .02516        .08841      .28  .7759    -.14811      .19844
BLACKU74|    2.13433***     .42694     5.00  .0000    1.29753    2.97112
Constant|   -7.63663***    2.42743    -3.15  .0017  -12.39431   -2.87895
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

(Note: BI coefficients on *re*74 and *re*75 are multiplied by 10,000, and coefficients on *re*742 and *re*752 are multiplied by 100,000,000. Some additional logit results are omitted.)

Fit Measures for Binomial Choice Model
Logit model for variable T

	Y=0	Y=1	Total
Proportions	.93084	.06916	1.00000
Sample Size	2490	185	2675

Log Likelihood Functions for BC Model

	P=0.50	P=N1/N	P=Model
LogL =	-1854.17	-672.65	-205.13

Fit Measures based on Log Likelihood	
McFadden = 1-(L/L0) =	.69505
Estrella = 1-(L/L0)^(-2L0/n) =	.44968
R-squared (ML) =	.29499
Akaike Information Crit. =	.16308
Schwartz Information Crit. =	.19172

Fit Measures Based on Model Predictions	
Efron =	.66728
Ben Akiva and Lerman =	.95673
Veall and Zimmerman =	.77403
Cramer =	.66395

Predictions for Binary Choice Model. Predicted value is 1 when probability is greater than .500000, 0 otherwise. Note, column or row total percentages may not sum to 100% because of rounding. Percentages are of full sample.

Actual Value	Predicted Value 0	Predicted Value 1	Total Actual
0	2463 (92.1%)	27 (1.0%)	2490 (93.1%)
1	51 (1.9%)	134 (5.0%)	185 (6.9%)
Total	2514 (94.0%)	161 (6.0%)	2675 (100.0%)

Crosstab for Binary Choice Model. Predicted probability vs. actual outcome. Entry = Sum[Y(i,j)*Prob(i,m)] 0,1. Note, column or row total percentages may not sum to 100% because of rounding. Percentages are of full sample.

Actual Value	Predicted Probability Prob(y=0)	Prob(y=1)	Total Actual
y=0	2432 (90.9%)	57 (2.1%)	2490 (93.0%)
y=1	57 (2.1%)	127 (4.7%)	185 (6.9%)
Total	2489 (93.0%)	185 (6.9%)	2675 (99.9%)

The first set of matching results use the kernel estimator for the neighbors, lists all intermediate results, and uses only the observations in the common support. The output below is annotated.

MATCH ; Lhs = re78 ; Kernel ; List ; Common Support $

The estimated propensity score function is echoed first. This merely reports the earlier estimated model binary choice model for the treatment assignment. The treatment assignment model is not reestimated. (The **; Hold** in the **LOGIT** or **PROBIT** command stores the estimated model for this use.)

```
+----------------------------------------------------+
| ******* Propensity Score Matching Analysis ****** |
| Treatment variable = T       , Outcome = RE78     |
| Sample In Use                                     |
| Total number of observations     =   2675         |
| Number of valid (complete) obs.  =   2675         |(BI report 1342. Their
| Number used (in common support)  =   1347         | estimated probabilities
| Sample Partitioning of Data In Use                | are slightly different.
|                     Treated   Controls     Total  | Their reported counts are
| Observations            185       1162      1347  | 185 1157 1342)
| Sample Proportion    13.73%     86.27%   100.00%  |
+----------------------------------------------------+

+---------------------------------------------------------------+
| Propensity Score Function = Logit  based on T                 |
| Variable   Coefficient  Standard Error  t statistic           |
| AGE             .32965       .12042910       2.737            |
| AGE2           -.00633       .00185630      -3.409            |
| EDUC            .88403       .34149595       2.589            |
| EDUC2          -.05215       .01702488      -3.063            |
| MARR         -1.89160        .29919332      -6.322            |
| BLACK         1.13696        .35195110       3.230            |
| HISP          1.96830        .56695463       3.472            |
| RE74         -1.04742        .35896340      -2.918            |
| RE75         -2.18585        .41826670      -5.226            |
| RE742           .23048       .08230843       2.800            |
| RE752           .02516       .08840602        .285            |
| BLACKU74      2.13433        .42694403       4.999            |
| ONE          -7.63663       2.42743171      -3.146            |
| Note:Estimation sample may not be the sample analyzed here. |
| Observations analyzed are restricted to the common support =|
| only controls with propensity in the range of the treated.  |
+---------------------------------------------------------------+
```

The note in the reported logit results reports how the common support is defined, that is, as the range of variation of the scores for the treated observations.

The next set of results reports the iterations which partition the range of probabilities. The report includes the results of the F tests within the partitions as well as the details of the full partition itself. Becker and Ichino do not report the results of this search for their data, but do report that they ultimately found seven blocks whereas we find eight. They do not report the means by which the test of equality is carried out within the blocks or the critical value used. The method used here is reported in the mathematical details in Section E57.6.

Partitioning the range of propensity scores

Iteration 1. Partitioning range of propensity scores into 5 intervals.

Range		Controls			Treatment					
		# Obs.	Mean PS	S.D. PS	# obs.	Mean PS	S.D. PS	F	Prob	
.00059	.19544	1086	.02108	.03352	18	.08025	.06307	15.77	.0010	*
.19544	.39029	41	.28559	.05967	24	.30771	.05508	2.29	.1361	
.39029	.58514	15	.49623	.05068	21	.48810	.06451	.18	.6748	
.58514	.77999	13	.68860	.04677	19	.64604	.04682	6.39	.0179	
.77999	.97484	7	.96228	.00706	103	.92986	.05425	29.44	.0000	*

Iteration 1 Mean scores are not equal in at least one cell

Iteration 2. Partitioning range of propensity scores into 6 intervals.

Range		Controls			Treatment					
		# Obs.	Mean PS	S.D. PS	# obs.	Mean PS	S.D. PS	F	Prob	
.00059	.09802	1030	.01509	.02117	11	.03650	.03263	4.72	.0550	
.09802	.19544	56	.13121	.02746	7	.14901	.02862	2.43	.1632	
.19544	.39029	41	.28559	.05967	24	.30771	.05508	2.29	.1361	
.39029	.58514	15	.49623	.05068	21	.48810	.06451	.18	.6748	
.58514	.77999	13	.68860	.04677	19	.64604	.04682	6.39	.0179	
.77999	.97484	7	.96228	.00706	103	.92986	.05425	29.44	.0000	*

Iteration 2 Mean scores are not equal in at least one cell

Iteration 3. Partitioning range of propensity scores into 7 intervals.

Range		Controls			Treatment					
		# Obs.	Mean PS	S.D. PS	# obs.	Mean PS	S.D. PS	F	Prob	
.00059	.09802	1030	.01509	.02117	11	.03650	.03263	4.72	.0550	
.09802	.19544	56	.13121	.02746	7	.14901	.02862	2.43	.1632	
.19544	.39029	41	.28559	.05967	24	.30771	.05508	2.29	.1361	
.39029	.58514	15	.49623	.05068	21	.48810	.06451	.18	.6748	
.58514	.77999	13	.68860	.04677	19	.64604	.04682	6.39	.0179	
.77999	.87741	0	.00000	.00000	17	.81657	.02822	.00	1.0000	
.87741	.97484	7	.96228	.00706	86	.95225	.01815	9.18	.0090	*

Iteration 3 Mean scores are not equal in at least one cell

Iteration 4. Partitioning range of propensity scores into 8 intervals.

Range		Controls			Treatment				
		# Obs.	Mean PS	S.D. PS	# obs.	Mean PS	S.D. PS	F	Prob
.00059	.09802	1030	.01509	.02117	11	.03650	.03263	4.72	.0550
.09802	.19544	56	.13121	.02746	7	.14901	.02862	2.43	.1632
.19544	.39029	41	.28559	.05967	24	.30771	.05508	2.29	.1361
.39029	.58514	15	.49623	.05068	21	.48810	.06451	.18	.6748
.58514	.77999	13	.68860	.04677	19	.64604	.04682	6.39	.0179
.77999	.87741	0	.00000	.00000	17	.81657	.02822	.00	1.0000
.87741	.92612	0	.00000	.00000	7	.90719	.00944	.00	1.0000
.92612	.97484	7	.96228	.00706	79	.95625	.01244	4.01	.0732

Mean PSCORES are tested equal within the blocks listed below.

After partitioning the range of the propensity scores, we report the empirical distribution of the propensity scores and the boundaries of the blocks estimated above. The values below show the percentiles that are also reported by Becker and Ichino. Finally, the reported search algorithm notwithstanding, the block boundaries shown are rough.

```
+------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used  | BI results
|   Percent      Lower   Upper  Sample size =   1347         |(Percentiles)
|   0% -   5%   .000591  .000783 Average score  .137238      |(.0006426)
|   5% -  10%   .000787  .001061 Std.Dev score  .274079      |.0008025)
|  10% -  15%   .001065  .001377 Variance       .075119      |(.0010932)
|  15% -  20%   .001378  .001748 Blocks used to test balance |
|  20% -  25%   .001760  .002321      Lower    Upper   # obs |
|  25% -  30%   .002340  .002956  1  .000591  .098016   1041 |(.0023546)
|  30% -  35%   .002974  .004057  2  .098016  .195440     63 |
|  35% -  40%   .004059  .005272  3  .195440  .390289     65 |
|  40% -  45%   .005278  .007486  4  .390289  .585138     36 |
|  45% -  50%   .007557  .010451  5  .585138  .779986     32 |
|  50% -  55%   .010563  .014643  6  .779986  .877411     17 |(.0106667)
|  55% -  60%   .014686  .022462  7  .877411  .926123      7 |
|  60% -  65%   .022621  .035060  8  .926123  .974835     86 |
|  65% -  70%   .035075  .051415                             |
|  70% -  75%   .051415  .076188                             |(.0757115)
|  75% -  80%   .076376  .134189                             |
|  80% -  85%   .134238  .320638                             |
|  85% -  90%   .321233  .616002                             |
|  90% -  95%   .624407  .949418                             |(.6250823)
|  95% - 100%   .949418  .974835                             |(.949302 .970598)
+------------------------------------------------------------+
```

Becker and Ichino report the following blocks and sample sizes:

	Lower	Upper	Observations
1	0.0006	0.05	931
2	0.05	0.10	106
3	0.10	0.20	63
4	0.20	0.40	69
5	0.40	0.60	35
6	0.60	0.80	33
7	0.80	1.00	105

The next set of results reports the analysis of the balancing property for the independent variables. Note that a test is reported for each variable in each block as listed in the table above. The lines marked (by the program) with ‘*’ show cells in which one or the other group had no observations, so the F test could not be carried out. This was treated as a ‘success’ in each analysis. Lines marked with an ‘o’ note where the balancing property failed. There are relatively few of these, but those we do find are not borderline. Becker and Ichino report their finding that the balancing property is satisfied. Note that this finding does not prevent the further analysis. It merely suggests to analysts that they might want to consider a richer specification of the propensity function model.

```
Examining exogenous variables for balancing hypothesis
* Indicates no observations, treatment and/or controls, for test.
o Indicates means of treated and controls differ significantly.
=================================================================
Variable  Interval Mean Control   Mean Treated    F     Prob
--------  -------- ------------   ------------  ------  -----
AGE              1    31.426214      30.363636     .38  .5489
AGE              2    28.196429      28.714286     .02  .8978
AGE              3    27.902439      28.583333     .09  .7611
AGE              4    26.800000      24.809524     .60  .4458
AGE              5    24.846154      24.210526     .10  .7544
AGE              6      .000000      30.823529     .00 1.0000 *
AGE              7      .000000      28.857143     .00 1.0000 *
AGE              8    23.285714      23.392405     .02  .8843
AGE2             1  1078.808738     953.454545    1.37  .2659
AGE2             2   854.089286     923.857143     .07  .7932
AGE2             3   854.829268     891.416667     .07  .7997
AGE2             4   774.400000     676.523810     .36  .5553
AGE2             5   644.230769     623.789474     .03  .8568
AGE2             6      .000000    1003.058824     .00 1.0000 *
AGE2             7      .000000     884.000000     .00 1.0000 *
AGE2             8   543.857143     570.506329     .59  .4496
EDUC             1    11.216505      11.545455     .35  .5666
EDUC             2    10.339286      10.714286     .20  .6665
EDUC             3    10.634146       9.875000    1.59  .2135
EDUC             4    10.200000      10.190476     .00 1.0000
EDUC             5    10.230769      11.000000    1.03  .3218
EDUC             6      .000000      11.058824     .00 1.0000 *
EDUC             7      .000000      10.142857     .00 1.0000 *
EDUC             8    10.571429      10.037975     .88  .3729
EDUC2            1   132.542718     136.636364     .11  .7477
EDUC2            2   112.946429     119.000000     .12  .7413
EDUC2            3   117.609756     103.541667    1.70  .1983
EDUC2            4   108.066667     108.285714     .00 1.0000
EDUC2            5   109.923077     124.263158     .83  .3703
EDUC2            6      .000000     124.705882     .00 1.0000 *
EDUC2            7      .000000     105.285714     .00 1.0000 *
EDUC2            8   113.714286     104.215190     .70  .4258
MARR             1      .833010        .818182     .02  .9013
MARR             2      .571429        .857143    3.73  .0821
MARR             3      .268293        .250000     .03  .8712
MARR             4      .200000        .047619    1.81  .1935
MARR             5      .153846        .210526     .17  .6821
MARR             6      .000000        .529412     .00 1.0000 *
MARR             7      .000000        .000000     .00 1.0000 *
MARR             8      .000000        .000000     .00 1.0000
BLACK            1      .356311        .636364    3.69  .0811
BLACK            2      .625000        .571429     .07  .7935
BLACK            3      .756098        .750000     .00 1.0000
BLACK            4      .866667        .523810    6.00  .0194
BLACK            5      .846154        .947368     .81  .3792
BLACK            6      .000000        .941176     .00 1.0000 *
BLACK            7      .000000        .428571     .00 1.0000 *
BLACK            8     1.000000       1.000000     .00 1.0000
```

```
HISP          1         .048544        .000000   52.44  .0000 o
HISP          2         .071429        .285714    1.51  .2583
HISP          3         .048780        .000000    2.10  .1547
HISP          4         .066667        .142857     .58  .4508
HISP          5         .153846        .052632     .81  .3792
HISP          6         .000000        .058824     .00 1.0000 *
HISP          7         .000000        .571429     .00 1.0000 *
HISP          8         .000000        .000000     .00 1.0000
RE74          1        1.235202       1.214261     .01 1.0000
RE74          2         .572655        .203166   12.23  .0019 o
RE74          3         .597151        .524593     .22  .6437
RE74          4         .253634        .361641     .77  .3866
RE74          5         .154631        .197888     .44  .5108
RE74          6         .000000        .002619     .00 1.0000 *
RE74          7         .000000        .000000     .00 1.0000 *
RE74          8         .000000        .000000     .00 1.0000
RE75          1        1.050114        .896447     .44  .5197
RE75          2         .409156        .325001     .59  .4610
RE75          3         .271518        .296956     .15  .6984
RE75          4         .286058        .168348    2.54  .1213
RE75          5         .137276        .139118     .00 1.0000
RE75          6         .000000        .061722     .00 1.0000 *
RE75          7         .000000        .000000     .00 1.0000 *
RE75          8         .012788        .023447     .53  .4798
RE742         1        2.400191       2.335453     .00 1.0000
RE742         2         .651190        .079029    9.34  .0034 o
RE742         3         .652245        .684379     .01 1.0000
RE742         4         .127254        .360581    2.27  .1439
RE742         5         .040070        .095745    1.31  .2647
RE742         6         .000000        .000117     .00 1.0000 *
RE742         7         .000000        .000000     .00 1.0000 *
RE742         8         .000000        .000000     .00 1.0000
RE752         1        1.796624       1.446671     .53  .4761
RE752         2         .276672        .072511    8.78  .0048 o
RE752         3         .200186        .224688     .08  .7781
RE752         4         .082652        .091302     .04  .8366
RE752         5         .016499        .028328    1.06  .3127
RE752         6         .000000        .000019     .00 1.0000 *
RE752         7         .000000        .000000     .00 1.0000 *
RE752         8         .000000        .000000     .00 1.0000
BLACKU74      1         .014563        .000000   15.12  .0001 o
BLACKU74      2         .071429        .142857     .27  .6173
BLACKU74      3         .121951        .166667     .24  .6280
BLACKU74      4         .200000        .095238     .74  .3969
BLACKU74      5         .230769        .315789     .29  .5952
BLACKU74      6         .000000        .941176     .00 1.0000 *
BLACKU74      7         .000000        .428571     .00 1.0000 *
BLACKU74      8        1.000000       1.000000     .00 1.0000
Variable BLACKU74 is unbalanced in block  1
Other variables may also be unbalanced
You might want to respecify the index function for the P-scores
```

This part of the analysis ends with a recommendation that the analyst reexamine the specification of the propensity score model. Since this is not a numerical problem, the analysis continues with estimation of the average treatment effect on the treated.

The first example below shows estimation using the kernel estimator to define the counterpart observation from the controls. This stage consists of *nbot* + 1 iterations. The first is the actual estimation, which is reported in the intermediate results. Then the *nboot* repetitions are reported. (These will be omitted if **; List** is not included in the command.)

Recall, we divided the income values by 10,000. The value of .157435 reported below thus corresponds to $1,574.35. Becker and Ichino report a value (see their Section 6.4) of $1537.94 based on the 185 treateds and 1,157 controls. The result below is based on the same 185 treateds and 1,162 controls. Note that the kernel estimator is the most time consuming of the three approaches. For this sample of over 2,600 observations, the entire procedure required less than one second.

```
+--------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Kernel            Using Epanechnikov kernel with bandwidth =  .0600  |
| Note, controls may be reused in defining matches.                   |
| Number of bootstrap replications used to obtain variance    =    25 |
+--------------------------------------------------------------------+

  Estimated average treatment effect =       .157435
  Begin bootstrap iterations *******************************************
  Bootstrap estimate   1             =       .017963
  Bootstrap estimate   2             =       .267056
  Bootstrap estimate   3             =       .023318
  Bootstrap estimate   4             =       .082595
  Bootstrap estimate   5             =       .102630
  Bootstrap estimate   6             =       .011022
  Bootstrap estimate   7             =       .095340
  Bootstrap estimate   8             =       .131663
  Bootstrap estimate   9             =       .227142
  Bootstrap estimate  10             =       .048036
  (Iterations 11 – 20 omitted)
  Bootstrap estimate  21             =       .203207
  Bootstrap estimate  22             =       .006060
  Bootstrap estimate  23             =       .123456
  Bootstrap estimate  24             =       .120571
  Bootstrap estimate  25             =       .044657
  End bootstrap iterations   *******************************************
```

(Note, the values reported in parentheses below for the average treatment effect and the estimated asymptotic standard error are Becker and Ichino's estimates, not part of the output of the program. Their counterpart to the confidence interval shown below is (-.0479755 to 3.555643).

```
+--------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =   1162 |
| Estimated Average Treatment Effect   =       .157435  (.1537943)    |
| Estimated Asymptotic Standard Error  =       .096927  (.1016874)    |
| t statistic (ATT/Est.S.E.)           =      1.624276                |
| 95% Confidence Interval for ATT      = -.032541 to .347411) 95%     |
| Average Bootstrap estimate of ATT    =       .119419                |
| ATT - Average bootstrap estimate     =       .038017                |
+--------------------------------------------------------------------+
Elapsed time:     0 hours,  0 minutes,  0.75 seconds.
```

The next set of estimates is based on all of the program defaults. The single nearest neighbor is used for the counterpart observation; 25 bootstrap replications are used to compute the standard deviation, and the full range of propensity scores (rather than the common support) is used. Intermediate output is also suppressed.

MATCH ; Rhs = re78 $

```
+---------------------------------------------------+
| ******* Propensity Score Matching Analysis ****** |
| Treatment variable = T       , Outcome = RE78     |
| Sample In Use                                     |
| Total number of observations     =   2675         |
| Number of valid (complete) obs.  =   2675         |
| Number used                      =   2675         |
| Sample Partitioning of Data In Use                |
|                     Treated   Controls     Total  |
| Observations            185       2490      2675  |
| Sample Proportion     6.92%     93.08%   100.00%  |
+---------------------------------------------------+

+-------------------------------------------------------------+
| Propensity Score Function = Logit  based on T                |
| Variable   Coefficient  Standard Error  t statistic          |
| AGE             .32965       .12042909       2.737           |
| AGE2           -.00633       .00185630      -3.409           |
| EDUC            .88403       .34149593       2.589           |
| EDUC2          -.05215       .01702488      -3.063           |
| MARR          -1.89160       .29919331      -6.322           |
| BLACK          1.13696       .35195111       3.230           |
| HISP           1.96830       .56695458       3.472           |
| RE74          -1.04742       .35896354      -2.918           |
| RE75          -2.18585       .41826677      -5.226           |
| RE742           .23048       .08230861       2.800           |
| RE752           .02516       .08840623        .285           |
| BLACKU74       2.13433       .42694404       4.999           |
| ONE           -7.63663      2.42743151      -3.146           |
+-------------------------------------------------------------+
```

The reported estimated propensity score function is the same as before, as it is simply an echo of the earlier function. All of the subsequent results will be different because the previous example restricted the sample in use to those in the common support while the results to follow use all observations in the sample. Becker and Ichino do not report a like set of results, so we can only compare the final results here. The partitioning of the range of propensity scores once again produces eight blocks.

```
Partitioning the range of propensity scores
Iteration  1  Mean scores are not equal in at least one cell
Iteration  2  Mean scores are not equal in at least one cell
Iteration  3  Mean scores are not equal in at least one cell
Mean PSCORES are tested equal within the blocks listed below.
```

```
+-------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used  |
|   Percent      Lower   Upper   Sample size =   2675         |
|   0% -   5%   .000000  .000000 Average score  .069159       |
|   5% -  10%   .000000  .000002 Std.Dev score  .206222       |
|  10% -  15%   .000002  .000006 Variance       .042527       |
|  15% -  20%   .000006  .000015 Blocks used to test balance  |
|  20% -  25%   .000015  .000031      Lower    Upper   # obs  |
|  25% -  30%   .000032  .000062  1  .000000  .097484   2369  |
|  30% -  35%   .000062  .000122  2  .097484  .194967     63  |
|  35% -  40%   .000122  .000205  3  .194967  .389934     65  |
|  40% -  45%   .000206  .000365  4  .389934  .584901     36  |
|  45% -  50%   .000367  .000609  5  .584901  .779868     32  |
|  50% -  55%   .000613  .001111  6  .779868  .877352     17  |
|  55% -  60%   .001123  .001813  7  .877352  .926094      7  |
|  60% -  65%   .001824  .003037  8  .926094  .974835     86  |
|  65% -  70%   .003054  .005404                              |
|  70% -  75%   .005431  .011012                              |
|  75% -  80%   .011029  .023221                              |
|  80% -  85%   .023327  .051415                              |
|  85% -  90%   .051471  .135404                              |
|  90% -  95%   .135611  .624407                              |
|  95% - 100%   .627957  .974835                              |
+-------------------------------------------------------------+

Examining exogenous variables for balancing hypothesis
Variable BLACKU74 is unbalanced in block  1
Other variables may also be unbalanced
You might want to respecify the index function for the P-scores

+----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Nearest Neighbor  Using average of  1 closest neighbors              |
| Note, controls may be reused in defining matches.                    |
| Number of bootstrap replications used to obtain variance    =     25 |
+----------------------------------------------------------------------+
  Estimated average treatment effect =       .141870
+----------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =     55 |
| Estimated Average Treatment Effect   =        .141870                |
| Estimated Asymptotic Standard Error  =        .118904                |
| t statistic (ATT/Est.S.E.)           =       1.193147                |
| Confidence Interval for ATT = (     -.091182  to        .374921) 95% |
| Average Bootstrap estimate of ATT    =        .177653                |
| ATT - Average bootstrap estimate     =       -.035783                |
+----------------------------------------------------------------------+
```

Using the full sample in this fashion produces an estimate of $1,418.70 for the treatment effect with an estimated standard error of $1,189.04. Note that from the results above, we find that only 55 of the 2490 control observations were used as nearest neighbors for the 185 treated observations. In comparison, using the 1,342 observations in their estimated common support, and the same 185 treateds, Becker and Ichino report estimates of $1,667.64 and $2,113.59 for the effect and the standard error, respectively and use 57 of the 1,342 controls as nearest neighbors. Finally, this is the fastest of the three procedures. Computation based on the same sample now requires about a third of a second.

The next set of results uses the caliper form of matching and again restricts attention to the estimates in the common support.

MATCH ; Rhs = re78 ; Range = .0001 ; Common Support $

```
+---------------------------------------------------+
| ******* Propensity Score Matching Analysis ****** |
| Treatment variable = T       , Outcome = RE78     |
| Sample In Use                                     |
| Total number of observations     =   2675         |
| Number of valid (complete) obs.  =   2675         |
| Number used (in common support)  =   1347         |
| Sample Partitioning of Data In Use                |
|                     Treated   Controls     Total  |
| Observations            185       1162      1347  |
| Sample Proportion    13.73%     86.27%   100.00%  |
+---------------------------------------------------+

+-------------------------------------------------------------+
| Propensity Score Function = Logit  based on T               |
| Variable   Coefficient  Standard Error  t statistic         |
| AGE             .32965       .12042909       2.737          |
| AGE2           -.00633       .00185630      -3.409          |
| EDUC            .88403       .34149593       2.589          |
| EDUC2          -.05215       .01702488      -3.063          |
| MARR          -1.89160       .29919331      -6.322          |
| BLACK          1.13696       .35195111       3.230          |
| HISP           1.96830       .56695458       3.472          |
| RE74          -1.04742       .35896354      -2.918          |
| RE75          -2.18585       .41826677      -5.226          |
| RE742           .23048       .08230861       2.800          |
| RE752           .02516       .08840623        .285          |
| BLACKU74       2.13433       .42694404       4.999          |
| ONE           -7.63663      2.42743151      -3.146          |
| Note:Estimation sample may not be the sample analyzed here. |
| Observations analyzed are restricted to the common support =|
| only controls with propensity in the range of the treated.  |
+-------------------------------------------------------------+

Partitioning the range of propensity scores
Iteration  1  Mean scores are not equal in at least one cell
Iteration  2  Mean scores are not equal in at least one cell
Iteration  3  Mean scores are not equal in at least one cell
Mean PSCORES are tested equal within the blocks listed below.
```

```
+------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used |
|   Percent      Lower   Upper   Sample size =   1347        |
|   0% -   5%   .000591  .000783 Average score  .137238      |
|   5% -  10%   .000787  .001061 Std.Dev score  .274079      |
|  10% -  15%   .001065  .001377 Variance       .075119      |
|  15% -  20%   .001378  .001748 Blocks used to test balance |
|  20% -  25%   .001760  .002321       Lower    Upper   # obs |
|  25% -  30%   .002340  .002956  1  .000591  .098016   1041 |
|  30% -  35%   .002974  .004057  2  .098016  .195440     63 |
|  35% -  40%   .004059  .005272  3  .195440  .390289     65 |
|  40% -  45%   .005278  .007486  4  .390289  .585138     36 |
|  45% -  50%   .007557  .010451  5  .585138  .779986     32 |
|  50% -  55%   .010563  .014643  6  .779986  .877411     17 |
|  55% -  60%   .014686  .022462  7  .877411  .926123      7 |
|  60% -  65%   .022621  .035060  8  .926123  .974835     86 |
|  65% -  70%   .035075  .051415                             |
|  70% -  75%   .051415  .076188                             |
|  75% -  80%   .076376  .134189                             |
|  80% -  85%   .134238  .320638                             |
|  85% -  90%   .321233  .616002                             |
|  90% -  95%   .624407  .949418                             |
|  95% - 100%   .949418  .974835                             |
+------------------------------------------------------------+

Examining exogenous variables for balancing hypothesis
Variable BLACKU74 is unbalanced in block  1
Other variables may also be unbalanced
You might want to respecify the index function for the P-scores
```

Results to this point will be identical to the first set as the same sample and the same procedures are used to partition the range of propensity scores and test the balancing property. The estimated treatment effects are very different. We see that only 28 of the 185 controls had a neighbor within a range (radius in the terminology of Becker and Ichino) of 0.0001. The treatment effect is estimated to be only $167.16 with a standard error of $294.66. In contrast, using this procedure, and this radius, Becker and Ichino report a nonsense result of -$5,546.10 with a standard error of $2,388.72. They note that this illustrates the sensitivity of the estimator to the choice of radius, which is certainly the case. To examine this aspect, we recomputed the estimator using a range of 0.01 instead of 0.0001. This produces the expected effect, as seen in the second set of results below. The estimated treatment effect rises to $1,552.11 which is comparable to the other results already obtained.

```
+----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Caliper           Using distance of  .00010 to locate matches        |
| Note, controls may be reused in defining matches.                    |
| Number of bootstrap replications used to obtain variance    =     25 |
+----------------------------------------------------------------------+
  Estimated average treatment effect =       .016716
+----------------------------------------------------------------------+
| Number of Treated observations =     28  Number of controls =     74 |
| Estimated Average Treatment Effect   =        .016716                |
| Estimated Asymptotic Standard Error  =        .029466                |
| t statistic (ATT/Est.S.E.)           =        .567303                |
| Confidence Interval for ATT = (     -.041037  to       .074469) 95% |
| Average Bootstrap estimate of ATT    =        .011175                |
| ATT - Average bootstrap estimate     =        .005541                |
+----------------------------------------------------------------------+
```

The final results are produced by the command:

MATCH ; Rhs = re78 ; Range = .01 $

```
+----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Caliper           Using distance of  .01000 to locate matches        |
| Note, controls may be reused in defining matches.                    |
| Number of bootstrap replications used to obtain variance    =     25 |
+----------------------------------------------------------------------+
  Estimated average treatment effect =       .155211
  Begin bootstrap iterations *******************************************
  End bootstrap iterations   *******************************************
+----------------------------------------------------------------------+
| Number of Treated observations =    141  Number of controls =   1119 |
| Estimated Average Treatment Effect   =        .155211                |
| Estimated Asymptotic Standard Error  =        .068781                |
| t statistic (ATT/Est.S.E.)           =       2.256596                |
| Confidence Interval for ATT = (      .020400  to        .290022) 95% |
| Average Bootstrap estimate of ATT    =        .123361                |
| ATT - Average bootstrap estimate     =        .031850                |
+----------------------------------------------------------------------+
```

Finally, we examine the effect of using a probit model instead of a logit for the propensity scores. The first set of results below repeats the first set computed above. The second set is otherwise the same, save for the change to a probit model. The effect on the estimated treatment is very small, only $70 or about 4%. The standard error falls noticeably, but this is probably not a general result. The logit results are

```
+----------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =   1162 |
| Estimated Average Treatment Effect   =        .157435                |
| Estimated Asymptotic Standard Error  =        .096927                |
| t statistic (ATT/Est.S.E.)           =       1.624276                |
| Confidence Interval for ATT = (     -.032541  to        .347411) 95% |
| Average Bootstrap estimate of ATT    =        .119419                |
| ATT - Average bootstrap estimate     =        .038017                |
+----------------------------------------------------------------------+
```

The same set of computations based on a probit model for the propensity scores produces the following:

```
+----------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =   1042 |
| Estimated Average Treatment Effect   =        .150392                |
| Estimated Asymptotic Standard Error  =        .077791                |
| t statistic (ATT/Est.S.E.)           =       1.933289                |
| Confidence Interval for ATT = (     -.002078  to        .302862) 95% |
| Average Bootstrap estimate of ATT    =        .160183                |
| ATT - Average bootstrap estimate     =       -.009791                |
+----------------------------------------------------------------------+
```

E57.6 Mathematical Details of the Procedure

The following are the computations used to estimate the average treatment effect.

Propensity Score Model

The propensity scores are calculated using an estimated probit or logit binary choice model. The model estimates

$$
\begin{aligned}
P(T{=}1|\mathbf{x}) &= P(\mathbf{x}) \\
&= \Phi(\boldsymbol{\beta}'\mathbf{x}) \text{ for a probit model} \\
\text{and} \qquad &= \Lambda(\boldsymbol{\beta}'\mathbf{x}) \text{ for a logit model.}
\end{aligned}
$$

The model need not be estimated with the sample analyzed to compute the treatment effects. Any subsample, or a different sample entirely may be used. The specification of the treatment model should contain sufficient richness, perhaps with quadratic or interaction terms, to capture, as fully as possible, the underlying drivers of assignment to treatment. The next series of steps are applied to the sample to be used to estimate the average treatment effect.

Balancing Hypothesis

The data are examined to see if they satisfactorily meet the balancing hypothesis of equal means of treatment and controls – that is, to see if the treatment assignment between treated and controls is random given the characteristics.

Let the sample of propensity scores for the full sample be denoted P_i, those for the treated as P_t and for the controls P_c. It is decided at the outset whether to examine all individuals in the sample, or those whose propensities lie in the 'common support.' The common support consists of the range of propensities defined by

$$P_{min} = \text{Min}_t\, P_t \text{ to } P_{max} = \text{Max}_t\, P_t.$$

Thus, the sample consists of all the treated observations and the subset of controls whose propensity scores lie in this range.

The range of propensity scores is divided into a set of K intervals and the average propensity scores of treated and controls are tested for equality using the observations whose propensities lie within these ranges. *LIMDEP* uses the standard F test of equality of means,

$$\text{F}_k[1,d] = (\overline{P}_C^k - \overline{P}_T^k)^2 / (s_{C,k}^2 / N_C^k + s_{T,k}^2 / N_T^k),\ k = 1,...,K.$$

For the degrees of freedom for the denominator, we use the Satterthwaite approximation,

$$d = \frac{(s_{C,k}^2 / N_C^k + s_{T,k}^2 / N_T^k)^2}{[(s_{C,k}^2 / N_C^k)^2 / (N_C^k - 1)] + [(s_{T,k}^2 / N_T^k)^2 / (N_T^k - 1)]}\,.$$

We use a critical p value of 0.01 for the test. The default number of ranges is five. If the test fails in any cell, we use a finer partition of the range of scores. Two strategies are used.

1. Becker and Ichino recommend halving the range of the cell in which the test fails and repeating the test in the halves. Thus, if the initial ranges are .1-.2, .2-.3, .3-.4, .4-.5, .5-.6, and the test fails in the third cell, they convert the .3-.4 cell to two cells, .3-.35 and .35-.4 and repeating. By this calculation, a single cell can be partitioned into many small parts. In our first pass, we use this strategy, up to a maximum of 15 cells in total.

2. If the maximum of 15 cells is reached in pass one, we then start again with five equal length intervals, and if a cell fails the equal means test, we increase the number of cells to six and repeat the testing. In the case examined above, the second iteration would start again with cells .1000-.1833, .1833-.2666, .2666-.3500, .3500-.4333, .4333-.5167, .5167-.6000, and so on. By this strategy, the range of propensity scores is divided into finer, still equal length intervals. Once again, the iterations continue up to a maximum.

If a cell has insufficient observations to carry out the test, treat the F statistic as zero – that is, the means test passes for such a cell.

The outcome of this search will either be an indication that the overall test appears to pass, or if it persistently fails, a recommendation that the propensity score function is insufficiently specified. Either way, this failure does not prevent further processing. The result of this step is a mesh of points,

$$[P^*] = [P_1, P_2, ..., P_{K+1}]$$

that is then used in the next step. This mesh is the set of ranges shown at the right of the sample output below (which was reported earlier with our first set of results).

```
+---------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used    |
|   Percent      Lower    Upper   Sample size =   2675          |
|   0% -   5%   .000000  .000000 Average score  .069159         |
|   5% -  10%   .000000  .000002 Std.Dev score  .206222         |
|  10% -  15%   .000002  .000006 Variance       .042527         |
|  15% -  20%   .000006  .000015 Blocks used to test balance    |
|  20% -  25%   .000015  .000031      Lower    Upper   # obs    |
|  25% -  30%   .000032  .000062  1  .000000  .097484   2369    |
|  30% -  35%   .000062  .000122  2  .097484  .194967     63    |
|  35% -  40%   .000122  .000205  3  .194967  .389934     65    |
|  40% -  45%   .000206  .000365  4  .389934  .584901     36    |
|  45% -  50%   .000367  .000609  5  .584901  .779868     32    |
|  50% -  55%   .000613  .001111  6  .779868  .877352     17    |
|  55% -  60%   .001123  .001813  7  .877352  .926094      7    |
|  60% -  65%   .001824  .003037  8  .926094  .974835     86    |
|  65% -  70%   .003054  .005404                                |
|  70% -  75%   .005431  .011012                                |
|  75% -  80%   .011029  .023221                                |
|  80% -  85%   .023327  .051415                                |
|  85% -  90%   .051471  .135404                                |
|  90% -  95%   .135611  .624407                                |
|  95% - 100%   .627957  .974835                                |
+---------------------------------------------------------------+
```

Once the mesh [P*] is obtained, we then carry out a test of the balancing hypothesis for each variable in **x**, using the sets of observations that are contained in the ranges. A test of equality of means is carried for each variable in **x**, in each range defined by [P*]. The test statistic is computed in the same manner as above for the propensity scores, and, again, the 0.01 critical value is used. The outcome of this step is either a notification that the data are consistent with the balancing hypothesis (equal means), or they are not. Either way, this does not prevent further computation. The test of the variable age in our first analysis was reported as

```
Examining exogenous variables for balancing hypothesis
* Indicates no observations, treatment and/or controls, for test.
o Indicates means of treated and controls differ significantly.
=================================================================
Variable  Interval Mean Control   Mean Treated    F     Prob
--------  -------- ------------   ------------  ------  -----
AGE           1       31.426214      30.363636     .38  .5489
AGE           2       28.196429      28.714286     .02  .8978
AGE           3       27.902439      28.583333     .09  .7611
AGE           4       26.800000      24.809524     .60  .4458
AGE           5       24.846154      24.210526     .10  .7544
AGE           6         .000000      30.823529     .00 1.0000 *
AGE           7         .000000      28.857143     .00 1.0000 *
AGE           8       23.285714      23.392405     .02  .8843
```

Note that there are no control observations in the sixth and seventh blocks. These are taken to represent successes of the hypothesis.

Average Treatment Effect

The average treatment effect on the treated is now estimated. For each treated observation/outcome, O_t, we locate the counterpart control observations, O_c* that are similar in the characteristics, by using closeness of the propensity scores. The treatment effect for this observation is then estimated with O_t - O_c*. The average over the treated observations is then our estimate of the effect of treatment on the treated.

Note that this computation proceeds regardless of the outcome of the data examination in the previous step. However, a negative outcome in that step might call the results of this computation into question.

We offer three methods of locating the counterpart observation, O_c*:

Single Nearest Neighbor

The counterpart observation is the one that has the nearest propensity score. O_c* is the outcome for the person for whom $|P_t - P_c|$ is minimized. Note that a particular control observation may be the nearest neighbor to more than one treated observation. And, some controls may not be the closest neighbors to any treated observations.

Caliper

The counterpart observation is constructed by averaging all control observations whose propensity scores fall in a given range. Thus, we first locate the set $[C_t^*]$ = the set of control observations for which $|P_t - P_c| \leq r$, where the user specifies the value of r in the command. (The distance may be specified to be a specific value, such as .01, or a percentage of the range of propensity scores, such as P_t +/- 2% of $(P_{max} - P_{min})$. By this construction, the neighbor may be an average of several control observations. It may also not exist, if no observations are close enough. As in the single nearest neighbor computation, control observations may be used more than once, or they might not be used at all. (E.g., if the caliper is .01, a control observation has propensity .5 and the nearest treated observations have propensities of .45 and .55, then this control will never be used.)

Kernel

The counterpart observation is obtained by constructing a kernel estimator,

$$O_c^* = \frac{\sum_{c=1}^{N_c} O_c K\left[(P_t - P_c)/h\right]}{\sum_{c=1}^{N_c} K\left[(P_t - P_c)/h\right]} = \sum_{c=1}^{N_c} w_c O_c$$

where $K[.]$ is a kernel weighting function and h is the bandwidth. Three kernel functions are supported

Epanechnikov[z]	=	$.75\,(1 - z^2/5) / 5^{1/2}$ for $\lvert z\rvert \leq 5$
Normal	=	$\phi(z)$ = standard normal density
Logistic	=	$\Lambda(z)[1-\Lambda(z)]$ = logistic density
		$\Lambda(z) = \exp(z)/[1+\exp(z)]$.

The bandwidth may be specified by the user. The default value is 0.6; any positive value less than .25 may be specified. (The kernel function becomes unstable if the bandwidth is too large.) Note that this is a weighted average of the outcomes for all control observations, where the weights sum to one and are

$$w_c = \frac{K\left[(P_t - P_c)/h\right]}{\sum_{c=1}^{N_c} K\left[(P_t - P_c)/h\right]},\ 0 < w_c < 1,\quad \Sigma_c\, w_c = 1$$

In this instance, the neighbor is an average of all control observations.

Estimated Standard Error for the Average Treatment Effect

The variance of the estimator is estimated by using bootstrapping. The entire process is repeated *nboot* times, specified by the user. The default number is 25; up to 1,000 may be requested. The mean squared deviation around the actual estimator is used as the variance estimator. The square root is reported as the estimated asymptotic standard error.

Computation Time

Searching for the neighbors could be time consuming in a very large sample. The procedure is limited to samples of 200,000 observations or less. *LIMDEP's* algorithms are quite fast. The search is optimized by sorting the observations on propensity scores before any searching is done. Thus, for example, the search for the single nearest neighbor, which might involve searching the entire data set if the data are unsorted, is a trivial inspection of the few adjacent observations with the sorted data. Doing this entire analysis with a sample of 2,500 observations, and using the kernel estimator and 25 bootstrap iterations takes about 0.5 seconds, including estimating the probit equation, on a recent vintage desktop computer. Computation time will generally not be a substantive constraint.

E58: Nonparametric Analysis of Duration Data

E58.1 Introduction

This and the next two chapters will document *LIMDEP*'s programs for analyzing duration or lifetime (sometimes called 'survival' or 'failure time') data. This chapter presents the nonparametric (life table) methods. Chapter E59 narrows the analysis to 'semiparametric' models, which make minimal, but nonetheless substantive assumptions about the underlying distribution. Finally, Chapter E60 presents models which make explicit distributional assumptions for the duration data. Principal references for the developments in these chapters are: Kalbfleisch and Prentice (1980), Cox and Oakes (1984), Gross and Clark (1975), and Kiefer (1988).

The techniques and *LIMDEP* routines described here are used for analyzing duration data such as survival times, length of time until failure, lengths of spells of unemployment, strike duration, and so on. The data consist of measurements on the length of survival and, possibly, a set of regressors (covariates). In addition, the data may be 'right censored.' That is, the time measured may represent only the last observation of an individual who had not yet 'exited' the process being studied. For example, in studying spells of unemployment, the observed duration time may represent the full length of the spell, i.e., the length of time it took for the individual to find a job. Alternatively, at the time of measurement, the individual might have still been seeking a job. The duration datum in this case is censored; we know only (assume) that the individual left unemployment at some time after the measurement. The methods described here assume that observations are homogeneous in the probability distribution over duration times, with the exception of the measured covariates, if any. This is relaxed in Chapter E60. In addition, it is assumed that any censoring in the data is unrelated to the duration values themselves.

This chapter will show how to compute, store, and plot simple life tables for duration data. Treatment for a single sample is shown first. A method of stratifying the sample is shown at the end of the chapter.

E58.2 Life Tables

If only the duration times are available (i.e., no covariates), then life tables and survival curves can be derived by actuarial methods. In addition, if any observations are censored, the data must contain an indicator (binary) variable indicating which observations 'exited' (indicator = 1) and which observations are censored (indicator = 0).

NOTE: The maximum number of observations which can be analyzed is 75,000.

Suppose, then that observations consist of survival times, t_1, t_2, ..., t_N. Survival times are ordered low to high by the program. Your data need not be ordered; they are sorted internally. Also, let c_1, c_2, ..., c_N be the censoring indicator equal to zero for censored, or one for exited observations.

The following are computed:

Table 1. Life table based on the method of Cutler and Ederer (1958): The range of t is divided into K equal intervals. For each interval, $j = 1,\ldots,K$, we compute:

a. the number of observations, n_j,
b. size of the risk set, $r_j = n_j - C_j/2$, where C_j is the number of censored observations,
c. the number of observations which 'exit,' m_j ,
d. the proportion of observations in the risk set which exited, $q_j = m_j/r_j$,
e. the proportion surviving (the survival function) = the cumulative proportion of individuals surviving to the beginning of the interval,

$$p_j \quad = (1-q_{j-1})P_{j-1}, \text{ where } P_1=1,$$

f. standard error for estimated survival rate,

$$\text{se}(P_j) = P_j[\textstyle\sum_{j=1}^{k-1} q_k/(r_k(1-q_k))]^{1/2},$$

g. hazard rate,

$$\lambda_j \quad = 2q_j/(h(2-q_j)), \text{ where } h \text{ is the interval width,}$$

h. standard error for the hazard function,

$$\text{se}(\lambda_j) = \lambda_j[(1-(h\lambda_j/2)^2)/(r_j q_j)]^{1/2}.$$

The survival function and hazard function are then plotted. The median survival time is reported with the plotted survival function.

Table 2. If requested, the survival rates may be computed for the individual observations. The observations are sorted from low to high and the following are reported in a table:

a. observation,
b. survival time,
c. status – either exited or censored,
d. the cumulative survival rate to the time at which this individual was measured,
e. estimated standard error of the survival rate,
f. total number of observations which have exited up to that duration,
g. total number of observations censored at or less than that duration,
h. size of the risk set at the beginning of the period. (This is the number of observations whose duration is at least as large as that of this observation.)

Table 3. If the results in Table 2 are requested, the same information is reported for each distinct exit time in the sample. This will differ from the preceding if there are ties in the data. In the first table, when the data are sorted, the survival rate is computed based simply on observation number, so that only the last observation in a set of ties is meaningful.

E58.3 Commands for Life Tables

The basic command to request the nonparametric survival analysis is

SURVIVAL ; Lhs = time variable $

If the data are censored, the censoring indicator is given as a second Lhs variable. For example,

SURVIVAL ; Lhs = time,status $

The censoring indicator must be coded 1 for actual exit times and 0 for censored observations. The number of intervals in the life table is set at 10. This may be changed by using

; Int = number desired

The number given may be any value from 10 to 120.

The estimated hazard function and survival function can be plotted after the life table is displayed by adding

; Plot

to the command.

E58.3.1 Tables for Individuals and Specific Exit Times

The default output for the program is Table 1. Tables 2 and 3 are requested by adding

; List

to the command. The results in Table 3 may be kept permanently as data with the following: (Note, these are for the distinct exit times, not the observations.)

; Res = name saves the integrated hazard

The integrated hazard is a form of 'generalized residual,' which can be used to analyze model specification. (See Lancaster (1985).) (At this point, there is no model as such.) When the density, $f(t)$ and survival rate, $S(t)$ are defined, the integrated hazard is computed as

$$\int \lambda_i(t)dt = \int [f(t)/S(t)]d\mathrm{t} = -\log S(t).$$

We approximate this function with our estimated survival rates.

; Keep = name saves the hazard rate
; Fill keeps the distinct duration times

Since these variables are not specific to the observation, they are placed in the first K rows of the data area, where K is the number of distinct exit times. A scalar called '*numexit*' contains the number of exit times observed in the data set.

HINT: The **; Fill** option replaces the duration variable with its results. So, if you intend to use this option, use **CREATE** to make a copy of the original duration variable and use the copy as your duration variable in the command.

Separate plots will show the hazard and survival rates. An example is given below.

E58.3.2 Stratification

The survival tables and other analysis may be based on stratified data. The stratification must be provided by a variable which takes values 1, 2, ... As many as nine strata may be analyzed. A full set of results is provided for each stratum identified. At the end, two statistics, the log-rank and generalized Wilcoxon, are computed for testing homogeneity of the survival distributions across the strata. The command is:

```
SURVIVAL    ; Lhs = time [,status]
            ; Str = stratification indicator variable
            ; other options $
```

The strata must be identified explicitly. If you have a variable whose values you wish to use to define the strata, it is only necessary to use **RECODE** to create the stratification variable. For example, suppose *age* in the ranges 18-24, 25-45, and 46-99 is used to define three strata. You could use

```
CREATE      ; strat = age $
RECODE      ; strat ; 18 / 24 = 1 ; 25 / 45 = 2 ; 46 / 99 = 3 $
SURVIVAL    ; Lhs = time ; Str = strat $
```

E58.4 Applications

The estimators are illustrated with two data sets, one on strike duration from Kennan (1985) and the other a simulated data set with covariates.

E58.4.1 Strike Duration Data

The data listed below are the durations of 62 strikes reported by Kennan (1985). The last 12 are censored at 80 weeks.

```
READ        ; Nobs = 62 ; Nvar = 1 ; By Variables ; Names = time $
  1   2   2   2   2   3   3   3   9   12   21   27   43   49
 52   3   4   5   7   8   9  10  11   12   13   14   15   17
 19  21  22  23  25  26  27  28  29   32   33   35   37   38
 41  42  43  44  49  52  61  72  80   80   80   80   80   80
 80  80  80  80  80  80
CREATE      ; status = time < 80 $
SURVIVAL    ; Lhs = time,status ; Int = 20 ; List ; Plot $
```

```
Estimated Survival Function
Duration variable is        TIME
Status is given by variable STATUS
Number of observations in stratum =    62
Number of observations exiting    =    50
Number of observations censored   =    12
```

Survival	Enter	Cnsrd	At Risk	Exited	Survival Rate	Hazard Rate
.0- 4.0	62	0	62	9	1.0000 (.000)	.0391 (.013)
4.0- 8.0	53	0	53	3	.8548 (.045)	.0146 (.008)
8.0- 12.0	50	0	50	5	.8065 (.050)	.0263 (.012)
12.0- 16.0	45	0	45	5	.7258 (.057)	.0294 (.013)
16.0- 20.0	40	0	40	2	.6452 (.061)	.0128 (.009)
20.0- 24.0	38	0	38	4	.6129 (.062)	.0278 (.014)
24.0- 28.0	34	0	34	4	.5484 (.063)	.0313 (.016)
28.0- 32.0	30	0	30	2	.4839 (.063)	.0172 (.012)
32.0- 36.0	28	0	28	3	.4516 (.063)	.0283 (.016)
36.0- 40.0	25	0	25	2	.4032 (.062)	.0208 (.015)
40.0- 44.0	23	0	23	4	.3710 (.061)	.0476 (.024)
44.0- 48.0	19	0	19	1	.3065 (.059)	.0135 (.014)
48.0- 52.0	18	0	18	2	.2903 (.058)	.0294 (.021)
52.0- 56.0	16	0	16	2	.2581 (.056)	.0333 (.024)
56.0- 60.0	14	0	14	0	.2258 (.053)	.0000 (.000)
60.0- 64.0	14	0	14	1	.2258 (.053)	.0185 (.019)
64.0- 68.0	13	0	13	0	.2097 (.052)	.0000 (.000)
68.0- 72.0	13	0	13	0	.2097 (.052)	.0000 (.000)
72.0- 76.0	13	0	13	1	.2097 (.052)	.0200 (.020)
76.0- 80.0	12	12	6	0	.1935 (.050)	.0000 (.000)

Individual Survival Data

Observation	Survival	Status	Srv.rate (S.E.)	Exited	Censored	# at risk
1	1.000	Exited	1.0000 (.0000)	1	0	62
2	2.000	Exited	.9839 (.0160)	2	0	61
3	2.000	Exited	.9677 (.0224)	3	0	60
4	2.000	Exited	.9516 (.0273)	4	0	59
5	2.000	Exited	.9355 (.0312)	5	0	58
6	3.000	Exited	.9194 (.0346)	6	0	57
7	3.000	Exited	.9032 (.0375)	7	0	56
8	3.000	Exited	.8871 (.0402)	8	0	55
16	3.000	Exited	.8710 (.0426)	9	0	54
17	4.000	Exited	.8548 (.0447)	10	0	53
18	5.000	Exited	.8387 (.0467)	11	0	52
19	7.000	Exited	.8226 (.0485)	12	0	51
20	8.000	Exited	.8065 (.0502)	13	0	50
9	9.000	Exited	.7903 (.0517)	14	0	49
21	9.000	Exited	.7742 (.0531)	15	0	48
22	10.000	Exited	.7581 (.0544)	16	0	47
23	11.000	Exited	.7419 (.0556)	17	0	46
10	12.000	Exited	.7258 (.0567)	18	0	45
24	12.000	Exited	.7097 (.0576)	19	0	44
25	13.000	Exited	.6935 (.0585)	20	0	43
26	14.000	Exited	.6774 (.0594)	21	0	42
27	15.000	Exited	.6613 (.0601)	22	0	41
28	17.000	Exited	.6452 (.0608)	23	0	40
29	19.000	Exited	.6290 (.0613)	24	0	39
11	21.000	Exited	.6129 (.0619)	25	0	38
30	21.000	Exited	.5968 (.0623)	26	0	37
31	22.000	Exited	.5806 (.0627)	27	0	36
32	23.000	Exited	.5645 (.0630)	28	0	35
33	25.000	Exited	.5484 (.0632)	29	0	34
34	26.000	Exited	.5323 (.0634)	30	0	33
12	27.000	Exited	.5161 (.0635)	31	0	32
35	27.000	Exited	.5000 (.0635)	32	0	31
36	28.000	Exited	.4839 (.0635)	33	0	30
37	29.000	Exited	.4677 (.0634)	34	0	29

```
      38        32.000   Exited     .4516 (.0632)     35         0         28
      39        33.000   Exited     .4355 (.0630)     36         0         27
      40        35.000   Exited     .4194 (.0627)     37         0         26
      41        37.000   Exited     .4032 (.0623)     38         0         25
      42        38.000   Exited     .3871 (.0619)     39         0         24
      43        41.000   Exited     .3710 (.0613)     40         0         23
      44        42.000   Exited     .3548 (.0608)     41         0         22
      13        43.000   Exited     .3387 (.0601)     42         0         21
      45        43.000   Exited     .3226 (.0594)     43         0         20
      46        44.000   Exited     .3065 (.0585)     44         0         19
      14        49.000   Exited     .2903 (.0576)     45         0         18
      47        49.000   Exited     .2742 (.0567)     46         0         17
      15        52.000   Exited     .2581 (.0556)     47         0         16
      48        52.000   Exited     .2419 (.0544)     48         0         15
      49        61.000   Exited     .2258 (.0531)     49         0         14
      50        72.000   Exited     .2097 (.0517)     50         0         13
      51        80.000   Censored   .1935 (.0502)     50         1         12
      ...
      62        80.000   Censored   .1935 (.0502)     50        12          1

Summary of Duration Data
Observation  Survival   Status   Srv.rate (S.E.) Exited  Censored  # at risk
       1       1.000      .0161  1.0000 (.0000)      1         0         62
       2       2.000      .0656   .9839 (.0160)      4         0         61
       3       3.000      .0702   .9194 (.0346)      4         0         57
       4       4.000      .0189   .8548 (.0447)      1         0         53
       5       5.000      .0192   .8387 (.0467)      1         0         52
       6       7.000      .0196   .8226 (.0485)      1         0         51
       7       8.000      .0200   .8065 (.0502)      1         0         50
       8       9.000      .0408   .7903 (.0517)      2         0         49
       9      10.000      .0213   .7581 (.0544)      1         0         47
      10      11.000      .0217   .7419 (.0556)      1         0         46
      11      12.000      .0444   .7258 (.0567)      2         0         45
      12      13.000      .0233   .6935 (.0585)      1         0         43
      13      14.000      .0238   .6774 (.0594)      1         0         42
      14      15.000      .0244   .6613 (.0601)      1         0         41
      15      17.000      .0250   .6452 (.0608)      1         0         40
      16      19.000      .0256   .6290 (.0613)      1         0         39
      17      21.000      .0526   .6129 (.0619)      2         0         38
      18      22.000      .0278   .5806 (.0627)      1         0         36
      19      23.000      .0286   .5645 (.0630)      1         0         35
      20      25.000      .0294   .5484 (.0632)      1         0         34
      21      26.000      .0303   .5323 (.0634)      1         0         33
      22      27.000      .0625   .5161 (.0635)      2         0         32
      23      28.000      .0333   .4839 (.0635)      1         0         30
      24      29.000      .0345   .4677 (.0634)      1         0         29
      25      32.000      .0357   .4516 (.0632)      1         0         28
      26      33.000      .0370   .4355 (.0630)      1         0         27
      27      35.000      .0385   .4194 (.0627)      1         0         26
      28      37.000      .0400   .4032 (.0623)      1         0         25
      29      38.000      .0417   .3871 (.0619)      1         0         24
      30      41.000      .0435   .3710 (.0613)      1         0         23
      31      42.000      .0455   .3548 (.0608)      1         0         22
      32      43.000      .0952   .3387 (.0601)      2         0         21
      33      44.000      .0526   .3065 (.0585)      1         0         19
      34      49.000      .1111   .2903 (.0576)      2         0         18
      35      52.000      .1250   .2581 (.0556)      2         0         16
      36      61.000      .0714   .2258 (.0531)      1         0         14
      37      72.000      .0769   .2097 (.0517)      1         0         13
      38      80.000      .0000   .1935 (.0502)      0        12         12
```

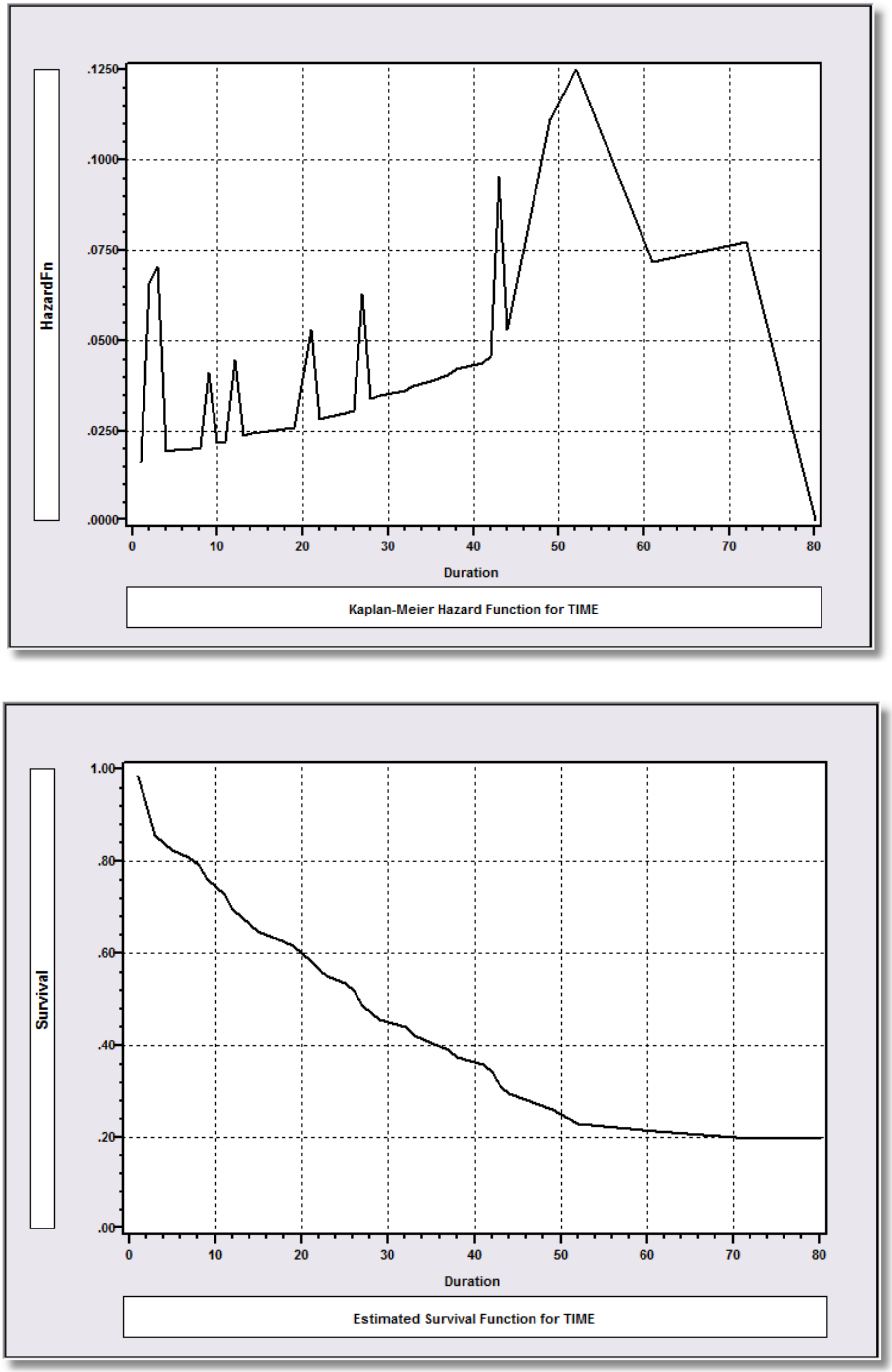


Figure E58.1 Nonparametric Estimates of Hazard and Survival Functions

E58.4.2 An Example with Stratification

The following reads an artificial set of data on censoring and duration and requests an analysis stratified on marital status.

```
READ        ; Nobs = 22 ; Nvar = 4
            ; Names = time,status,sex,married ; By Variables $

  11  3 19 32 2 14 8 21 16 5 2 8 14 18 18 21 10 1 9 23 19 7
   1  1  0  1 1  1 1  1  0 1 1 1  1  1  1  1  0 1 0  1  1 1
   0  0  1  0 1  0 1  1  1 0 0 1  1  0  1  1  0 0 1  1  0 1
   1  1  2  2 1  1 1  1  2 2 2 1  1  1  2  1  2 2 1  1  2 1

SURVIVAL    ; Lhs = time,status ; Str = married ; List $
```

```
 +-------------------------------------------------+
 | Estimated Survival Function                     |
 | Duration variable is        TIME                |
 | Status is given by variable STATUS              |
 | Stratification variable is  MARRIED             |
 | Number of strata is         2                   |
 | Counts are:    Stratum      Count               |
 |                   1           13                |
 |                   2            9                |
 +-------------------------------------------------+

Estimation results for stratum MARRIED  = 1
Number of observations in stratum =      13
Number of observations exiting    =      12
Number of observations censored   =       1
    Survival   Enter  Cnsrd  At Risk  Exited   Survival Rate     Hazard Rate
    .0-   2.3     13      0       13       1  1.0000 ( .000)   .0348 ( .035)
   2.3-   4.6     12      0       12       1   .9231 ( .074)   .0378 ( .038)
   4.6-   6.9     11      0       11       0   .8462 ( .100)   .0000 ( .000)
   6.9-   9.2     11      1       10       3   .8462 ( .100)   .1449 ( .083)
   9.2-  11.5      7      0        7       1   .6044 ( .138)   .0669 ( .067)
  11.5-  13.8      6      0        6       0   .5181 ( .143)   .0000 ( .000)
  13.8-  16.1      6      0        6       2   .5181 ( .143)   .1739 ( .120)
  16.1-  18.4      4      0        4       1   .3454 ( .138)   .1242 ( .123)
  18.4-  20.7      3      0        3       0   .2590 ( .128)   .0000 ( .000)
  20.7-  23.0      3      0        3       3   .2590 ( .128)   .8696 ( .000)
Individual Survival Data
Observation  Survival   Status   Srv.rate (S.E.) Exited  Censored  # at risk
       5        2.000   Exited    1.0000 (.0000)      1         0         13
       2        3.000   Exited     .9231 (.0739)      2         0         12
      22        7.000   Exited     .8462 (.1001)      3         0         11
       7        8.000   Exited     .7692 (.1169)      4         0         10
      12        8.000   Exited     .6923 (.1280)      5         0          9
      19        9.000   Censored   .6154 (.1349)      5         1          8
       1       11.000   Exited     .6154 (.1349)      6         1          7
       6       14.000   Exited     .5275 (.1414)      7         1          6
      13       14.000   Exited     .4396 (.1426)      8         1          5
      14       18.000   Exited     .3516 (.1385)      9         1          4
       8       21.000   Exited     .2637 (.1288)     10         1          3
      16       21.000   Exited     .1758 (.1119)     11         1          2
      20       23.000   Exited     .0879 (.0836)     12         1          1
```

```
Summary of Duration Data
Observation  Survival   Status   Srv.rate (S.E.) Exited  Censored  # at risk
       1       3.000      .0909  1.0000 (.0000)      2         0         13
       2       7.000      .0500   .8462 (.1001)      1         0         12
       3       8.000      .1053   .7692 (.1169)      2         0         11
       4       9.000      .0000   .6154 (.1349)      0         1          9
       5      11.000      .0625   .6154 (.1349)      1         0          8
       6      14.000      .1333   .5275 (.1414)      2         0          7
       7      21.000      .2308   .3516 (.1385)      3         0          5
       8      23.000      .1000   .0879 (.0836)      1         0          2

Estimation results for stratum MARRIED  = 2
Number of observations in stratum =       9
Number of observations exiting    =       6
Number of observations censored   =       3
    Survival   Enter  Cnsrd  At Risk  Exited   Survival Rate     Hazard Rate
    .0-   3.2      9      0        9       2  1.0000 ( .000)   .0781 ( .055)
   3.2-   6.4      7      0        7       1   .7778 ( .139)   .0481 ( .048)
   6.4-   9.6      6      0        6       0   .6667 ( .157)   .0000 ( .000)
   9.6-  12.8      6      1        5       0   .6667 ( .157)   .0000 ( .000)
  12.8-  16.0      5      1        4       0   .6667 ( .157)   .0000 ( .000)
  16.0-  19.2      4      1        3       2   .6667 ( .157)   .2500 ( .162)
  19.2-  22.4      1      0        1       0   .2857 ( .189)   .0000 ( .000)
  22.4-  25.6      1      0        1       0   .2857 ( .189)   .0000 ( .000)
  25.6-  28.8      1      0        1       0   .2857 ( .189)   .0000 ( .000)
  28.8-  32.0      1      0        1       1   .2857 ( .189)   .6250 ( .000)
Individual Survival Data
Observation  Survival   Status   Srv.rate (S.E.) Exited  Censored  # at risk
      18       1.000   Exited    1.0000 (.0000)      1         0          9
      11       2.000   Exited     .8889 (.1048)      2         0          8
      10       5.000   Exited     .7778 (.1386)      3         0          7
      17      10.000   Censored   .6667 (.1571)      3         1          6
       9      16.000   Censored   .6667 (.1571)      3         2          5
      15      18.000   Exited     .6667 (.1571)      4         2          4
      21      19.000   Exited     .5000 (.1863)      5         2          3
       3      19.000   Censored   .3333 (.1843)      5         3          2
       4      32.000   Exited     .3333 (.1843)      6         3          1
Summary of Duration Data
Observation  Survival   Status   Srv.rate (S.E.) Exited  Censored  # at risk
       1       1.000      .0455  1.0000 (.0000)      1         0          9
       2       2.000      .0476   .8889 (.1048)      1         0          8
       3       5.000      .0500   .7778 (.1386)      1         0          7
       4      10.000      .0000   .6667 (.1571)      0         1          6
       5      16.000      .0000   .6667 (.1571)      0         1          5
       6      18.000      .0588   .6667 (.1571)      1         0          4
       7      19.000      .0625   .5000 (.1863)      1         1          2
       8      32.000      .0714   .3333 (.1843)      1         0          1

           +------------------------------------------------------+
           | Homogeneity tests: Degrees of freedom=   1           |
           |                                                      |
           | Log-Rank (LM) =     .93355    ,  Prob.   .33394     |
           |                                                      |
           | Gen. Wilcoxon =     .15433    ,  Prob.   .69443     |
           +------------------------------------------------------+
```

E58.5 Technical Details for the Homogeneity Tests

The log-rank and generalized Wilcoxon tests are both used for testing the hypothesis of homogeneity of the strata. They are computed as follows: Let

K = the number of strata, strata denoted $k = 1,...,K$,
N = the number of distinct exit times,
T_i = the exit time at time 'i,'
n_{ik} = the number of individuals in stratum k with exit time $t_{ik} \geq T_i$,
$n_{i.}$ = $\Sigma_k n_{ik}$ = number of individuals in the sample with $t_{ik} \geq T_i$,
x_{ik} = number of individuals who exit stratum k at time T_i,
$x_{i.}$ = $\Sigma_k x_{ik}$ = number of individuals in the sample who exit at time T_i,
$\mathbf{x}_i$ = $[x_{i1}, x_{i2}, ..., x_{iK}]'$.

Under the assumption of homogeneity, conditioned on the sums n_{ik} and $x_{i.}$, the vector $\mathbf{x}_i$ has a (K-1) dimensional hypergeometric distribution with mean vector

$$E[x_{ik}] = n_{ik} x_{i.} / n_{i.}, \; k = 1,...,K,$$

and covariances

$$\mathrm{Cov}[x_{ik}, x_{il}] = n_{ij} (\delta_{kl} - n_{il} / n_{i.}) x_{i.} (n_{i.} - x_{i.}) / [n_{i.} (n_{i.} - 1)], \; \delta_{kl} = \mathbf{1}(k = l).$$

Let $\mathbf{x} = \Sigma_i \mathbf{x}_i$, $\mathbf{E} = \Sigma_i E[\mathbf{x}_i]$, and $\mathbf{V} = \Sigma_i \mathrm{Var}[\mathbf{x}_i]$.

The log-rank statistic is

$$\mathrm{LR} = (\mathbf{x} - \mathbf{E})'\mathbf{V}^{-1}(\mathbf{x} - \mathbf{E}).$$

This has a limiting chi squared distribution with K-1 degrees of freedom. Since $\mathbf{V}$ is short ranked, its ordinary inverse does not exist. We use a G2 inverse to compute the statistic.

The generalized Wilcoxon statistic is a slight modification. Let

$$w_{ik} = n_{i.} (x_{ik} - x_{i.} n_{ik} / n_{i.}),$$

$$w_k = \Sigma_i w_{ik},$$

and

$$\mathbf{w} = [w_1, w_2, ..., w_K]'.$$

This vector has mean $\mathbf{0}$ and covariance matrix

$$\mathbf{Q} = \Sigma_i n_{i.}^2 \mathrm{Var}[\mathbf{x}_i].$$

The statistic is

$$\mathrm{GW} = \mathbf{w}'\mathbf{Q}^{-1}\mathbf{w}.$$

Once again, a generalized inverse is needed to compute the statistic.

E59: Proportional Hazard Models

E59.1 Introduction

Two models for duration data are presented in this chapter. Cox's proportional hazards model has proven useful for modeling duration data with only minimal assumptions about the underlying distribution. The shortcoming, however, is that this approach can be rather inflexible. The second model considered here, Han and Hausman's ordered logit model is, like the Cox model, semiparametric in that it only assumes a basic form for the hazard function.

E59.2 The Proportional Hazards Model

If one or more covariates are observed with the duration data, a regression-like model derived by Cox (1972) may be estimated. The formal model is based on the hazard rate at time t,

$$h(t,\mathbf{x}) = h(t,\mathbf{0})e^{\boldsymbol{\beta}'\mathbf{x}},$$

where $h(t,\mathbf{0})$ is the baseline hazard rate at time t for covariate vector $\mathbf{0}$. Assumptions for the model are presented in Cox (1972, 1975) and the related references cited there. The parameters are estimated as follows: We allow for ties and censored data in the measured durations. Let $T_1,...,T_K$ be the set of K distinct times in the N observations. Let R_j be the index set of the individuals at risk just prior to time T_j (i.e., the set of individuals with duration greater than or equal to t_j). For every individual i in R_j, $t_i \geq T_j$. The probability that an individual 'exits' (dies, leaves, etc.) at time T_j, given that exactly this one individual exits at time T_j, is

$$\text{Prob(exit at time } T_j) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_j)}{\sum_{i\in R_j}\exp(\boldsymbol{\beta}'\mathbf{x}_i)}.$$

The conditioning eliminates the baseline hazard. If exactly one individual exits at each time and no observations are censored, the partial log likelihood (see Cox (1975)) is

$$\log L = \sum_{j=1}^{K} \boldsymbol{\beta}'\mathbf{x}_j - \log\left[\sum_{i\in R_j}\exp(\boldsymbol{\beta}'\mathbf{x}_i)\right].$$

If $m_j \geq 1$ individuals exit at the same t_j, the partial log likelihood is the sum of the individual likelihoods,

$$\log L = \sum_{j=1}^{K} \boldsymbol{\beta}'\sum_{r\in T_j}\mathbf{x}_r - m_j\log\left[\sum_{i\in R_j}\exp(\boldsymbol{\beta}'\mathbf{x}_i)\right].$$

Censored observations enter the risk set at each observation but do not contribute to the numerator of the partial likelihood.

The partial log likelihood is maximized using Newton's method. Options available for this model include stratification, time dependent covariates, and fixed values of the parameters.

E59.2.1 Commands for the Proportional Hazards Model

The minimal command for this model is

SURVIVAL ; Lhs = time [,status] ; Rhs = list of covariates $

This differs from the model of the previous chapter only in the list of Rhs variables.

NOTE: This model is homogeneous of degree zero in **x**. Any variable which does not vary over individuals will simply multiply both numerator and denominator of the partial likelihood, and hence drop out of it. If it is found, the variable *one* is automatically removed from your Rhs list. But, if there are other covariates which are constant over individuals, the Hessian will become singular and the estimation process will break down.

A censoring indicator variable is provided exactly as before, as a second Lhs variable. This is indicated as the optional **[,status]** variable above. If you provide a status variable, code it as one for complete observations and zero for censored observations.

Standard Model Specifications for the Cox Proportional Hazards Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps survival rates as a new (or replacement) variable in data set.
; Res = name	keeps integrated hazards as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.
; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

The algorithm is preset to Newton's method. This form of the partial likelihood is quite well behaved, and convergence is normally routine after only a few iterations. If you do not find this to be the case, the problem is probably multicollinearity among the covariates. Time dependent covariates can also be problematic. However, there may be cases in which BFGS is a preferable algorithm. In general, Newton's method should be best.

There are no predictions or residuals produced for this model, since the baseline hazard is not estimated. A number of related variables are computed. These are described below.

E59.2.2 Plotting the Survival and Integrated Hazard Functions

As part of the results for this model, *LIMDEP* will display plots of the survival function and the integrated hazard rate, computed at the means of the covariates. Use

; Plot

to request the figures. You may also request additional plots at specified values of the Rhs variables. To use other values which you provide:

Step 1. Load these values as the rows of a matrix using the **MATRIX** command. Load one row for each set of values you wish to use for a plot.

Step 2. Add **; Plot = name of the matrix** to the **SURVIVAL** command.

This will produce an additional pair of plots for each row of the matrix, i.e., for each set of values.

E59.2.3 Keeping the Survival and Integrated Hazard Functions

You can also keep an estimate of the survival and/or integrated hazard function in your data area. Use

; Keep = name of variable for survival function
and **; Res = name of variable for integrated hazard**

(The integrated hazard function is a 'generalized residual.') These are computed for each observation in the sample. The formulas used for these computations are as follows: (Note, no correction is made for censoring. This can be problematic if the data are heavily censored.)

1. Obtain N distinct exit times, $t_1, t_2, ..., t_N$; $t_0 = 0$.

2. Obtain N counts of exit times; m_i = number of exits at time t_i.

3. The baseline hazard at t_i is

$$h(t_i,0) \quad = \quad h(t_i,0) = \frac{m_i}{(t_i - t_{i-1})\sum_{j\in \mathrm{R_i}\ \text{such that}\ T_j \geq t_i} \exp(\boldsymbol{\beta}'\mathbf{x}_j)}.$$

4. The baseline integrated hazard is

$$H(t_i,0) \quad = \quad \sum_{j=1}^{i}(t_j - t_{j-1})h(t_j,0)$$

5. For each individual,

$$H(t_i,\mathbf{x}_i) \; = \; \exp(\boldsymbol{\beta}'\mathbf{x}_i)H(t_i,\mathbf{0}).$$

   (This is the generalized residual for individual i. It is kept by **; Res**.)

6. The survival probability is

$$S(t_i,\mathbf{x}_i) \quad = \exp[-H(t_i,\mathbf{x}_i)].$$

   This estimated survival rate is not necessarily monotonic in t_i in the sample because $\mathbf{x}_i$ differs across observations.

E59.2.4 Time Dependent Covariates

It may be useful to include in the model covariates which change over time or are functions of time and other variables. *LIMDEP* allows a large amount of flexibility in specifying these. The feature described here is for covariates which are explicit functions of time, i.e., time *dependent* covariates. Time *varying* covariates, such as marital status, cannot be handled in the proportional hazards model. They are, however, permissible in the parametric models discussed in the next chapter. To include time *dependent* covariates (TVCs), add the following specification, once for each one you want to specify:

; TVC = specification
; TVC = specification ...

The covariates in the model will now be the Rhs variables plus the TVCs. Thus, each one adds a new coefficient to the model. The Rhs variables must exist in the data set. The TVCs are computed during the iterations. (An example is given below.)

The specifications which may be used are as follows:

TVC = *expression*,
TVC = Log(*expression*) - natural log,
TVC = Exp(*expression*) - e raised to the power,
TVC = Abs(*expression*) - absolute value.

Expressions are any algebraic function of the data and 'time' (see below) with the following restrictions:

- Expressions may not contain parentheses except for the special notation described below.
- Expressions are evaluated strictly from left to right, with multiplication, division, and the other operators described below taking precedence over addition and subtraction.

The basic operators which may be used in the expression are +, -, *, /, and ^. (The last means raise to the following power). Operands in the expressions may be any variable in the data set, whether in the model elsewhere or not, or any scalar, either given explicitly, i.e., 3.14159265, or in a scalar referred to by name, such as *pi* (also 3.14159265) or *myown* (which you would have calculated earlier). Before describing the entry of time into the expression, we note an important aspect of the computational rule. The left to right rule will only depart from the obvious when the '^' operator is used. When calculated in this fashion,

$$x \wedge 2 * y = x^2 y,$$

but

$$y * x \wedge 2 = (xy)^2 = x^2 y^2$$

because in the second case, $y * x$ is calculated first. For other operations, the rules of arithmetic apply. Thus,

$$TVC = x * y + z * w * r / c + 1.1 / var$$

is evaluated exactly as it appears. You may not have analytic functions (such as log(.) or cos(.)) of the variables directly in the expression. If you need them, just use **CREATE** to produce them beforehand, then use the created variables in the expression.

Time is entered into the expression by using the name, enclosed in parentheses. As described below, you can use any calculable function of time in the expression. For the simplest case, note that '*time*,' itself is the Lhs variable in the model. A model with a TVC defined as $z(time) = z_1 * time$ might appear as follows:

```
SURVIVAL   ; Lhs = time ; Rhs = x1,x2
           ; TVC = z1 * (time) $
```

Be sure to remember the parentheses! The variable *time*, which is fixed for each observation (at its respective value) is a valid variable in this expression. It is only by including the parentheses in the expression that you insure that $z_i(t)$ is computed as a function of time as it varies and not time for the ith individual. That is, as the partial likelihood is evaluated, at each observation, '(time)' is the value of '*time*' that applies for the specific value for which the risk set is being defined. Recall the partial likelihood is computed over the K distinct exit times in the N observations, t_i , $i = 1,...,K$. With a TVC, $z_i(t)$, the partial likelihood becomes

$$\log L = \sum_{j=1}^{K} \boldsymbol{\beta}'\mathbf{x}_j + \gamma z_j(t) - \log\left[\sum_{i \in R_j} \exp(\boldsymbol{\beta}'\mathbf{x}_i + \gamma z_i(t))\right].$$

You can also enter any function of time by computing the variable using the **CREATE** command to obtain the function of time you want. (*LIMDEP* will keep track and use the correct values in computing $z_i(t)$.) For example, suppose instead of time itself, you wish to use

$$z_i(t) \;=\; z_{1i} \times \exp(-0.01\times time)$$

for the effect of a covariate whose influence fades over time. The commands could be

```
CREATE      ; time1 = Exp(-.01 * time) $
SURVIVAL    ; Lhs = time ; Rhs = x1,x2
            ; TVC = x1 * (time1) $
```

There are several other operations which can be used. These are all of the form

$$result \;=\; x \text{ operator } y,$$

where x, because of the left to right rule, may already be a function of one or more other variables. The operations are:

$x\ @\ y = e^{xy}$,

$x\ !\ y \;= \max(x,y)$ [note, $x!y!w... = \max(x,y,w...)$],

$x\ \#\ y \;= \min(x,y)$,

$x\ \&\ y \;= 1$ if $x > y$ and 0 else,

$x\ \%\ y = \max(x\text{-}y,0)$ (useful for splines),

$x\ \_\ y \;= \min(x\text{-}y,0)$.

The first of these would allow you to specify the TVC shown in the earlier example without having to create it. We could have used

$$z(i,t) \;=\; z_1 * \exp(-.01\times time) = -.01\ @\ (time) * z_1.$$

(Because of the left to right rule, z_1 must appear last, not first.)

With the dummy variable operator, '&,' you have some limited capability for 'time *varying* covariates,' that is covariates that vary, perhaps discretely, over time. For example, consider creating the variable

$$z(i,t) \;=\; age \text{ if } t > T \text{ and 0 otherwise},$$

where T is some threshold. You could obtain this with

; TVC = (time) & t * age

The other logical operators, '%' and '_' give some additional possibilities, but they are fairly limited. In particular, although this allows discrete jumps at points in time, it cannot be computed for the specific individual; it is computed as the same function of time for all individuals. Once again, direct handling of true time varying covariates is accomplished with the parametric models described in the next chapter.

You may also add an 'IF' sort of construction to the TVC specification. The syntax of the conditional TVC specification is

; TVC = alternative value: [condition] expression

where:

'*expression*' is exactly as shown above.

'*alternate value*:' (with its trailing colon) is the value to give the TVC if the condition is false. This is optional. If you do not provide this, the TVC equals zero if the condition is false.

'[*condition*]' (enclosed in brackets) is a logical condition which is evaluated to determine whether or not to set the TVC equal to the value of the expression.

The alternate value may be any of:

- a number, e.g., 1.234,
- a calculator scalar, e.g., *rho*,
- any variable existing in the data set, in which case the value for that individual is used,
- functions of (*time*) enclosed in parentheses to indicate that this is the value obtained as we move through the risk set for this observation.

The [condition] is a logical expression of the form

entity relation entity +/& entity relation entity ...

Entities may be any of those listed above. Relations are >, >=, <, <=, =, and #. Use '+' for 'OR' and '&' for 'AND.' This may be as involved as you like, but the compiler will run out of space if the number of operations (relation or +/&) exceeds 10 in any TVC. For examples:

Set TVC = 1 if (*time*) is greater than or equal *warranty* and 0 otherwise.

; TVC = [(time) >= warranty] 1

Set TVC = (*time*) if (*time*) is less than *warranty*, and exp(-.01(*time*)) otherwise.

; TVC = (time) : [(time) >= warranty] -.01@ (time)

Set TVC = 1 if (*time*) >= 12 and (*time*) < *age* or if (*time*) < *retire*, and 0 otherwise.

; TVC = [(time) >= 12 & (time) < age + (time) < retire] 1

An error occurs and estimation is halted if a TVC cannot be calculated for any observation. For example, '(*time*)@1' = exp(time) will cause an overflow error if (*time*) exceeds 308. The observation and sequence number of the TVC are given at the point at which the error occurs. It is not possible to anticipate such conditions; they will only be found during estimation.

E59.2.5 Stratification

The sample may be stratified with separate baseline hazard functions for each stratum by specifying a stratification variable. This is done as follows:

; Str = stratification variable

It is assumed that this variable is coded 1,2,3,... The expanded specification given below can be used if some other scheme is desired. But, it is important to be sure that this variable does not contain zeros. Instead of stratifying on the values of a variable, you might wish to stratify based on limit values. Specify the variable as above and add the **; Limits** specification as shown below. For example, suppose the stratification is based on weight, with separate strata for the following classes:

1 = weight # 150 pounds,
2 = 150 < weight # 200,
3 = weight > 200.

We use

```
SURVIVAL    ; Lhs = time ; Rhs = ...
            ; Str = weight
            ; Limits = 150,200 $
```

Note that only two limits need to be specified. The number of strata will be one more than the number of limits you give since you need not give the extreme end values. The strata in this formulation are always defined as 'greater than lower limit and less than or equal to upper limit.' The low end of the first class is always negative infinity, and the highest limit is plus infinity.

E59.2.6 Cox Model with Fixed Effects

Suppose the data can be divided into G groups, possibly strata, for example. The model for stratification assumes that the model is the same in all strata, but the risk set and partial likelihood are recomputed for each stratum. A type of fixed effects model would allow variation of the model itself across the groups. A fixed effects approach, for example would be

$$\text{Prob(exit at time } T_j) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_j + \alpha_g d_{j,g})}{\sum_{i \in R_j} \exp(\boldsymbol{\beta}'\mathbf{x}_i + \alpha_g d_{i,g})}$$

Thus, the probabilities shift based on which group the individual is in. But, the risk set is computed as usual as if there were no stratification. This model assumes that the baseline hazard is

$$h_g(t_i,0) = \gamma_g h(t_i,0)$$

where $\gamma_i = \exp(\alpha_i)$. Since the baseline hazards are not estimated, their scale is unknown. As such, the individual group effects must be normalized, which we do by setting $\alpha_G = 0$. After estimation,

$$\gamma_g = \exp(\alpha_g) \;/\; \Sigma_g \exp(\alpha_g).$$

The model with stratification and the one with fixed effects differ in the treatment of the grouping in the data. The log likelihood, neglecting ties, for the stratification case is

$$\log L = \sum_{s=strata} \left\{ \sum_{j=1}^{K_s} \boldsymbol{\beta}'\mathbf{x}_{j,s} - \log\left[\sum_{i \in R_{j,s}} \exp(\boldsymbol{\beta}'\mathbf{x}_{i,s})\right]\right\}.$$

For the fixed effects case, it is

$$\log L = \sum_{s=strata} \sum_{j=1}^{K_s} \boldsymbol{\beta}'\mathbf{x}_{js} + \gamma_s d_{js} - \log\left[\sum_{i \in R_j} \exp(\boldsymbol{\beta}'\mathbf{x}_{i,s} + \gamma_s d_{is})\right].$$

The risk set in the stratification case is composed only of the individuals in the stratum. In the fixed effects case, the risk set is composed of the entire sample.

To request this estimator, it is necessary to have a group variable, which will appear exactly the same as a stratification variable. Then,

```
SURVIVAL    ; Lhs = duration ; Rhs = the desired variables
            ; Str = the group variable ; Fixed Effects $
```

This estimator is limited to 150 - K groups. The parameters are estimated by creating the dummy variables and augmenting the model. An example appears below.

E59.2.7 Output from the Proportional Hazards Model

Initial output from this estimator contains a tally of the number of observations, number of distinct exit times, number of censored observations, and number of observations which exited (were not censored). If any TVCs have been specified, the specification is echoed in the initial output.

After the iterations end, the report includes the partial log likelihood and the value of the partial log likelihood evaluated at the starting values. If you do not provide starting values, these are zero. A chi squared test of the hypothesis that the coefficients equal the starting values is given next. The log-rank test is a Lagrange multiplier test of the same hypothesis. Tabulated output includes estimates, standard errors and descriptive statistics for the regressors.

A listing of the estimates of 10 points from the survival distribution and integrated hazard function is given. Finally, the estimated survival function and integrated hazard (negative log-survival) function are plotted at the means of the regressors and at any additional points that you have specified.

Results saved automatically by this procedure are only scalar *logl*, matrices *b* and *varb*, and *Last Model* labels *b_variables*. If your model included TVCs, the additional labels would be *tvc*1, *tvc*2, ...

E59.2.8 Applications of the Proportional Hazards Model

To illustrate the technique, we apply Cox's proportional hazard model to the data used in the previous chapter.

```
READ        ; Nobs = 22 ; Nvar = 4
            ; Names = time,status,sex,married ; By Variables $

  11  3 19 32 2 14 8 21 16 5 2 8 14 18 18 21 10 1 9 23 19 7
   1  1  0  1 1  1 1  1  0 1 1 1  1  1  1  1  0 1 0  1  1 1
   0  0  1  0 1  0 1  1  1 0 0 1  1  0  1  1  0 0 1  1  0 1
   1  1  2  2 1  1 1  1  2 2 2 1  1  1  2  1  2 2 1  1  2 1
```

This sets up a matrix for plotting the survival function. The matrix command loads two rows in the matrix, so there will be two additional plots of the survival function. The models have censoring, stratification, a fixed coefficient, and a time varying covariate, respectively.

```
MATRIX      ; mf = [0 / 1] $
SURVIVAL    ; Lhs = time,status ; Rhs = sex ; Plot = MF $
SURVIVAL    ; Lhs = time,status ; Rhs = sex ; Str = married $
SURVIVAL    ; Lhs = time,status ; Rhs = sex,married ; Rst = b1 , 0.01 $
SURVIVAL    ; Lhs = time,status ; Rhs = married ; TVC = -.1*(time)*sex $
```

```
+----------------------------------------------------+
| Cox Proportional Hazard Model                      |
| Duration variable is                    TIME       |
| Status is given by variable             STATUS     |
| Total Number of Observations          =    22      |
| Total Number of Observations Exiting  =    18      |
| Total Number of Observations Censored =     4      |
| Total Number of Distinct Exit Times   =    13      |
| Number of Observed Times Incl. Cnsrd. =    16      |
+----------------------------------------------------+
-----------------------------------------------------------------------------
Cox Proportional Hazard Model
Dependent variable                  TIME
Log likelihood function        -39.79330
Restricted log likelihood      -40.09036
Log-rank test with  1 degrees of freedom:
Chi-squared =     .609, Prob  =    .4353
--------+--------------------------------------------------------------------
        |                 Standard             Prob.      95% Confidence
    TIME|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     SEX|    -.38029         .49028     -.78  .4379    -1.34122    .58063
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

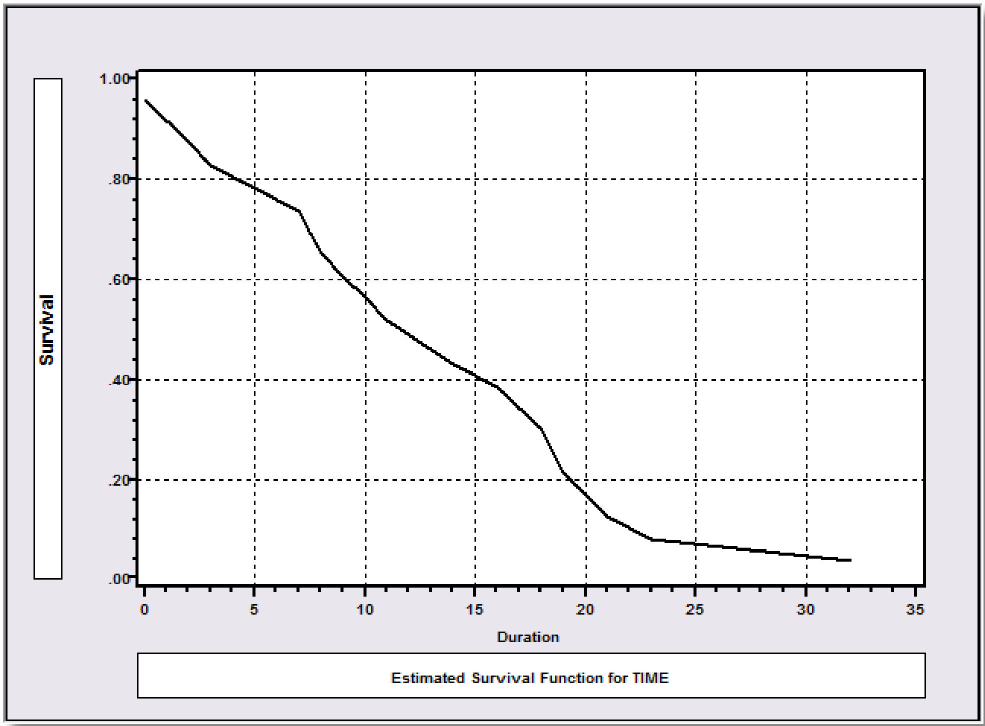


Figure E59.1 Estimated Survival Function

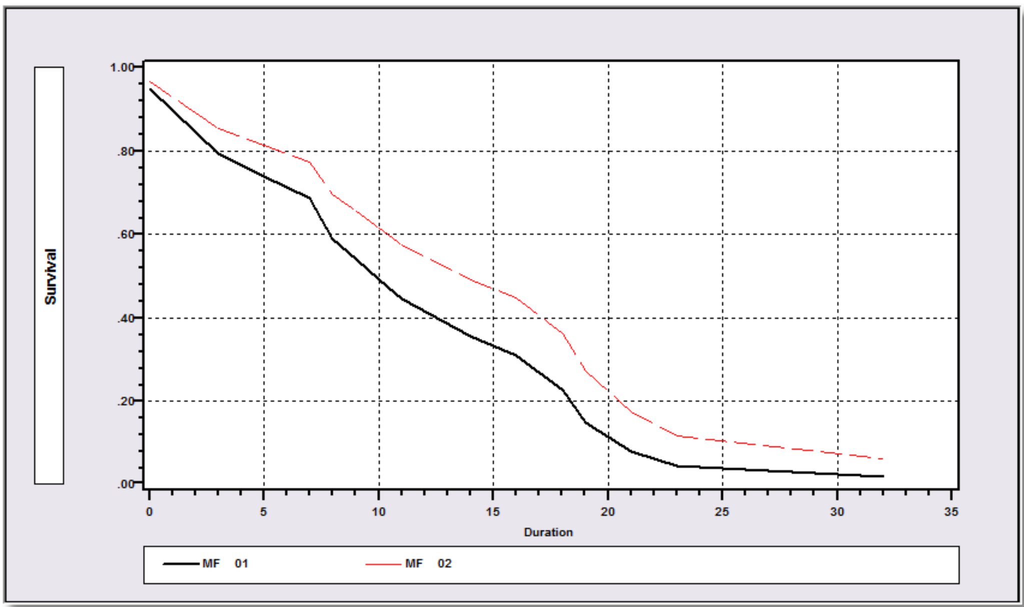


Figure E59.2 Comparison of Two Survival Functions

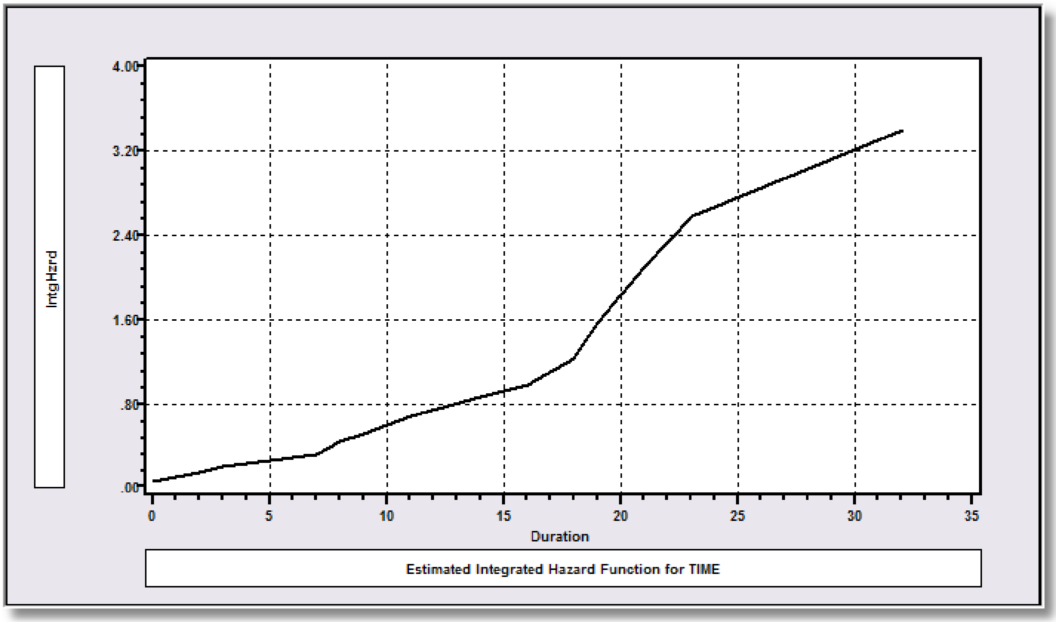


Figure E59.3 Estimated Integrated Hazard Function

```
+-----------------------------------------------------------------+
| Cox Proportional Hazard Model                                   |
| Duration variable is                       TIME                 |
| Status is given by variable                STATUS               |
| Total Number of Observations             =     22               |
| Total Number of Observations Exiting     =     18               |
| Total Number of Observations Censored =         4               |
| Total Number of Distinct Exit Times      =     13               |
| Number of Observed Times Incl. Cnsrd. =        16               |
| Stratification is based on                 MARRIED              |
| Stratum   Lower Limit  Upper Limit  Observations  Proportion    |
|    1         .0000        1.000          13.          .5909     |
|    2         1.000        2.000           9.          .4091     |
| (Range: greater than lower and less than or equal to upper.)    |
+-----------------------------------------------------------------+
-----------------------------------------------------------------------------
Cox Proportional Hazard Model
Dependent variable                  TIME
Log likelihood function        -29.16864
Log-rank test with  1 degrees of freedom:
Chi-squared =     1.436, Prob  =     .2308
--------+--------------------------------------------------------------------
        |                  Standard             Prob.       95% Confidence
    TIME|   Coefficient      Error        z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     SEX|     -.66598         .56356    -1.18  .2373     -1.77054    .43858
--------+--------------------------------------------------------------------
```

```
+----------------------------------------------------+
| Cox Proportional Hazard Model                      |
| Duration variable is                      TIME     |
| Status is given by variable               STATUS   |
| Total Number of Observations            =    22    |
| Total Number of Observations Exiting    =    18    |
| Total Number of Observations Censored =       4    |
| Total Number of Distinct Exit Times     =    13    |
| Number of Observed Times Incl. Cnsrd. =      16    |
+----------------------------------------------------+
-----------------------------------------------------------------------------
Cox Proportional Hazard Model
Dependent variable                  TIME
Log likelihood function        -39.81778
Log-rank test with  2 degrees of freedom:
Chi-squared =    2.347, Prob  =    .3093
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    TIME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     SEX|    -.37656         .49024     -.77  .4424    -1.33742    .58429
 MARRIED|     .01000    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------

+----------------------------------------------------+
| Cox Proportional Hazard Model                      |
| Duration variable is                      TIME     |
| Status is given by variable               STATUS   |
| Total Number of Observations            =    22    |
| Total Number of Observations Exiting    =    18    |
| Total Number of Observations Censored =       4    |
| Total Number of Distinct Exit Times     =    13    |
| Number of Observed Times Incl. Cnsrd. =      16    |
| Total Number of time dependent covariates= 1       |
|  1.  -.1*(TIME)*SEX                                |
+----------------------------------------------------+
-----------------------------------------------------------------------------
Cox Proportional Hazard Model
Dependent variable                  TIME
Log likelihood function        -38.91896
Log-rank test with  2 degrees of freedom:
Chi-squared =    2.312, Prob  =    .3148
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    TIME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
 MARRIED|    -.73891         .57805    -1.28  .2012    -1.87186    .39405
T.V.C.-1|    -.00645         .00533    -1.21  .2263     -.01688    .00399
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

To illustrate the fixed effects estimator, we have arbitrarily divided the 22 observations into six unequal sized groups with the variable group, which is defined by

$$group = [1,1,1,2,2,2,2,3,3,4,4,4,4,5,5,5,5,6,6,6,6,6]$$

The command and results are, then,

SURVIVAL ; Lhs = time,status ; Rhs = married,sex
; Str = group ; Fixed $

```
+-----------------------------------------------------+
| Cox Proportional Hazard Model                       |
| Duration variable is                    TIME        |
| Status is given by variable             STATUS      |
| Total Number of Observations          =    22       |
| Total Number of Observations Exiting  =    18       |
| Total Number of Observations Censored =     4       |
| Total Number of Distinct Exit Times   =    13       |
| Number of Observed Times Incl. Cnsrd. =    16       |
+-----------------------------------------------------+
-----------------------------------------------------------------------------
Cox Proportional Hazard Model
Dependent variable                  TIME
Log likelihood function       -36.37568
Log-rank test with  8 degrees of freedom:
Chi-squared =    7.248, Prob  =    .5102
Model includes group fixed effects
Mean and Variance = 1.0 and  .8546
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    TIME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
 MARRIED|    -.87065         .64656    -1.35  .1781    -2.13788    .39658
     SEX|    -.95204         .67628    -1.41  .1592    -2.27752    .37344
        |Model for group specific effects
GROU=001|     .91582         .63922     1.43  .1519     -.33703   2.16868
GROU=002|     .54904         .39406     1.39  .1635     -.22331   1.32138
GROU=003|     .39708         .41032      .97  .3332     -.40713   1.20129
GROU=004|    2.84684***     1.02810     2.77  .0056      .83179   4.86188
GROU=005|     .50994         .33666     1.51  .1298     -.14989   1.16978
GROU=006|     .78128*        .44281     1.76  .0777     -.08662   1.64917
--------+--------------------------------------------------------------------
```

E59.2.9 Cox Model with Time Varying Covariates

The Cox proportional hazard (semiparametric) model may be modified to allow time varying covariates. In order to construct this form, the estimator requires one record of data for each interval within which the covariates are constant, and T records in total for each person, if covariates change T - 1 times (altogether – more than one covariate may be changing). The command changes as follows:

SURVIVAL ; Lhs = time
; Rhs = the list of covariates
; Entry = a variable which gives t0, the time of the beginning of the interval
; Pds = the number of records $

The interval described by a particular data record are interval (*t*0 to *t*1) measured by *T* to *time*, i.e., 'entry' to 'Lhs.' Note that in records after the first, *t*0 will be *time* on the previous record. All records but the last are treated as censored. The last may be also, in which you would also include a censoring indicator as a second Lhs variable, as usual. Note, also, that this is the same setup that is currently used in the TVC versions of the parametric survival models. Other options available with this estimator include all previous features, as well as

; Robust to request the sandwich estimator
; Cluster = ... specification
; Wts = a weighting variable

The Cox model also now creates a matrix named *cox_bsln* with five columns containing

1. exit times,
2. baseline survival rates,
3. hazard functions,
4. cumulative hazard functions,
5. integrated hazard function = -log(survival function).

The Cox model may be fit with group 'fixed effects' by specifying

; Str = the group identifier
; Fixed Effects

(In other treatments, this is labeled a 'frailty' model. That is probably not appropriate here, as the usual random effects 'frailty model' is not identified in this context.) With this option in use, the estimated effects are renormalized to have mean 1.0 while the variance is left unrestricted.

The output results reported for this model may be modified to include 'hazard ratios,' which are, for a specific coefficient b_k, equal to exp(b_k). Add

; Hazard Ratios

to the command to request this treatment.

E59.3 The Ordered Extreme Value Model

Han and Hausman (1988) have devised a semiparametric estimator for the proportional hazards model. They cite three virtues:

1. It is suited to discrete data.
2. It is unhindered by large numbers of ties.
3. It circumvents problems associated with heterogeneity.

In addition, they argue that an advantage of the technique is that the parameters of the covariates are invariant to the length of time intervals chosen. As such, the grid of intervals, which need not be of equal length, can be made finer as the sample size increases.

The hazard rate is

$$\lambda_i(\tau) = \lim_{\Delta \to 0} \frac{\text{Prob}[\tau < t_i < \tau + \Delta]}{\Delta} = \lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)\ .$$

They specify this as

$$\log \int_0^{t_i} \lambda_0(\tau)d\tau = \boldsymbol{\beta}'\mathbf{x}_\text{i} + \varepsilon_\text{i},$$

where $F(\varepsilon_i) = \exp(-\exp(-\varepsilon_i))$ (extreme value, or Gompertz).

Let $\log \int_0^{t} \lambda_0(\tau)d\tau = R_t,\ t = 1,...,T.$

The probability of failure in period t by individual i is

$$\text{Prob}[T_{t\text{-}1} < t_i < T_t] = \int_{l_{t-1}-\boldsymbol{\beta}'\mathbf{x}_i}^{l_t-\boldsymbol{\beta}'\mathbf{x}_i} f(\varepsilon)d\varepsilon\ .$$

The logs of the integrated baseline hazards, R_t are treated as unknown parameters. (The authors observe that Cox's proportional hazard model treats them as nuisance parameters and conditions them out of the likelihood function.) Let $y_i = t - 1$ if t_i falls in interval t. Then, the probability defined above, with the extreme value distribution for ε, defines exactly the ordered probability model described in Chapter E58 with an extreme value (Gompertz) probability model. The ls in the present context would be the μs in the ordered probability model discussed previously.

To estimate this model, therefore, it is necessary only to code the dependent variable appropriately and submit it with the **ORDERED PROBABILITY** (or just **ORDERED**) command. The data are assumed to be generated as observations on duration in intervals

$$t = \begin{matrix} 0 & T_1 & T_2 & T_3 & \cdots & T_J \\ 0 & 1 & 2 & 3 & \cdots & J \end{matrix}$$

The lower row shows the values taken by what will be the Lhs variable in the model. The *time* variable in the data must be recoded to conform to the preceding layout. **RECODE** may be used if necessary. The threshold values μ_0, μ_1, ... are then interpreted as the logs of the integrated baseline hazard functions. This may be estimated as an ordinary ordered Gompertz model, with up to 50 values (J = 49) taken by the Lhs variable. At the end of the estimation, *LIMDEP* computes estimates of the hazard rate at the means of the regressors by the computation

$$h(t) = \text{Prob}[t_j < t < t_j{+}1] \,/\, \text{Prob}(t \geq t_j).$$

This is computed by using the predicted cell probabilities for the ordered logit model at the means of the covariates. These probabilities are divided by the interval width if values are provided that allow these to be calculated.

The model command is simply

```
ORDERED     ; Lhs = ...
            ; Rhs = ...
            ; Hazard
            ; Model = Gompertz $
```

By this formulation, the intervals are assumed to be one period in length. The specification

; Endpoints = $\mathbf{T_1,T_2,...,T_J}$

can be used to provide the interior endpoints of the intervals. The authors discuss using the ordered probit model instead of the Gompertz model. This is a bit ambiguous, however. Nonetheless, the hazard rates are computed using whichever distribution has been used to fit the model.

If desired, the Rhs may contain only a constant term, *one*. That is, it is not necessary to have covariates in the model. This produces a semiparametric alternative to the Kaplan-Meier estimator of the previous chapter. The program first estimates the ordered probability model. All results saved are the same as the ordered probability model discussed earlier, except that the matrix *mu* which is normally saved for the ordered Gompertz model is now replaced with a matrix named *hazard* which contains the estimated hazard rates. There is one hazard rate computed for each interval. The last one is assumed to be the same as the second to last one. Suppose your dependent variable takes values 0,1,...,7. This is eight values, and eight hazard rates will be computed. You can then plot the hazard rates against the left endpoints of the intervals, which can be defined separately. An example is given below.

Application

We apply Han and Hausman's technique to the Kennan strike data used earlier. Since there are no covariates, the estimated hazard function compares directly to the one computed earlier.

```
READ      ; Nobs = 62 ; Nvar = 1 ; Names = t ; By Variables $

  1  2  2  2  2  3  3  3  9 12 21 27 43 49 52
  3  4  5  7  8  9 10 11 12 13 14 15
 17 19 21 22 23 25 26 27 28 29 32 33
 35 37 38 41 42 43 44 49 52 61 72 80
 80 80 80 80 80 80 80 80 80 80 80

CREATE    ; yt = t $
RECODE    ; yt
          ;  0 / 4  = 0  ; 5  / 10 = 1 ; 11 / 13 = 2 ; 14 / 17 = 3
          ; 18 / 23 = 4  ; 24 / 28 = 5 ; 29 / 40 = 6 ; 41 / 60 = 7
          ; 61 / 80 = 8 $
MATRIX    ; endt = [1,5,11,14,18,24,29,41,61] $
ORDERED   ; Lhs = yt ; Rhs = one
          ; Model = Gompertz
          ; Hazard
          ; Endpoints = endt $
MPLOT     ; Lhs = endt ; Rhs = hazard
          ; Fill
          ; Grid
          ; Title = Estimated Gompertz Hazard Function $
```

```
-----------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                     YT
Log likelihood function      -129.69815
Underlying probabilities based on Gompertz
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      YT|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .60133***     .14017     4.29  .0000      .32661    .87605
        |Threshold parameters for index
   Mu(1)|     .29787***     .10668     2.79  .0052      .08878    .50696
   Mu(2)|     .47788***     .12574     3.80  .0001      .23142    .72433
   Mu(3)|     .60973***     .13711     4.45  .0000      .34100    .87846
   Mu(4)|     .83083***     .15420     5.39  .0000      .52860   1.13307
   Mu(5)|    1.06237***     .17229     6.17  .0000      .72470   1.40005
   Mu(6)|    1.37013***     .19818     6.91  .0000      .98170   1.75855
   Mu(7)|    1.96417***     .26111     7.52  .0000     1.45241   2.47594
--------+--------------------------------------------------------------------
```

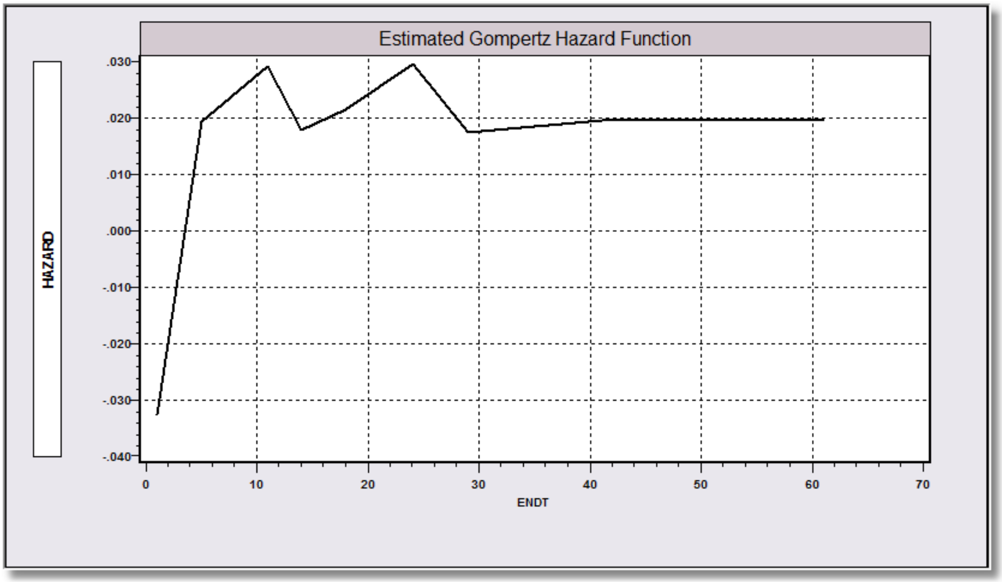


Figure E59.4 Han-Hausman Estimated Gompertz Hazard Function

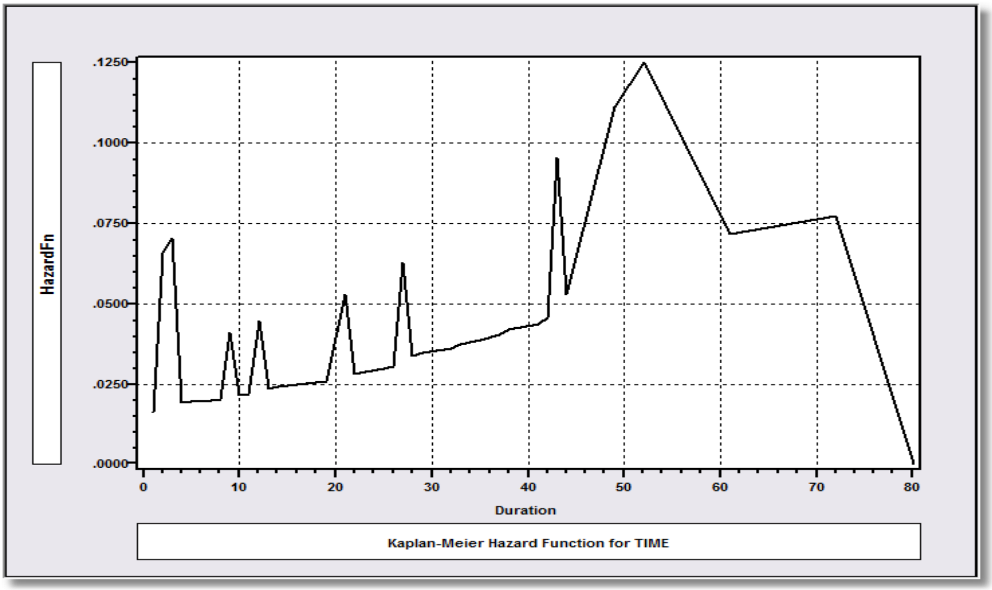


Figure E59.5 Kaplan-Meier Estimated Hazard Function

E60: Parametric Models for Duration

E60.1 Introduction

The models discussed in this chapter embody specific assumptions about the distribution of duration or failure times. *LIMDEP* includes many of the models which have been proposed in the literature. (See, e.g., Kalbfleisch and Prentice (1980), Lancaster (1990) and Cox and Oakes (1984).) This chapter presents a large number of variations of the models. The choice among the various models is sometimes made on the basis of the shape of the hazard function. As shown below, this can vary widely. In order to arrange the material in a convenient fashion, the basic formulations are first presented in full. Some of the more esoteric extensions are collected in later sections. We begin with basic models for duration, without covariates or heterogeneity. Later sections will extend the models.

The essential model command for estimating parametric survival models is

SURVIVAL ; Lhs = log of time variable [, censoring indicator (optional)]
; Rhs = covariates
; Model = type $

Type is one of Weibull, exponential, normal, loglogistic, inverse Gaussian, gamma, F, or Gompertz. Plots of hazard functions, integrated hazards and survival functions may also be requested. 'Residuals' in this model are the integrated hazard function. 'Fitted values' are the estimated hazard function values. Other specifications which may each (alone) modify the basic model include: latent heterogeneity in the location of the distribution, time varying covariates, panel data estimators, (fixed effects, random effects, random parameters, and latent class), split population models, truncation, variance heterogeneity, and sample selection. These extensions are treated in Chapter E61.

E60.2 Parametric Models for Survival Data

We denote by '*t*' the nonnegative random variable 'time until transition' (using Lancaster's term). In many familiar applications, *t* is time until failure, which produces the term 'failure time models.' But, other applications, for example, strike duration, involve time until recovery from a disease, elapsed time until a merger takes place or length of a spell of unemployment. In each of these, Lancaster's term seems more appropriate than 'failure time' so we will use this.

Parametric models may be defined in terms of the density, $f(t)$, the survival function, $S(t)$ = Prob[$T \geq t$], or the hazard function, $h(t) = f(t)/S(t)$. Note that $f(t) = -\text{d}S(t)/\text{d}t$ and $h(t) = -\text{dlog}S(t)/\text{d}t$. In *LIMDEP*'s set of specifications, each model is characterized, at minimum, by a positive location or rate parameter, λ, and a positive scale parameter, p. The parametric distributions supported by this program in *LIMDEP* are listed in Table E60.1.

The survival function for the gamma model must be written in implicit form because of the incomplete gamma integral. The hazard rate is likewise complicated. Note that the gamma model is an encompassing model for the Weibull and exponential models. If θ equals one, the Weibull model results. If θ and p both equal one, the exponential model results. As described below, the generalized *F* model is even more broad, as it encompasses all the above save for the Gompertz and inverse Gaussian models.

Authors differ on how useful the parametric models are. All assume a specific functional form, which is not necessarily good. On the other hand, all save for the exponential model allow for both positive and negative duration dependence, and all are fairly flexible functional forms. Moreover, some of the options discussed below, such as variance heterogeneity, latent heterogeneity and the presence of covariates, allow you a large amount of room to accommodate different patterns in the data that cannot be accommodated with the semiparametric or Kaplan-Meier approaches.

Distribution	**Density and Survival**	**Hazard Function**
exponential	$\lambda\exp(-\lambda t)$ $\exp(-\lambda t)$	λ
Weibull	$\lambda p(\lambda t)^{p-1}\exp[-(\lambda t)^p]$ $\exp[-(\lambda t)^p]$	$\lambda p(\lambda t)^{p-1}$
lognormal	$[p/(\lambda t)]\phi(-p\log(\lambda t))$ $\Phi(-p\log(\lambda t))$	$\phi(-p\log(\lambda t))/\Phi(-p\log(\lambda t))$
loglogistic	$\dfrac{\lambda p(\lambda t)^{p-1}}{[1+(\lambda t)^p]^2}$ $\dfrac{1}{[1+(\lambda t)^p]}$	$\dfrac{\lambda p(\lambda t)^{p-1}}{[1+(\lambda t)^p]}$
gamma	$\dfrac{(\lambda p)(\lambda t)^{P\theta-1}}{\Gamma(\theta)}\exp(-(\lambda t)^p)$ no closed form	no closed form
Gompertz	$p\exp(\lambda t)\exp\{(-p/\lambda)[\exp(\lambda t)-1]\}$ $\exp\{(-p/\lambda)[\exp(\lambda t)-1)]\}$	$p\exp(\lambda t)$
inverse Gaussian	$\phi\left[-(\lambda t-p)/\sqrt{t}\right]\dfrac{p}{t^{1.5}}$ $\Phi\left[-(\lambda t-p)/\sqrt{t}\right](1-\exp(2\lambda p)$	$\dfrac{\phi\left[-(\lambda t-p)/\sqrt{t}\right]}{\Phi\left[-(\lambda t-p)/\sqrt{t}\right]}\dfrac{p}{t^{1.5}(1-\exp(-2\lambda p))}$
generalized F	see below	
Weibull or exponential (p = 1) with gamma heterogeneity	$\{S(t)=\exp[-(\lambda t)^p]\}$ $\left[S(t)^{\theta+1}\right]\lambda p(\lambda t)^{p-1}\left[1+\theta(\lambda t)^p\right]^{-1/\theta}$ $\left[S(t)^{\theta}\right]\lambda p(\lambda t)^{p-1}$	$S(t)\left[1+\theta(\lambda t)^p\right]^{-1/\theta}$

Table E60.1 Parametric Survival Models

E60.2.1 Loglinear Models and Estimation Strategies

For the first five models listed above, estimation is facilitated by the transformation,

$$w = (\log t - \beta) / \sigma,$$

where $\lambda = e^{-\beta}$ and $p = 1/\sigma$.

With this change of variable, the densities and survival functions for w for the five distributions are as listed in Table E60.2.

Distribution	Density	Survival Function
Weibull:	$\exp(w - e^{w})$	$\exp(-e^{w})$
exponential:	$\exp(-e^{w})$	e^{-w}
lognormal:	$\phi(w)$	$\Phi(-w)$
loglogistic:	$e^{w}(1 + e^{w})^{-2}$	$1/(1 + e^{w})$
gamma:	$\exp(\theta w - e^{w} - \log\theta)$	$1- \gamma(\theta,e^{w})$, $\gamma(\theta,t)$ = incomplete gamma integral

Table E60.2 Loglinear Survival Models for Transformed Variables

The Gompertz model, $S(t) = \exp((-p/\lambda)(e^{\lambda t}-1))$ and $h(t) = pe^{\lambda t}$, is not loglinear, so we adopt a somewhat different estimation strategy. The inverse Gaussian model is log linear, but the transformation above does not produce a convenient functional form to use for optimization. The inverse Gaussian survival model is estimated as a particular form of the general loglinear model – see Chapter E55 for details.

The generalized F model, like the inverse Gaussian model, is not easily transformed to a simple functional form. Let $z = (\lambda t)^{p}$. Assume that z has a central F distribution with degrees of freedom parameters $2M1$ and $2M2$. ($M1$ and $M2$ need not be integers.) By the change of variable technique, the density of t is

$$f(t) = [(\lambda p)(\lambda t)^{(p-1)} / \beta(M1,M2)]\,[K(t)]^{M1}\{1 + [K(t)]\}^{-(M1+M2)}$$

where $K(t) = (M1/M2)(\lambda t)^{p}$

and $\beta(M1,M2)$ = the beta function, $\Gamma(M1)\Gamma(M2)/\Gamma(M1+M2)$.

The generalized F distribution has four structural parameters, λ, $p = 1/\sigma$, $M1$ and $M2$. The other parametric models have two (λ and p; lognormal, loglogistic, Weibull, Gompertz), one (λ; exponential), three (λ, θ, p; Weibull/heterogeneity), or three (λ, p, γ; gamma), so this is more general than the other models. Also, all of those listed except the Gompertz, mixed Weibull and inverse Gaussian models are special cases of the generalized F listed in Table E60.3.

Distribution	Form of the Generalized F Distribution		
loglogistic:	$M1 = 1$	$M2 = 1$	p free
lognormal:	$M1 \to +\infty$	$M2 \to +\infty$	p free
Weibull:	$M1 = 1$	$M2 \to +\infty$	p free
exponential:	$M1 = 1$	$M2 \to +\infty$	$p = 1$
gamma:	$M1$ free	$M2 \to +\infty$	p free

Table E60.3 Special Cases of the Generalized F Distribution

The survival and hazard functions do not exist in closed form and must be approximated. The survival function is computed using the CDF of the beta distribution:

$$S(t) = \text{Bds}\,[\,1/(1+K), M2, M1]$$

from which the hazard function may then be estimated. Lancaster (1990) has analyzed this model at length. Among his results is that the generalized F is a gamma weighted mixture of gamma models, which suggests that it can be interpreted as a gamma model with latent gamma distributed heterogeneity. Other models with heterogeneity are detailed below.

E60.2.2 Covariates and Log Likelihood Functions

The effect of external covariates, $\mathbf{x}_i$ on the survival rate or the hazard function can be incorporated by writing

$$\lambda_i = e^{-\boldsymbol{\beta}'\mathbf{x}_i}.$$

(This is labeled the 'accelerated failure time' model.) The model is otherwise the same as before. After transformation, the covariates enter w_i linearly, which, once again, makes estimation relatively simple. This formulation is used in all the models listed above, including those not handled as loglinear.

We note at this point a possible inconsistency in the literature. The formulations shown above correspond to Kalbfleisch and Prentice. Elsewhere, the signs and normalizations of the parameters may be different. For example, in terms of the original models, Kiefer (1988) writes the densities and hazard rates in terms of a 'γ' which would correspond to λ^p in our models. He also reverses the signs on the coefficients in the models. With both of these changes, where we have

$$\log t_i = \boldsymbol{\beta}'\mathbf{x}_i + \sigma w_i$$

Kiefer would have

$$\log t_i = -(1/\sigma)\boldsymbol{\beta}'\mathbf{x}_i - \sigma w_i.$$

For the present, we assume that the covariates, $\mathbf{x}_i$ have been fixed for the individual from time $T = 0$ to $T = t_i$, when we make our observation. Section E60.7 generalizes these models to allow $\mathbf{x}_i$ to evolve as a step function from time zero to the time of observation.

Data on observed transition times may be complete or censored. If an observation is censored, then t_i marks the time, relative to the origin, that the observation was made, not when the transition occurred. There is a presumption, dropped in the split population model, that the transition would occur some time after time t_i (but that for certain it would occur).

For the loglinear models, the likelihood function for N observations on

$$y_i = \log t_i = \sigma w_i + \beta$$

and right censoring indicator δ_i (one for complete observations, zero for censored) is

$$L \;=\; \Pi_i[\sigma^{-1} f(w_i)]^{\delta_i}[S(w_i)]^{1-\delta_i}.$$

Note that $\log L$ may be written

$$\log L \;=\; \Sigma_i\,[\delta_i\,(-\log\sigma + \log h(w_i)) \;+\; \log S(w_i)]$$

where $\quad h_i\,(w_i) \;=\; f(w_i)/S(w_i) \;=\;$ the hazard function.

Log likelihood functions are maximized by BFGS or Newton's method. The choice is discussed below. For discussions of interpretation of the parameters and the distributions, the reader is referred to Kalbfleisch and Prentice (1980), Lancaster (1990) or to Kiefer's (1988) survey.

The log likelihood for the loglinear models is the sum of individual terms of the form

$$\log L_i \;=\; \delta_i \log[f(w_i)/\sigma] \;+\; (1 - \delta_i)\log S(w_i)$$

where $\quad w_i \;=\; (\log t_i - \boldsymbol{\beta}'\mathbf{x}_i)/\sigma.$

The derivatives are: $\quad \partial\log L_i/\partial\boldsymbol{\beta} \;=\; [\delta_i\,\partial\log f/\partial w_i + (1-\delta_i)\partial\log S/\partial w_i\,](-\mathbf{x}_i/\sigma)$

$$\partial\log L_i/\partial\sigma \;=\; [\delta_i\,\partial\log f/\partial w_i + (1-\delta_i)\partial\log \mathrm{S}/\partial w_i\,](-1/\sigma) \;-\; \delta_i/\sigma.$$

Note that $\quad \delta_i \log f(w_i) \;+\; (1 - \delta_i)\log S(w_i) \;=\; \delta_i \log h_i\,(w_i) \;+\; \log S(w_i),$

where $h(w_i)$ is the hazard function. Let the bracketed term in the derivatives be denoted A_i. It follows from the first term that $\Sigma_i\,A_i\,\mathbf{x}_i = \mathbf{0}$ at the maximum of the log likelihood for the sample. Therefore, terms involving A_i times constants not involving w_i will fall out of the second derivatives matrix at the maximum. Making use of this result (and skipping some algebra), we have, at the maximum of the log likelihood,

$$\partial^2\log L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \;=\; -\Sigma_i\,[\delta_i\,\partial^2\log h/\partial w_i^{\,2} + \partial^2\log S/\partial w_i^{\,2}]\mathbf{z}_i\,\mathbf{z}_i' \;+\; \mathbf{K},$$

where $\quad \mathbf{z}_i' \;=\; (1/\sigma)[\mathbf{x}_i', -w_i\,]$

$$\boldsymbol{\theta}' \;=\; [\boldsymbol{\beta}',\sigma]$$

and $\quad \mathbf{K}_{ij} \;=\; \Sigma_i\,\delta_i/\sigma^2$

in the lower right corner and zero everywhere else. We use this matrix as the weighting matrix in Newton's method. This is, therefore, a hybrid of Newton's method and the method of scoring, since we use the exact expectation for one part of the Hessian and estimate the expectation with the mean for the rest of the terms.

What differs from model to model are the bracketed terms in the second derivatives. For the four models for which this procedure is available, these terms denoted by C_i are listed in Table E60.4.

Model	Second Derivative
Weibull and exponential:	$C_i \;=\; \mathrm{e}^{\,w_i},$
loglogistic:	$C_i \;=\; [S(w_i)]^2(1 + \delta_i)\mathrm{e}^{w_i},$
lognormal:	$C_i \;=\; [h(w_i) - \delta_i\,w_i\,][h(w_i) - w_i)] + \delta_i\,.$

Table E60.4 Second Derivatives

The remaining models, mixed Weibull, gamma, generalized F, Gompertz and inverse Gaussian models are estimated using the original log likelihood in terms of observed t_i rather than log t_i. The log likelihood takes the same general form as for the loglinear models. For these cases, the BHHH method is used to estimate the asymptotic covariance matrix of the MLE.

E60.3 Commands for Parametric Duration Models

The model command is

SURVIVAL **; Lhs = logt,delta** (delta is optional and may be omitted)
; Rhs = one
; Model = Weibull $

Weibull may be replaced by **Loglogistic, Exponential, Normal, Gamma, InverseGauss** or **F**. To add additional covariates, simply add them to the Rhs list. (There is no default model. If you do not specify a particular model, the estimator reverts to the Cox model discussed in Chapter E59.) The censoring indicator, *delta*, is optional. The censoring indicator is set up as noted earlier, taking values one for complete observations, zero for censored.

NOTE: The Lhs variable in all models except the Gompertz model is the log of time. You must **CREATE** and use log time for the Lhs variable. The command, itself, may compute the log, as in **; Lhs = log(time)**.

When there are covariates on the Rhs, β in the preceding becomes $\boldsymbol{\beta}'\mathbf{x}$ where $\mathbf{x}$ is the covariate vector, including *one*. The first set of maximum likelihood estimates given is the complete parameter vector, $[\boldsymbol{\beta}',\sigma]$. Since λ now depends on $\mathbf{x}$ $[\lambda=\exp(-\boldsymbol{\beta}'\mathbf{x})]$, we compute it at the means of the variables. This value of λ, p (which is $1/\sigma$), and the median and several percentiles of this distribution are also displayed.

'Predictions' for survival models are computed as follows:

; List requests a listing of

- actual observation on t_i (not logt_i)
- prediction of $t_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$ (Note, in general this is neither the mean nor the median.)
- generalized residual = integrated hazard = $-\log S(t_i)$
- hazard function
- survival function

The survival probabilities or hazard rates may also be retained as new variables in the data set by including

; Keep = name to retain the hazard function

and **; Res = name** to retain the integrated hazard rate.

The preceding provide generic model forms that can be used for the exponential, Weibull, loglogistic, lognormal and inverse Gaussian models. The gamma, Gompertz and generalized F models require specific forms of the model commands. These are given below in Section E60.6.

Standard Model Specifications for the Parametric Duration Models

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps ancillary parameter σ with main parameter $\boldsymbol{\beta}$ vector in *b*.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

; Cluster = spec requests computation of the cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

You may provide your own starting values for the models if you wish. In all cases, the values you provide are parameters = $-\boldsymbol{\beta}$, σ. You may also impose fixed value and equality restrictions with

; Rst = list...

In the exponential model, σ equals one, so you should not include it in the list. Note that during estimation, *LIMDEP* is using the negative of the parameter vector during the iterations.

E60.4 Results for Parametric Models

Ordinary least squares is used to obtain the starting values. The iterations follow, then the maximum likelihood estimates are displayed. The estimates presented first are β and σ. In a subsequent table, estimates of λ, *p*, the median of the distribution, and four percentiles (.25, .50, .75, and .95) of the distribution of time (not log time) with these values of λ and *p* are listed. These are computed at the sample mean values of the covariates.

The CDFs of all models given above save the gamma and inverse Gaussian can be inverted (the lognormal distribution requires an approximation) to obtain percentiles of the distributions. In particular, where α is the probability of surviving to time *t* or longer, these are given in Table E60.5.

Model	Inverse CDF
Weibull	$t = [(-\log\alpha)^{1/p}]/\lambda$
Gompertz	$t = [\log(1.0 - \lambda\log\alpha)/p)]/\lambda$
lognormal	$t = \exp(-\Phi^{-1}(\alpha)/p)]/\lambda$
loglogistic	$t = [((1.0 - \alpha)/\alpha)^{1/p}]/\lambda$
exponential	$t = [-\log\alpha]/\lambda$

Table E60.5 Inverse CDF

The median of the gamma distribution is obtained by inverting the corresponding chi squared distribution (with noninteger degrees of freedom parameters). The median of each distribution above is obtained by setting $\alpha = .5$. This estimate is presented with an estimate of its asymptotic standard error with the earlier estimates of λ and p. The displayed results contains the values of *t* corresponding to $\alpha = .25$, .50, .75, and .95.

When you estimate any of the parametric survival models (Weibull, loglogistic, Gompertz, exponential, normal, inverse Gaussian or F), plots of the survival function, hazard function, and integrated hazard function can be produced by adding the following to your command,

; Plot

Results saved by the loglinear models are:

Matrices: *b* and *varb* contain the estimate of β and the asymptotic covariance matrix. **; Par** adds the ancillary parameters, σ, and for the gamma model, θ to *b* and *varb*.

Scalars: $s = \sigma$,
ybar and *sy* are descriptive statistics for $\log t_i$,
kreg and *nreg* give the dimensions of the estimation problem,
logl contains the log likelihood,
theta is the value of γ for the gamma model, θ for the heterogeneity models.

Last Model: *b_variables* and *sigma*.

Last Function: None

E60.5 Applications

We will illustrate several of the parametric models with Kennan's strike data. These are reported in Greene (2011). The two variables are *t* = duration in days of major strikes in several years and *prod*, a measure of 'unexpected' output in the economy in that year. Note that *prod* is the same for all observations in a given year. The data are listed in Table E60.6.

t	*prod*	*t*	*prod*
7.00000	.0113800	3.00000	.0742700
9.00000	.0113800	10.0000	.0742700
13.0000	.0113800	1.00000	.0645000
14.0000	.0113800	2.00000	.0645000
26.0000	.0113800	3.00000	.0645000
29.0000	.0113800	3.00000	.0645000
52.0000	.0113800	3.00000	.0645000
130.000	.0113800	4.00000	.0645000
9.00000	.0229900	8.00000	.0645000
37.0000	.0229900	11.0000	.0645000
41.0000	.0229900	22.0000	.0645000
49.0000	.0229900	23.0000	.0645000
52.0000	.0229900	27.0000	.0645000
119.000	.0229900	32.0000	.0645000
3.00000	-.0395700	33.0000	.0645000
17.0000	-.0395700	35.0000	.0645000
19.0000	-.0395700	43.0000	.0645000
28.0000	-.0395700	43.0000	.0645000
72.0000	-.0395700	44.0000	.0645000
99.0000	-.0395700	100.000	.0645000
104.000	-.0395700	5.00000	-.104430
114.000	-.0395700	49.0000	-.104430
152.000	-.0395700	2.00000	-.00700000
153.000	-.0395700	12.0000	-.00700000
216.000	-.0395700	12.0000	-.00700000
15.0000	-.0546700	21.0000	-.00700000
61.0000	-.0546700	21.0000	-.00700000
98.0000	-.0546700	27.0000	-.00700000
2.00000	.00535000	38.0000	-.00700000
25.0000	.00535000	42.0000	-.00700000
85.0000	.00535000	117.000	-.00700000

Table E60.6 Kennan (1985) Data on Strike Duration

The following compares four of the model formulations.

```
CREATE    ; logt = Log(t) $
SURVIVAL  ; Lhs = logt ; Rhs = one,prod ; Model = Exponential $
SURVIVAL  ; Lhs = logt ; Rhs = one,prod ; Model = Weibull ; Plot $
SURVIVAL  ; Lhs = logt ; Rhs = one,prod ; Model = Loglogistic $
SURVIVAL  ; Lhs = logt ; Rhs = one,prod ; Model = InverseGauss ; Plot $
```

```
-----------------------------------------------------------------------------
Loglinear survival model: EXPONENTIAL
Dependent variable                   LOGT
Log likelihood function         -97.28844
Estimation based on N =      62, K =   2
Inf.Cr.AIC  =    198.6 AIC/N =    3.203
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    LOGT|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.77651***      .13909    27.15  .0000     3.50390   4.04912
    PROD|   -9.33381***     2.97787    -3.13  .0017   -15.17033  -3.49730
        |Ancillary parameters for survival
   Sigma|        1.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------------------

+---------------------------------------------------------------------------+
| Alternative representations of survival model:                            |
| Accelerated Failure Time:  b(k)  =  beta(k)    (Given above)              |
+---------+-----------------------------+-----------------------------------+
|         | Proportional Hazards        | Hazard Ratios                     |
|         | c(k) = -1/sigma*beta(k)     | h(k)=exp[(-1/sigma)*beta(k)]      |
|         | z  =  c(k)/Std.Err.[c(k)]   | z  = [h(k)-1]/Std.Err.[h(k)]      |
+---------+-----------------------------+-----------------------------------+
|Variable|     c(k)   Std.Err.   |z|    |     h(k)       Std.Err     |z|    |
+---------+-----------------------------+-----------------------------------+
|PROD    |    9.3338    2.9779   3.134 |   11314.1995  33692.1826    .336 |
+---------------------------------------------------------------------------+

+------------------------------------------------------------------+
| Parameters of underlying density at data means:                  |
| Parameter   Estimate   Std. Error     Confidence Interval        |
| ------------------------------------------------------------     |
| Lambda        .02538       .00339       .0187 to       .0320     |
| P            1.00000       .00000      1.0000 to      1.0000     |
| Median      27.30615      3.64417     20.1636 to     34.4487     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25       .50       .75       .95                 |
| Time         54.61     27.31     11.33      2.02                 |
+------------------------------------------------------------------+

-----------------------------------------------------------------------------
Loglinear survival model: WEIBULL
Dependent variable                   LOGT
Log likelihood function         -97.28542
Estimation based on N =      62, K =   3
Inf.Cr.AIC  =    200.6 AIC/N =    3.235
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.77977***      .13833    27.32  .0000     3.50865   4.05090
    PROD|   -9.33220***     2.95428    -3.16  .0016   -15.12249  -3.54191
        |Ancillary parameters for survival
   Sigma|     .99220***      .12064     8.22  .0000      .75576   1.22865
-------------------------------------------------------------------------------
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+--------------------------------------------------------------------------+
| Alternative representations of survival model:                           |
| Accelerated Failure Time:  b(k)  =  beta(k)    (Given above)             |
+---------+-----------------------------+----------------------------------+
|         | Proportional Hazards        | Hazard Ratios                    |
|         | c(k) = -1/sigma*beta(k)     | h(k)=exp[(-1/sigma)*beta(k)]     |
|         | z  =  c(k)/Std.Err.[c(k)]   | z  = [h(k)-1]/Std.Err.[h(k)]     |
+---------+-----------------------------+----------------------------------+
|Variable|     c(k)   Std.Err.   |z|   |     h(k)       Std.Err     |z|   |
+---------+-----------------------------+----------------------------------+
|PROD    |    9.4055    3.3030   2.848 |   12155.3802  40149.5959    .303 |
+--------------------------------------------------------------------------+

+-----------------------------------------------------------------+
| Parameters of underlying density at data means:                 |
| Parameter   Estimate   Std. Error     Confidence Interval       |
| ------------------------------------------------------------    |
| Lambda         .02530       .00337       .0187 to       .0319   |
| P             1.00786       .12254       .7677 to      1.2480   |
| Median       27.47425      3.66307     20.2946 to     34.6539   |
| Percentiles  of  survival  distribution:                        |
| Survival       .25        .50        .75        .95             |
| Time          54.65     27.47      11.48       2.07             |
+-----------------------------------------------------------------+

-----------------------------------------------------------------------------
Loglinear survival model: LOGISTIC
Dependent variable                    LOGT
Log likelihood function         -101.34034
Estimation based on N =      62, K =   3
Inf.Cr.AIC  =    208.7 AIC/N =    3.366
Model estimated: Aug 11, 2011, 11:01:41
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.29228***      .16511    19.94  .0000     2.96866   3.61589
    PROD|   -9.54360***     3.07860    -3.10  .0019   -15.57755  -3.50964
        |Ancillary parameters for survival
   Sigma|     .70847***      .09732     7.28  .0000      .51773    .89921
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------------+
| Alternative representations of survival model:                             |
| Accelerated Failure Time:  b(k)  =  beta(k)    (Given above)               |
+--------+------------------------------+------------------------------------+
|        | Proportional Hazards         | Hazard Ratios                      |
|        | c(k) = -1/sigma*beta(k)      | h(k)=exp[(-1/sigma)*beta(k)]       |
|        | z  =  c(k)/Std.Err.[c(k)]    | z  = [h(k)-1]/Std.Err.[h(k)]       |
+--------+------------------------------+------------------------------------+
|Variable|     c(k)   Std.Err.   |z|    |     h(k)       Std.Err     |z|    |
+--------+------------------------------+------------------------------------+
|PROD    |   13.4707    4.7901   2.812 |   708368.36  3389322.32     .209 |
+----------------------------------------------------------------------------+
```

```
+------------------------------------------------------------------+
| Parameters of underlying density at data means:                  |
| Parameter   Estimate   Std. Error     Confidence Interval        |
| ------------------------------------------------------------     |
| Lambda        .04129       .00662       .0283 to       .0543     |
| P            1.41149       .19389      1.0315 to      1.7915     |
| Median      24.21755      3.88060     16.6116 to     31.8235     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25       .50       .75       .95                 |
| Time         52.74     24.22     11.12      3.01                 |
+------------------------------------------------------------------+
```

```
-----------------------------------------------------------------------------
Loglinear survival model: INVERSE GAUSSIAN
Dependent variable                     LOGT
Log likelihood function      -298.24556
Restricted log likelihood    -331.34838
Chi squared [   2 d.f.]        66.20563
Significance level               .00000
McFadden Pseudo R-squared      .0999034
Estimation based on N =     62, K =   3
Inf.Cr.AIC  =    602.5 AIC/N =    9.718
Model estimated: Aug 11, 2011, 11:05:33
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    2.53283***       .65173     3.89  .0001     1.25547   3.81019
    PROD|   -10.4031       10.93846     -.95  .3416    -31.8421   11.0359
        |Scale parameter for inverse gaussian model
   Sigma|     .28218         .24945     1.13  .2580     -.20674    .77110
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------+
| Percentiles  of  survival  distribution at data means    |
| Survival       .25       .50       .75       .95         |
| Time         41.50     16.13      7.10      2.80         |
+----------------------------------------------------------+
```

In order to compare the various models, one can examine the coefficients, log likelihoods, and various diagnostic statistics. A tool which can be used to diagnose model adequacy is the integrated hazard function, $ih(t)$ = -log$S(t)$ where $S(t)$ is the survival function. Under the hypothesis that the model specification is correct, the integrated hazard function should be a straight line emanating from the origin. Departures from this might signal model misspecification. For example, the figures below are produced by adding **; Plot** to the Weibull and inverse Gaussian models: The curvature of the integrated hazard for the inverse Gaussian model suggests (slightly) that the Weibull might be the preferred model.

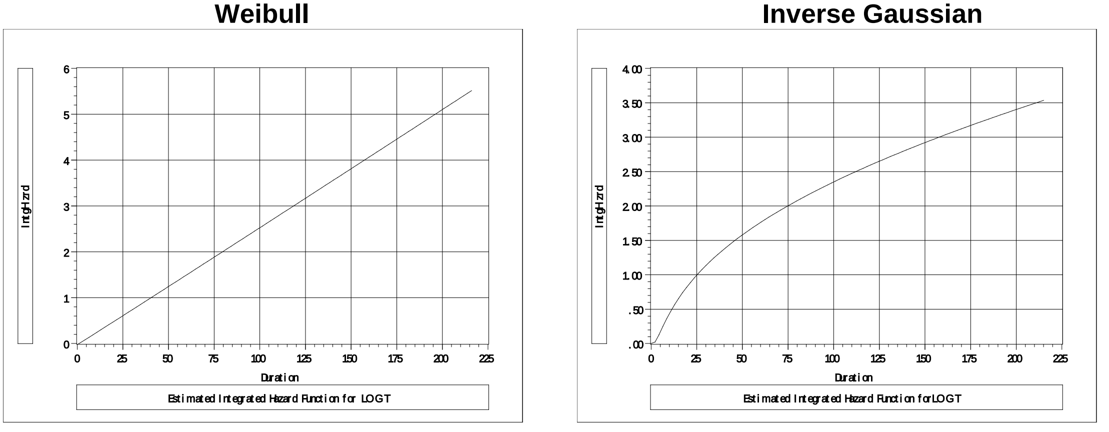


Figure E60.1 Comparison of Weibull and Inverse Gaussian Models

To examine the effect of censoring, we modify the data by censoring observations at T = 80. The results produced earlier with the uncensored data are repeated.

```
CREATE      ; ct = 80 ~ t ; d = (ct = 80) $
CREATE      ; logct = Log(ct) $
SURVIVAL    ; Lhs = logct ; Rhs = one,prod
            ; Model = Weibull
            ; Plot $
```

```
-----------------------------------------------------------------------------
Loglinear survival model: WEIBULL
Dependent variable                LOGCT
Log likelihood function       -89.64803
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   LOGCT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.61369***      .14190    25.47  .0000     3.33558   3.89180
    PROD|   -6.41060**      2.89543    -2.21  .0268   -12.08553   -.73567
        |Ancillary parameters for survival
   Sigma|     .85049***      .10611     8.01  .0000      .64251   1.05847
--------+--------------------------------------------------------------------
```

```
+----------------------------------------------------------------------------+
| Alternative representations of survival model:                             |
| Accelerated Failure Time:  b(k)  =  beta(k)    (Given above)               |
+--------+------------------------------+-----------------------------------+
|        | Proportional Hazards         | Hazard Ratios                     |
|        | c(k) = -1/sigma*beta(k)      | h(k)=exp[(-1/sigma)*beta(k)]      |
|        | z  =  c(k)/Std.Err.[c(k)]    | z  = [h(k)-1]/Std.Err.[h(k)]      |
+--------+------------------------------+-----------------------------------+
|Variable|     c(k)   Std.Err.   |z|    |     h(k)       Std.Err     |z|    |
+--------+------------------------------+-----------------------------------+
|PROD    |    7.5376    3.7119   2.031 |    1877.2373   6968.0266    .269 |
+----------------------------------------------------------------------------+
```

```
+------------------------------------------------------------------+
| Parameters of underlying density at data means:                  |
| Parameter   Estimate   Std. Error     Confidence Interval        |
| ------------------------------------------------------------     |
| Lambda        .02893       .00376       .0216 to       .0363     |
| P            1.17580       .14670       .8883 to      1.4633     |
| Median      25.31274      3.28804     18.8682 to     31.7573     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25       .50       .75       .95                 |
| Time          45.64     25.31     11.98      2.76                |
+------------------------------------------------------------------+
| Uncensored data                                                  |
| Lambda        .02530       .00337       .0187 to       .0319     |
| P            1.00786       .12254       .7677 to      1.2480     |
| Median      27.47425      3.66307     20.2946 to     34.6539     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25       .50       .75       .95                 |
| Time          54.65     27.47     11.48      2.07                |
+------------------------------------------------------------------+
```

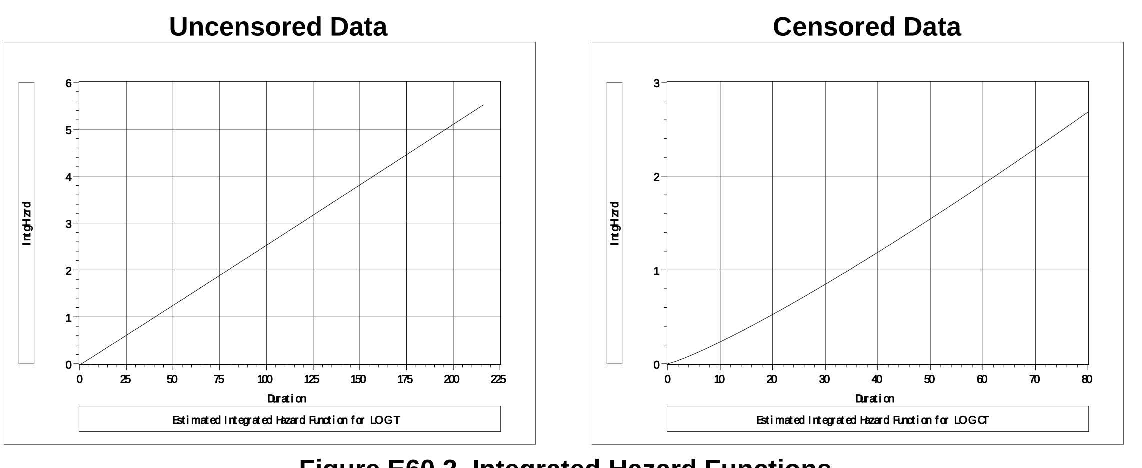


Figure E60.2 Integrated Hazard Functions

E60.6 Gamma, Gompertz and Generalized F Models

These three models generally require additional information specific to the model to set up the estimation command. The Gompertz model can be estimated using more than one procedure. Owing to the complexity of the model, a specific application may require use of different procedures to obtain a result.

E60.6.1 Estimating the Gamma Model

The gamma model must be treated differently from the other models. The parameter θ cannot be easily estimated simultaneously with the other parameters because of the difficulty of computing the derivative of the log likelihood. One method of estimation is to search over θ. You can provide the value of θ at the time the command is requested, as follows:

SURVIVAL ; Lhs = ... ; Rhs = ... ; Model = Gamma ; Theta = value $

Even though θ is supplied by you rather than searched for by the iterative algorithm, it is still treated as an unknown parameter. *LIMDEP* will compute an estimated standard error for the estimate of θ and factor this variance into the estimated covariance matrix for the other parameter estimates. I.e., it treats it just like the other parameters. Unlike the other parameters, though, we use a first difference approximation to estimate the derivative of the log likelihood with respect to θ.

You may specify that θ is to be treated as fixed in the preceding and not allow its variance to be factored into the estimated asymptotic covariance matrix. This will nearly always result in the remaining estimated standard errors being smaller than when θ is treated as having been estimated. To request this, add

; Fix

to the command. The output will clearly show the constraint.

Three alternative formulations may also be specified for the gamma model:

1. To fix σ at some value and allow θ to be freely estimated instead, use

   NAMELIST ; x = the set of Rhs variables $
   CALC ; k = Col(x) $
   SURVIVAL ; Lhs = ... ; Rhs = x ; Model = Gamma
   ; Rst = k_b, value for sigma, tt $

2. To fix both σ and θ, use the same as above, but instead of the free label *tt* above, insert the desired fixed value.

Experience suggests that estimation of the gamma model with one or both of the parameters fixed is fairly routine. The third approach is to allow both σ and θ to vary freely, and be estimated as free parameters. To request this, simply use

SURVIVAL ; Lhs = ... ; Rhs = ... ; Model = Gamma $

To continue, our experience with some carefully constructed data sets suggests that the model with free σ and θ is quite difficult to estimate. We found in most cases that θ and the slope parameters wandered off to extreme values, and ultimately, it was not possible to obtain convergence of the estimator. (It is quite well behaved with our sample data, however.)

A number of options discussed below are unavailable for the gamma model:

- truncation
- splitting model
- gamma heterogeneity
- time varying covariates

The following show several formulations of the gamma model. For the data analyzed here, the unrestricted model turns out to be fairly easily estimated.

SURVIVAL ; Lhs = logt ; Rhs = one,prod ; Model = Gamma $
SURVIVAL ; Lhs = logt ; Rhs = one,prod ; Model = Gamma ; Theta = .5 $
SURVIVAL ; Lhs = logt ; Rhs = one,prod ; Model = Gamma ; Theta = 1.5 $

```
-----------------------------------------------------------------------------
Loglinear survival model: GENRL.GAMMA
Dependent variable                 LOGT
Log likelihood function       -97.28530
Estimation based on N =     62, K =   4
Inf.Cr.AIC  =    202.6 AIC/N =    3.267
Model estimated: Aug 11, 2011, 13:25:31
Generalized GAMMA Model, Theta=   1.015
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.76120***     1.40938     2.67  .0076      .99887   6.52354
    PROD|   -9.33410***     2.98174    -3.13  .0017   -15.17820  -3.48999
        |Ancillary parameters for survival
   Sigma|    1.00138         .71035     1.41  .1586     -.39088   2.39364
   THETA|    1.01482        1.18452      .86  .3916    -1.30679   3.33643
--------+--------------------------------------------------------------------

+------------------------------------------------------------------+
| Parameters of underlying density at data means:                  |
| Parameter   Estimate   Std. Error     Confidence Interval        |
| ------------------------------------------------------------     |
| Lambda        .02578       .03642      -.0456 to       .0972     |
| P             .99862       .90798      -.7810 to      2.7783     |
| Median      27.43512       .00000     27.4351 to     27.4351     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25       .50       .75       .95                 |
| Time         54.60     27.44     11.48      2.09                 |
+------------------------------------------------------------------+

Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit from iterations. Exit status=0.
```

```
--------------------------------------------------------------------------------
Loglinear survival model: GENRL.GAMMA
Dependent variable                  LOGT
Log likelihood function        -97.63613
Generalized GAMMA Model, Theta=     .500
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |RHS of hazard model
Constant|    4.42078***      .57425     7.70  .0000     3.29527   5.54630
    PROD|   -9.17447***     2.78655    -3.29  .0010   -14.63602  -3.71292
        |Ancillary parameters for survival
   Sigma|     .63011         .39717     1.59  .1126     -.14832   1.40854
   THETA|     .50000         .47370     1.06  .2912     -.42844   1.42844
--------+-----------------------------------------------------------------------

+-------------------------------------------------------------------+
| Parameters of underlying density at data means:                   |
| Parameter   Estimate   Std. Error     Confidence Interval         |
| ------------------------------------------------------------      |
| Lambda        .01330       .00774      -.0019 to        .0285     |
| P            1.58701      1.08060      -.5310 to       3.7050     |
| Median      29.56565       .00000     29.5657 to      29.5657     |
| Percentiles  of  survival  distribution:                          |
| Survival       .25       .50       .75       .95                  |
| Time         57.94     29.57     11.49      1.48                  |
+-------------------------------------------------------------------+

--------------------------------------------------------------------------------
Loglinear survival model: GENRL.GAMMA
Dependent variable                  LOGT
Log likelihood function        -97.35987
Generalized GAMMA Model, Theta=    1.500
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.16619        2.41356     1.31  .1896    -1.56430   7.89668
    PROD|   -9.37034***     3.04648    -3.08  .0021   -15.34132  -3.39936
        |Ancillary parameters for survival
   Sigma|    1.27142        1.03610     1.23  .2198     -.75930   3.30213
   THETA|    1.50000        2.08518      .72  .4719    -2.58688   5.58688
--------+-----------------------------------------------------------------------

+-------------------------------------------------------------------+
| Parameters of underlying density at data means:                   |
| Parameter   Estimate   Std. Error     Confidence Interval         |
| ------------------------------------------------------------      |
| Lambda        .04675       .11290      -.1745 to        .2680     |
| P             .78652       .90435      -.9860 to       2.5590     |
| Median      26.48423       .00000     26.4842 to      26.4842     |
| Percentiles  of  survival  distribution:                          |
| Survival       .25       .50       .75       .95                  |
| Time         53.42     26.48     11.32      2.35                  |
+-------------------------------------------------------------------+
```

E60.6.2 Estimating the Gompertz Model

Estimation of λ and p for a (possibly censored) set of observations on t = time is done as follows:

SURVIVAL ; Lhs = t[,status] ; Model = Gompertz $

The status variable is optional. Note that there are no covariates. Starting values for the parameters are obtained from the Kaplan-Meier estimates of the hazard function, based on

$$\log h_t = \log p + \lambda t.$$

We use the least squares coefficients in the regression of the estimated hazards on a constant and '*t*.' The log of the hazard, which has been tabulated for 10 or more values of t, is, theoretically, linear in t. Then, λ and p are estimated directly by maximum likelihood. They and the estimated median of the distribution are presented with standard errors. Several percentiles of the distribution are also presented. A listing by observation of time, the survival rate, density of the distribution, and hazard rate is obtained by adding

; List

to the model command.

NOTE: The input (Lhs) variable is *time* for this model, not log(*time*).

When the Gompertz model is estimated with covariates, you must provide starting values. They are optional with the other models. It is difficult to obtain good starting values for this model – it depends partly on the data. Here are two strategies:

1. Use the procedure described above to fit the model without the covariates. Then, use **; Start = *** in your later command to specify the starting values for the expanded model. This will result in an initial assumption of zero for all coefficients but the constant term. Thus, for example,

   SURVIVAL ; Lhs = time ; Model = Gompertz $
   SURVIVAL ; Lhs = time ; Rhs = one,sex
   ; Start = * ; Model = Gompertz $

2. Estimate some other functional form, such as the Weibull and use the estimates as the starting values for the Gompertz model. These will not be particularly good, but they will probably be better than zero as they will provide some information about relative sizes. If you do this, you must use the **; Par** option to make sure that an estimate of σ = 1/P gets passed as well. Thus, for example,

   SURVIVAL ; Lhs = Log(time) ; Rhs = one,sex ; Model = Weibull ; Par $
   SURVIVAL ; Lhs = time ; Rhs = one,sex ; Start = b ; Model = Gompertz $

We note, within our experience, regardless of the strategy chosen, estimate of the Gompertz model will be a challenge in most cases.

The following shows an exercise in which the Weibull and Gompertz hazard functions are compared. The **MAXIMIZE** command is used to fit the Gompertz model.

```
SAMPLE      ; 1-62 $
SURVIVAL    ; Quietly ; Lhs = logt
            ; Rhs = one,prod ; Model = Weibull $
CALC        ; pw = 1/s $
CALC        ; prbar = Xbr(prod) $
CALC        ; lbarw = Exp(-b(1)-b(2)*prbar) $
SURVIVAL    ; Quietly ; Lhs = logt ; Rhs = one,prod ; Model = Exponential $
MATRIX      ; beta0 = b $
CALC        ; p0 = 1/s $
MAXIMIZE    ; Labels = b0,b1,pgomp ; Start = beta0,p0
            ; Fcn = al = Exp(-b0-b1*prod) |
              Log(pgomp)+al*t-(pgomp/al)*(Exp(al*t)-1) ; Output = 3 $
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function      -292.91074
Estimation based on N =     62, K =   3
Inf.Cr.AIC  =    591.8 AIC/N =    9.546
Model estimated: Aug 11, 2011, 13:38:11
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    6.86763        6.03419     1.14  .2551    -4.95917   18.69443
      B1|   -41.0662       91.48359     -.45  .6535   -220.3708  138.2384
   PGOMP|     .02100***      .00359     5.85  .0000      .01397    .02804
--------+--------------------------------------------------------------------
```

```
CALC        ; lbarg = Exp(-b(1)-b(2)*prbar) $
CALC        ; pg = b(3) $
SAMPLE      ; 1-200 $
CREATE      ; time = Trn(1,1) $
CREATE      ; ghazard = pg*Exp(lbarg*time) $
CREATE      ; whazard = pw*lbarw*(lbarw*time)^(pw-1) $
PLOT        ; Lhs = time ; Rhs = ghazard,whazard
            ; Endpoints = 0,200
            ; Title = Gompertz and Weibull Hazard Functions
            ; Grid ; Fill $
```

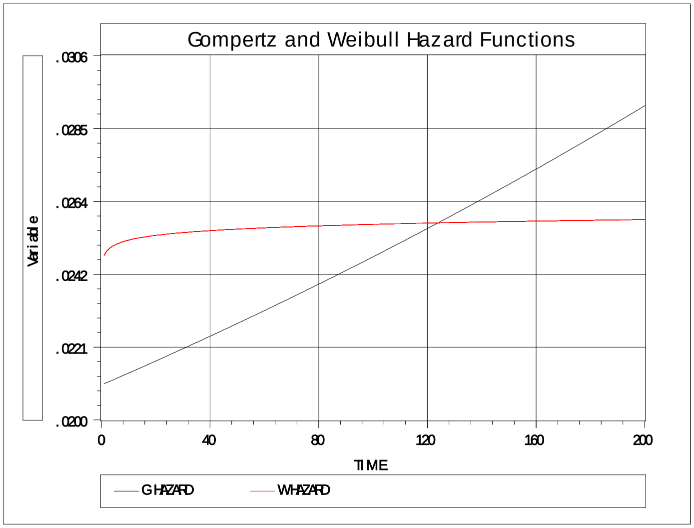


Figure E60.3 Gompertz and Weibull Hazard Functions

The hazard function for the Weibull model is essentially that of an exponential model. Since the estimated value of *p* is only 1.00786 for the Weibull model, this is to be expected.

E60.6.3 Estimating the Generalized F Model

This model is more general than the other models in the set, and may be used to help decide which is the best among the set to use as a modeling framework. The hazard function for the generalized F may display negative or positive duration dependence.

To request this model, the basic command is

```
SURVIVAL    ; Lhs  = ... as usual, including censoring if appropriate
            ; Rhs  = ... as usual
            [; Rh2= ... if the splitting model is desired (with the other
                        options for the splitting model)]
            [; Hfn = ... if the variance heterogeneity model is desired]
            ; Model  = F
            ... (other options) $
```

All options for the parametric survival models are available except the time varying covariates specification. The model may be estimated with *M*1 and *M*2 fixed or free. Two available setups are as follows:

1. *ML estimation of free M*1 *and M*2. If you wish to let the optimization procedure find *M*1 and *M*2, just use the **; Model = F** form of the command exactly as shown above.

2. *Fixed M*1 *and/or M*2: The likelihood for this model is a bit ill behaved in some data sets, as *M*1 and/or *M*2 begin to wander to the values for the special cases, i.e., +∞. You can fix one or both of the two parameters using a special form of the command. To fix both *M*1 and *M*2 at particular values, use

   ; Model = F (value1, value2)

For example, **; Model = F(1,1)** produces the loglogistic distribution. This will produce conditional maximum likelihood estimates of the other parameters in the model with $M1$ and $M2$ fixed at the specific values you give. To fix just one of them, use

; Model = F (M1, value2) to fix $M2$, for example, F($M1$,2.5)
; Model = F (value1, M2) to fit $M1$, for example, F(1.25,$M2$).

The estimator computes maximum likelihood estimates of all free model parameters.

The full parameter vector in the generalized F model is, in order,

β = slope parameters in index function (mandatory),
α = slope parameters in splitting model (optional),
γ = parameters in variance heterogeneity model (optional – see Section E61.4),
$M1,M2$ = degrees of freedom parameters in F distribution (mandatory),
σ = scale (variance parameter) for duration distribution (mandatory).

You may also fix $M1$ and/or $M2$ using the **; Rst** specification. But, you should not use this unless you are constraining other parameters in the model as well. For fixing only $M1$ and/or $M2$, **; Model = F(...,...)** does the same thing, and is much simpler.

The following are special cases of *LIMDEP*'s generalized F model.

Distribution	**Form of the Generalized F Distribution**		
loglogistic:	$M1 = 1$	$M2 = 1$	σ free
lognormal:	$M1 \to +\infty$	$M2 \to +\infty$	σ free
Weibull:	$M1 = 1$	$M2 \to +\infty$	σ free
exponential:	$M1 = 1$	$M2 \to +\infty$	$\sigma = 1$
gamma:	$M1$ free	$M2 \to +\infty$	σ free

Table E60.7 Special Cases of the Generalized F Distribution

To specify the limiting forms, you can use a large value for $M1$ and/or $M2$ (or, of course, use the form directly).

For an example of the generalized F, we use Kennan's strike data with the production data used earlier. Without the censoring, the results strongly support the Weibull specification – indeed the ML results are virtually identical to the Weibull model. $M2$ wanders off to an extreme value while $M1$ moves toward one, suggesting that the Weibull model gives the highest likelihood. However, the exponential model cannot be rejected based on its log likelihood of -97.285. Note that the standard errors for the estimated parameters of the Weibull model will be inflated by the presence of $M1$ and $M2$, compared to the estimates when the Weibull model is specified explicitly. Recall, the integrated hazard function plotted earlier also suggested that a Weibull model would be appropriate.

```
SURVIVAL    ; Lhs = logt ; Rhs = one,prod
            ; Model = F $
```

```
Warning   141: Iterations:current or start estimate of sigma nonpositive
Warning   141: Iterations:current or start estimate of sigma nonpositive
Maximum of    50 iterations. Exit iterations with status=1.

-----------------------------------------------------------------------------
Loglinear survival model: GENRLIZED F
Dependent variable                 LOGT
Log likelihood function        -97.28532
Generalized F(    1.01,96241.70)
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.77595***      .28074    13.45  .0000     3.22571   4.32620
    PROD|   -9.33410***     3.36300    -2.78  .0055   -15.92545  -2.74274
        |Ancillary parameters for survival
      M1|    1.01474        4.09810      .25  .8044    -7.01739   9.04688
      M2|    96241.7      .1463D+11      .00 1.0000 ***********  ***********
   Sigma|    1.00133        3.17303      .32  .7523    -5.21771   7.22036
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+---------------------------------------------------------------------------+
| Alternative representations of survival model:                            |
| Accelerated Failure Time:  b(k)  =  beta(k)    (Given above)              |
+--------+------------------------------+-----------------------------------+
|        | Proportional Hazards         | Hazard Ratios                     |
|        | c(k) = -1/sigma*beta(k)      | h(k)=exp[(-1/sigma)*beta(k)]      |
|        | z  =  c(k)/Std.Err.[c(k)]    | z  = [h(k)-1]/Std.Err.[h(k)]      |
+--------+------------------------------+-----------------------------------+
|Variable|     c(k)   Std.Err.   |z|    |     h(k)       Std.Err     |z|    |
+--------+------------------------------+-----------------------------------+
|PROD    |    9.3217   28.2834    .330 |   11178.4412 316164.5113    .035 |
+---------------------------------------------------------------------------+

+------------------------------------------------------------------+
| Parameters of underlying density at data means:                  |
| Parameter   Estimate   Std. Error     Confidence Interval        |
| -----------------------------------------------------------      |
| Lambda        .02540       .00732       .0111 to       .0397     |
| P             .99868      3.16464     -5.2040 to      7.2014     |
| Median      27.27760       .00000     27.2776 to     27.2776     |
| Percentiles  of  survival  distribution:                         |
| Survival       .25        .50        .75        .95              |
| Time         54.61      27.28      11.31       2.01              |
+------------------------------------------------------------------+
```

E60.7 Time Varying Covariates

As noted earlier, we have assumed thus far that the covariates are constant from the beginning of the measurement period, $T = 0$, to the time of the measurement, $T = t_i$. There will be circumstances in which this assumption must be relaxed. For example, a model of the duration of unemployment might include marital status, and the individual's marital status might change during the spell. A second example might be job tenure, during which rank or position in the firm might have changed. This section presents a method of incorporating true *time varying covariates* in the duration model. We draw heavily on Petersen (1986a, 1986b). Petersen and we differ on the definition of time varying as opposed to time dependent covariates. We use the latter to denote covariates which are written as explicit functions of time. We use the term *time varying covariates* to denote what he calls time dependent covariates.

Following Petersen, we formulate the model as follows: Let the interval 0 to t_i be divided into k exhaustive, nonoverlapping intervals, $t_0 < t_1 < ... < t_{k-1} < t_k$, where $t_0 = 0$ and $t_k = t_i$. The covariates are assumed to stay constant within each of the k intervals, but may change from one interval to the next. Let

$$h(t|\mathbf{x}_j) \; = \; \text{hazard function from time } t_{j-1} \text{ to } t_j,$$

since within that interval, the covariates are constant. We deviate slightly from Petersen's notation. His formulation includes both time varying covariates, denoted $\mathbf{x}(t)$ and time dependent covariates, $z(t)$. The models considered here allow only the former. Then, from the relationship between the hazard function and the survival rate,

$$h_t \; = \; -\text{dlog}S(t)/\text{d}t,$$

and

$$\text{Prob}[T \le t_j \mid T \ge t_{j-1}] \; = \; \exp -\int_{t_{j-1}}^{t_j} h(s \mid \mathbf{x}_j)\text{d}s.$$

The survival function for duration of t_k or more can then be written

$$S(t_k \mid \mathbf{x}_k) \; = \; \prod_{j=1}^{k} \text{Prob}[T \ge t_j \mid T \ge t_{j-1}].$$

Finally, the density at t_k is

$$f(t_k \mid \mathbf{x}_k) \; = \; h(t_k)S(t_k).$$

The log likelihood function for one observation is

$$\log L_i \; = \; \delta_i \log h(t_k \mid \mathbf{x}_k) \; + \; \log S(t_k).$$

Thus, each observation contributes the survivor function to the log likelihood function. For noncensored observations, we add the density, evaluated at the terminal point. Therefore,

$$\log L_i \; = \; \delta_i \log h(t_k \mid \mathbf{x}_k) \; - \; \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} h(s \mid \mathbf{x}_j)\text{d}s.$$

The hazard function is modeled as a step function, with different values of the covariates through several intervals between $t = 0$ and $t = t_i$, the terminal value in the observation, at which either censoring or exit takes place. This requires one or more lines of data per observation, since the covariates must be provided for each interval observed. The number may vary by observation. For example, suppose marital status and education are the covariates in a model of duration of unemployment. For an observation i, education is constant, $E = 12$, say, for the entire observation period, but marital status changes at $t = 10$. The observation period is 0 to 24. This would require two lines of data

	t	*e*	*m*	*status*	*nperiod*
1st	10	12	0	1	2
2nd	24	12	1	1	2

The censoring status is provided for all periods, even though only the one on the last record is needed. This is requested with **; Pds = nperiod**, where *nperiod* is a variable which tells how many lines of data are needed for the observation. The model command is

```
SURVIVAL    ; Lhs = time or logtime, censoring status
            ; Rhs = covariates
            ; Pds = nperiod
            ; Model = ... as usual, along with other options $
```

The only change is the addition of **; Pds = nperiod**. Note that this is the same setup as the discrete choice model with variable numbers of choices as well as the other panel data estimators in *LIMDEP*. As discussed earlier, only the Gompertz model uses actual time, as opposed to the logarithm of time as the Lhs variable.

NOTE: The number of periods, *nperiod*, was given as a third Lhs variable in previous versions of *LIMDEP*. (You may continue to use this format if you wish.)

This formulation is available for the following models: Weibull, loglogistic, exponential, Gompertz, and Weibull with heterogeneity. The excluded models are the normal, split population, gamma, and generalized F models.

Technical Details

For the model with time varying covariates, we construct the log likelihood as the sum of terms

$$\log L_i = \delta_i \log h(t_k \mid \mathbf{x}_k) - \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} h(s \mid \mathbf{x}_j) ds.$$

where, for the present, we have reverted back to expressing time in natural units. The hazard functions for the distributions which include this feature are listed below. To construct the second term in the log likelihood, we require the indefinite integrals of these functions. The terms in the likelihood functions for these models are listed in the table below.

Model	**Hazard Function**	**Indefinite Integral**
Weibull	$(\lambda p)(\lambda t)^{p-1}$	$(\lambda t)^p$
Weibull/gamma	$(\lambda p)(\lambda t)^{p-1} / [1 + \theta(\lambda t)^p]$	$(1/\theta)\log[1 + \theta(\lambda t)^p]$
exponential/gamma	$\lambda / [1 + \theta\lambda t]$	$(1/\theta)\log[1 + \theta(\lambda t)]$
exponential	λ	λt
loglogistic	$(\lambda p)(\lambda t)^{p-1} / [1 + (\lambda t)^p]$	$\log[1 + (\lambda t)^p]$
Gompertz	pe^{λ}	$(p/\lambda)e^{\lambda t}$

Table E60.8 Characteristics of Survival Distributions

For the Gompertz model, it is more convenient to leave the distribution in its original form. Petersen (1986a, 1986b) used a modified form of nonlinear least squares to estimate the parameters. We apply the BFGS method directly to the log likelihood. Derivatives with respect to $\boldsymbol{\beta}$ and σ are obtained as follows:

$$\partial \log L_i / \partial \boldsymbol{\beta} = (\partial \log L_i / \partial \lambda_i)\lambda_i \mathbf{x}_i,$$

$$\partial \log L_i / \partial \sigma = (\partial \log L_i / \partial p)(dp/d\sigma) = (\partial \log L_i / \partial p)(-1/\sigma^2).$$

E61: Panel Data and Heterogeneity in Parametric Duration Models

E61.1 Introduction

This chapter develops several extensions of the duration models for panel data and other nonstandard forms. The other extensions include heterogeneity, heteroscedasticity and a two part model for sample selection.

E61.2 Panel Data Models

The following are less natural for the parametric survival models than they are for, say, the linear regression model or the probit model. The notion of 'clustering' might be more applicable than panel data in this context. Consider a setting in which observations occur in naturally collected groups, in which group members have common attributes. For example, one might consider bank or business failures in which there is a strong regional or local influence shared within a group, but perhaps not between groups. Then, we consider a parametric model of the form

$$f(t_{ij}) = (1/\sigma)\, f(w_{ij})$$

where $$w_{ij} = (1/\sigma)\,(\log t_{ij} - \boldsymbol{\beta}_i'\mathbf{x}_{ij})$$

where 'i' indexes the group, 'j' indexes the member of group i, and there are N_i members of group i. *LIMDEP* provides four formulations for such a model:

1. Fixed effects
2. Random effects
3. Random parameters
4. Latent class

Extensive descriptions of these four modeling frameworks and how *LIMDEP* does the estimation appear elsewhere in this manual, e.g., in Chapters R22-R25, so at this juncture, we will present only the essential details. We emphasize, again, in this framework, the usual panel data interpretation is a bit ambiguous, since one would not normally observe the same individual repeatedly. The 'cluster of observations' seems a more appropriate application. As a consequence, covariance structures for observations should be 'exchangeable' – that is, any time sequencing operation, such as autocorrelation, would normally not be used in this context.

NOTE: These 'panel data models' are available only for the Weibull, loglogistic and lognormal models.

All models will contain the mandatory part of the specification

SETPANEL ; Group = identifier
; Pds = count variable $
SURVIVAL ; Lhs = the log of the time variable
; Rhs = the list of covariates
; Model = Weibull or Normal or Loglogistic
; Panel
; ... specification of the particular model form
; ... any other options $

Since the Kennan data are grouped by year, they can be viewed as a panel (by our earlier loose interpretation). To illustrate the different estimators, we will fit several different forms of the Weibull model which appears to be a good choice for these data.

E61.2.1 Fixed Effects Models

The fixed effects model would result from

$$\boldsymbol{\beta}_i'\mathbf{x}_i^* = \beta_i^0 + \boldsymbol{\beta}'\mathbf{x}_i$$

where $\mathbf{x}_i$* is the full covariate vector including a constant term while $\mathbf{x}_i$ is all variables not including a constant term, and where $\boldsymbol{\beta}_i$ is partitioned conformably. The fixed effects estimator is requested with

; FEM
; Panel

NOTE: The fixed effect cannot be estimated for any group in which all observations in the group are censored. These groups must be dropped from the analysis. The output for the estimator will indicate how many times this condition was encountered in the data set.

Since *prod* does not vary within each year, the fixed effects model must be fit with just the constant terms. Since there are no covariates, this might seem to be equivalent to fitting separate models for each year, however, the underlying variance parameter is constrained to be the same in every year. (We created the count variable, *ni*, by hand, since the sample is so small. Purely for this numerical example, we created nine groups with seven observations in each group save for the last, which has six.)

SURVIVAL ; Lhs = logt ; Rhs = one
; Model = Weibull
; Panel ; FEM ; Par $
MATRIX ; List ; alphafe $

```
-----------------------------------------------------------------------------
FIXED EFFECTS SWeibl Model
Dependent variable                   LOGT
Log likelihood function         -86.95720
Estimation based on N =      62, K =  10
Unbalanced panel has      9 individuals
Skipped    0 groups with inestimable ai
Weibull loglinear survival model
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variance parameter given is sigma
Std.Dev.|     .81949***      .08502     9.64  .0000      .65286    .98612
--------+--------------------------------------------------------------------

 ALPHAFE|             1
--------+--------------
       1|       2.13448
       2|       2.93709
       3|       3.65576
       4|       3.87759
       5|       3.26583
       6|       4.23517
       7|       4.67064
       8|       3.65549
       9|       4.00311
```

E61.2.2 Random Effects and Random Parameters Models

The random parameters model specifies that

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$$

where β is the mean of the distribution, Δ is a matrix of coefficients, and $\boldsymbol{\Gamma}$ is a lower triangular matrix. By this specification, each coefficient in the model is

$$\beta_{ik} = \beta_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \boldsymbol{\gamma}'\mathbf{v}_i.$$

Any coefficient can be assumed to be nonrandom. Correlation across parameters is achieved by having nonzero off diagonal elements in $\boldsymbol{\Gamma}$. The diagonal elements are the scales (not necessarily the standard deviations) for the random terms.

NOTE: A random effects model is obtained in this framework by allowing only the constant term to be random and not providing any '*z*' variables for the heterogeneous mean.

The random parameters models are specified by providing, for each random parameter desired:

; RPM [= list of variables in z if desired. This is optional.]
; Fcn = name of variable (n, u, t, etc.)
for normal, uniform, tent distribution, etc.

LIMDEP's Weibull and exponential models support heterogeneity with log-gamma density by writing the survival function as

$$S(t|v) = v \times \exp[(-\lambda t)^P]$$

where v has a gamma density with mean one.

The random parameters formulation of the parametric models allows the modeler to incorporate heterogeneity in the parametric survival models in the form of variation in the model parameters, β. Thus, for example, the hazard function for the Weibull model is

$$h(t_i) = \lambda_i \, p \, (\lambda_i t_i)^{p-1}$$

where $\lambda_i = \exp(\boldsymbol{\beta}_i{}'\mathbf{x}_i)$

and, we allow $\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\,\mathbf{v}_i$.

This builds individual heterogeneity into the hazard rate in a different manner than in the gamma model above. Note, however, that if only the constant term in $\boldsymbol{\beta}$ is so affected, then this random parameter model becomes the same as the gamma model above with a different distribution. Typically the heterogeneity would be assumed to arise from a normally distributed v_i, in which case, the gamma variable v in the earlier model would be changed to include a lognormal γv_i in this new formulation.

This formulation, available for the Weibull, exponential, lognormal and loglogistic survival models, adds these survival models to *LIMDEP*'s class of random parameter models. The full range of features for the random parameter models is available. The function definition may specify that any parameter in the model is random. The randomness may be 'pure' as in

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Gamma}\,\mathbf{v}_i$$

which replicates (at least in spirit) earlier formulations, or heterogeneous,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\,\mathbf{v}_i.$$

We should note, the preceding heterogeneity model assumes (as all random 'effects models' do) that the heterogeneity is uncorrelated with the covariates. If this is not true, then parameter estimates are inconsistent and inference about them from that model is possibly problematic. In other settings, the 'fixed effects' model is the preferable alternative. *LIMDEP*'s fixed effects estimator is available for these models as well. However, it is unclear how viable an option this will be, because the fixed effects model requires panel data, and one typically does not observe panels of duration data.

The random parameters model with a random constant term is comparable to the latent heterogeneity models discussed in Section E61.3. We will treat the data as a panel here – there are nine years of data. For these treatments, we assume that the latent effect is common to all observations in the given year, as it would be in the spirit of Kennan's treatment in which the *prod* variable represents a macroeconomic shock that hits all industries. Thus, the treatment here assumes that there are other shocks that do likewise.

The first model is a simple random effects formulation. The second allows both parameters to be random.

SURVIVAL ; Lhs = logt ; Rhs = one,prod
; Model = Weibull
; Panel ; RPM ; Fcn = one(n) ; Halton $
SURVIVAL ; Lhs = logt ; Rhs = one,prod
; Model = Weibull
; Panel ; RPM ; Fcn = one(n),prod(n) ; Halton ; Corr ; Pts = 200 $

```
-----------------------------------------------------------------------------
Random Coefficients  WiblSurv Model
Dependent variable                 LOGT
Log likelihood function       -95.33572
Estimation based on N =     62, K =   4
Simulation based on    100 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    PROD|   -8.84829***     1.81985    -4.86  .0000   -12.41513  -5.28144
        |Means for random parameters
Constant|    3.69841***      .13270    27.87  .0000     3.43832   3.95850
        |Scale parameters for dists. of random parameters
Constant|     .50916***      .13244     3.84  .0001      .24958    .76873
        |Scale parameter for survival distribution
ScalParm|     .88209***      .11850     7.44  .0000      .64983   1.11435
--------+--------------------------------------------------------------------
```

The next set of results extends this model by allowing the slope to be random as well. In view of the results above, one should not expect much improvement.

```
-----------------------------------------------------------------------------
Random Coefficients  WiblSurv Model
Dependent variable                 LOGT
Log likelihood function       -89.14375
Simulation based on    200 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|    3.84798***      .09445    40.74  .0000     3.66286   4.03309
    PROD|   -17.4211***     2.12527    -8.20  .0000    -21.5865  -13.2556
        |Diagonal elements of Cholesky matrix
Constant|     .58786***      .11122     5.29  .0000      .36986    .80585
    PROD|    15.2982***     2.32825     6.57  .0000     10.7349   19.8615
        |Below diagonal elements of Cholesky matrix
lPRO_ONE|    1.58832        2.16962      .73  .4641    -2.66406   5.84070
        |Scale parameter for survival distribution
ScalParm|     .67565***      .06457    10.46  .0000      .54909    .80222
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Implied covariance matrix of random parameters
Var_Beta|             1             2
--------+----------------------------
       1|       .345575       .933705
       2|       .933705       236.559
Implied standard deviations of random parameters
S.D_Beta|             1
--------+--------------
       1|       .587857
       2|       15.3805
Implied correlation matrix of random parameters
Cor_Beta|             1             2
--------+----------------------------
       1|       1.00000       .103269
       2|       .103269       1.00000
```

E61.2.3 Latent Class Models

The latent class model specifies a complete parameter vector, including σ, for each of *J* latent classes. The model is specified with

; LCM ; Pts = the number of classes desired

We fit a three class model to the strike data. In view of the previous results, these should not be expected to amount to much. Surprisingly, the latent class model seems to work quite well. However, a closer look at the results suggests that we have overfit it. The second and third classes, while different from the first, are nearly identical to each other. When we refit the model with two classes, the results are virtually identical to those below, with the second and third classes simply combined to one class.

SURVIVAL ; Lhs = logt ; Rhs = one,prod ; Model = Weibull
; Panel ; LCM ; Pts = 3 ; List $

```
-----------------------------------------------------------------------------
Latent Class / Panel WiblSurv Model
Dependent variable                 LOGT
Log likelihood function       -89.58133
Model fit with  3 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Model parameters for latent class 1
Constant|    4.02923***      .17348    23.23  .0000     3.68921   4.36925
    PROD|   -6.52161*       3.44670    -1.89  .0585   -13.27702    .23379
   Sigma|     .87649***      .11016     7.96  .0000      .66057   1.09241
        | Model parameters for latent class 2
Constant|    3.21546***      .48212     6.67  .0000     2.27053   4.16040
    PROD|   -24.3683       14.86458    -1.64  .1011    -53.5023    4.7658
   Sigma|     .67642***      .14614     4.63  .0000      .38998    .96286
        | Model parameters for latent class 3
Constant|    2.49119***      .39354     6.33  .0000     1.71986   3.26252
    PROD|   -9.19942        8.38600    -1.10  .2726   -25.63568   7.23683
   Sigma|     .52569**       .21591     2.43  .0149      .10252    .94885
        | Estimated prior probabilities for class membership
Class1Pr|     .68785***      .17967     3.83  .0001      .33570   1.04001
Class2Pr|     .24318         .27691      .88  .3798     -.29954    .78591
Class3Pr|     .06896         .18114      .38  .7034     -.28606    .42398
```

```
============================================================================
Predictions computed for the group with the largest posterior probability
Obs.  Periods Fitted outcomes
============================================================================
Ind.=    1  J* = 3  P(j)=  .001  .378  .621
       01-07   93.7   31.2   93.7   31.2   93.7   37.0   93.7
Ind.=    2  J* = 2  P(j)=  .004  .996  .000
       01-07   11.4   77.0   11.4   77.0   11.4   77.0   11.4
Ind.=    3  J* = 1  P(j)= 1.000  .000  .000
       01-07   91.1   61.4   83.6   61.4   83.6   61.4   83.6
Ind.=    4  J* = 1  P(j)= 1.000  .000  .000
       01-07   61.4   83.6   61.4   83.6   61.4   83.6   61.4
Ind.=    5  J* = 1  P(j)= 1.000  .000  .000
       01-07  133.1   61.4  133.1   61.4  133.1   61.4  133.1
Ind.=    6  J* = 1  P(j)= 1.000  .000  .000
       01-07   61.4  133.1   61.4  133.1   61.4  133.1  215.7
Ind.=    7  J* = 1  P(j)=  .985  .015  .000
       01-07  133.1  215.7  133.1  104.5  133.1  104.5  133.1
Ind.=    8  J* = 2  P(j)=  .202  .798  .000
       01-07  149.3  831.4  149.3  831.4  149.3  831.4  149.3
Ind.=    9  J* = 1  P(j)=  .999  .001  .000
       01-06   95.3  104.5   95.3  104.5   95.3  104.5
```

E61.3 Latent Heterogeneity

This section considers handling heterogeneity in the survival models. We first provide a robust covariance matrix for the case of uncorrected latent heterogeneity. Sections E61.3.2 to E61.3.4 consider explicit treatments for latent heterogeneity in survival models. The traditional approach has been to embed a gamma distributed latent effect in a Weibull or exponential model. In Section E61.3.2, we present a more general, flexible model that can be used in several of the parametric models. Sections E61.3.3 and E61.3.4 present the traditional models. The final part, discussed in Section E61.4, presents a model of scale heterogeneity, which would be a counterpart to heteroscedasticity in a regression context.

E61.3.1 A Heterogeneity Corrected Covariance Matrix

Under certain conditions (see Gourieroux, Monfort, and Trognon (1984)), an appropriate asymptotic covariance matrix for a ‘pseudo maximum likelihood estimator’ can be obtained by using

$$\mathbf{V} = \mathbf{H}^{-1}\,(\mathbf{BHHH})\,\mathbf{H}^{-1}$$

where **H** is the negative expected Hessian of the pseudo-log likelihood and **BHHH** is the expected outer product of the first derivatives, i.e., the inverse of the BHHH estimator of the asymptotic covariance matrix. The pseudo-log likelihood is the incorrectly assumed log likelihood for our purposes. This computation can be done for any of the parametric models by ‘tricking’ the cluster estimator. The right matrix emerges if the cluster estimator is invoked with clusters of one, so the command is

SURVIVAL ; Lhs = ... ; Rhs = ... ; Model = ... ; Cluster = 1 $

This is available for all of the parametric models described to this point.

To illustrate the estimator, we have reestimated the Weibull model with the correction to the covariance matrix. The evidence of heterogeneity below is mixed. However, the robust covariance matrix is substantially larger. The commands are

```
SURVIVAL    ; Lhs = logt
            ; Rhs = one,prod ; Model = Weibull
            ; Cluster = 1 $
SURVIVAL    ; Lhs = logt
            ; Rhs = one,prod ; Model = Weibull $
```

```
-----------------------------------------------------------------------------
Loglinear survival model: WEIBULL
Dependent variable                 LOGT
Log likelihood function       -97.28542
Estimation based on N =     62, K =   3
Inf.Cr.AIC  =    200.6 AIC/N =    3.235
Model estimated: Aug 11, 2011, 18:57:22
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
(Robust covariance matrix, <H>*OPG*<H> is used for the estimator.)
        |RHS of hazard model
Constant|    3.77977***      .34986    10.80  .0000     3.09407   4.46548
    PROD|   -9.33220       13.19748     -.71  .4795   -35.19879  16.53439
        |Ancillary parameters for survival
   Sigma|     .99220***      .31292     3.17  .0015      .37889   1.60552
--------+--------------------------------------------------------------------
(No correction)
Constant|    3.77977***      .13833    27.32  .0000     3.50865   4.05090
    PROD|   -9.33220***     2.95428    -3.16  .0016   -15.12249  -3.54191
        |Ancillary parameters for survival
   Sigma|     .99220***      .12064     8.22  .0000      .75576   1.22865
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E61.3.2 Parametric Models with Heterogeneity

If one assumes the survival distribution is homogeneous when it is not, there are two likely consequences:

- parameter estimates will be inconsistent and/or
- inferences will be based on inappropriate standard errors.

The results of Gourieroux et al. suggest that in many settings, the primary effect of the misspecification will be the second of these, but not the first. (For extensive discussions of heterogeneity, see Kiefer (1988) and Heckman and Singer (1984).) The following will describe two estimators that deal with both possibilities by specifically incorporating heterogeneity in the model.

A familiar, traditional model of heterogeneity in the parametric survival models (see the next section), is of the form of a multiplicative term in the Weibull survival model,

$$S(t|v) = v\{\exp[(-\lambda t)^p]\}.$$

where v has a gamma density with mean one. (See, e.g., Hui (1991).) The unconditional survival function is found by

$$\begin{aligned} S(t) &= \int_0^\infty vS(t|v)f(v)\mathrm{d}v. \\ &= \left[1 + \theta(\lambda t)^p\right]^{-1/\theta} \end{aligned}$$

where θ is the parameter of the gamma distribution. There is no natural mechanism that produces this model; it is devised as a reasonable approach that is mathematically convenient.

Consider the alternative approach

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i),\ E[\varepsilon_i] = 0,\ \mathrm{Var}[\varepsilon_i] = 1,$$

where the unobserved heterogeneity enters the model in the same fashion that the observed heterogeneity does. (The assumption of zero mean is innocent if the model contains a constant term, while the unit variance is simply a scaling – the variance is carried by σ^2.) We use maximum simulated likelihood to estimate the model parameters. The term $\lambda_i(\varepsilon_i)$ enters any of the models in Table E60.1 – the estimator described here is available for the exponential, Weibull, lognormal, loglogistic and inverse Gaussian. The log likelihood, conditioned on ε_i is

$$\log L|\varepsilon_1,\ldots,\varepsilon_N = \sum_{i=1}^N \delta_i \log f[y_i,\lambda_i(\varepsilon_i),p] + (1-\delta_i)\log S[y_i,\lambda_i(\varepsilon_i),p].$$

The unconditional log likelihood to be maximized is obtained by integrating ε_i out of the conditional log likelihood;

$$\log L = \sum_{i=1}^N \int_{-\infty}^\infty \left\{\delta_i \log f[y_i,\lambda_i(\varepsilon_i),p] + (1-\delta_i)\log S[y_i,\lambda_i(\varepsilon_i),p]\right\} f(\varepsilon_i)d\varepsilon_i.$$

The integral does not exist in closed form, but there are two approaches that can be used to provide a satisfactory approximation. If ε_i has a standard normal distribution, the integral can be computed using Hermite quadrature. Alternatively, if ε_i has a distribution that can be simulated (such as the standard normal), the integral can be computed using Monte Carlo methods. The simulated log likelihood that is used here is

$$\log L = \sum_{i=1}^N \frac{1}{R}\sum_{r=1}^R \left\{\delta_i \log f[y_i,\lambda_i(\varepsilon_{ir}),p] + (1-\delta_i)\log S[y_i,\lambda_i(\varepsilon_{ir}),p]\right\}$$

where ε_{ir} is a random sample of R draws from the appropriate population.

The maximum simulated likelihood estimator for this model is available for the Weibull, exponential, lognormal, loglogistic and inverse Gaussian models. The command is

SURVIVAL ; Lhs = dependent variable, censoring indicator if any
; Rhs = covariates (should include one)
; Model = the desired model
; RPM ; Fcn = one(n)
; Pts = 1
[; Halton ; Pts = desired value of R] $

(The model is fit as a random parameter model with only a random constant term.)

To illustrate, we have fit the Weibull model with normally distributed heterogeneity.

SURVIVAL ; Lhs = logt ; Rhs = one,prod ; Model = Weibull
; Pds = 1 ; RPM ; Fcn = one(n) ; Output = 3 ; Halton $

```
-----------------------------------------------------------------------------
Random Coefficients  WiblSurv Model
Dependent variable                 LOGT
Log likelihood function       -97.28525
Sample is  1 pds and     62 individuals
Weibull duration model
Simulation based on    100 Halton draws
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
    PROD|   -9.33205***     2.95417    -3.16  .0016   -15.12212  -3.54199
        |Means for random parameters
Constant|    3.77976***      .13833    27.32  .0000     3.50864   4.05088
        |Scale parameters for dists. of random parameters
Constant|     .00782         .13208      .06  .9528    -.25105     .26670
        |Scale parameter for survival distribution
ScalParm|     .99218***      .12063     8.22  .0000      .75574   1.22861
--------+--------------------------------------------------------------------
```

E61.3.3 Weibull Survival Model With Gamma Heterogeneity

A modification of the Weibull (or exponential) model suggested by Hui (1991) is

$$S(t|\mathrm{v}) = v\{\exp[(-\lambda t)^p]\}.$$

The random variable, v, is the heterogeneity effect. We assume that v is distributed as gamma with parameters k and R;

$$f(v) \quad = \left[k^R / \Gamma(R)\right] e^{-kv} v^{R-1}.$$

If the duration model contains a constant, then no generality is lost by assuming that the mean of v is one. Thus,

$$\mathrm{E}[v] = k/R = 1 \text{ or } k = R.$$

Now, we find

$$S(t) \quad = \int_0^{\infty} vS(t|v)f(v)\mathrm{d}v.$$

The result for the Weibull model is

$$S(t) \quad = \left[1 + \theta(\lambda t)^p\right]^{-1/\theta}$$

and $$h(t) \quad = S(t)^{\theta} \text{ times Weibull hazard,}$$

where $\theta = 1/k$. The variance of v is $1/k$, so, $\theta = 0$ corresponds to the Weibull model. The further θ deviates from zero, the greater is the effect of the heterogeneity. The Weibull survival function emerges if the limit of $S(t)$ as θ goes to zero is taken.

To request this variant of the Weibull model, use

SURVIVAL ; Lhs = ... ; Rhs = ...
; Model = Weibull
; Heterogeneity (or just ; Het) $

All other options and features described for the other parametric models apply equally. The only difference is that the BFGS algorithm is always used for estimation and the estimated covariance matrix for the parameter estimates is always the BHHH estimator. Since the exponential is a minor modification of the Weibull model, you may also specify **; Model = Exponential** in the command. The other models (loglogistic, normal, etc.) are not available with this specification.

NOTE: The log likelihood is somewhat volatile in the parameter θ. You may find the diagnostic 'Unable to compute function at current estimates' appearing in the output for your iterations. This means that the current trial value of θ is not positive. This is a recoverable error; *LIMDEP* will now try a new value.

The normal and gamma models are not directly comparable. The gamma model is not obtained by changing the normality assumption to the log gamma model in our formulation. The heterogeneity enters in a different form in the gamma model. The normal model turns out to be much easier to fit with these data, and appears to produce better results. The gamma model is actually inestimable, and we stopped the iterations at 12 to show the intermediate results. At 'convergence,' θ has gravitated to zero.

SURVIVAL ; Lhs = logt ; Rhs = one,prod
; Model = Weibull
; Het
; Output = 3
; Maxit = 25 $

This is the outcome when the estimator is allowed to iterate to completion:

```
Itr 12 F=  .9729D+02 gtHg=  .5460D+02 chg.F=  .8193D-04 max|db|=  .7173D+08
1st derivs.    -.45911D+00  -.22948D-01   .28633D+01   .13717D+00
Parameters:    -.37856D+01   .91768D+01   .24002D-06   .99166D+00
Itr 13 F=  .9729D+02 gtHg=  .6801D+01 chg.F=  .4096D-04 max|db|=  .1869D+08
1st derivs.     .00000D+00   .00000D+00   .00000D+00   .00000D+00
Parameters:    -.37856D+01   .91768D+01  -.16856D-07   .99166D+00
Itr 14 F=  .0000D+00 gtHg=  .0000D+00 chg.F=  .9729D+02 max|db|=  .0000D+00
                         * Converged                           * Converged
Normal exit:  14 iterations. Status=0, F=    .0000000
Function=  .10472246124D+03, at entry,  .00000000000D+00 at exit
```

The estimator claims convergence, but note that the estimate of θ is zero, and the derivatives matrix has essentially vanished – the BHHH estimator is singular, so estimation is halted. The results consist, within rounding error, of the original Weibull model plus a parameter that is zero.

```
-----------------------------------------------------------------------------
Loglinear survival model: WEIBULL
Dependent variable                 LOGT
Log likelihood function       -97.28851
Weibull Model with Gamma Heterogeneity
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    LOGT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |RHS of hazard model
Constant|    3.78562***      .28325    13.36  .0000     3.23046   4.34078
    PROD|   -9.17681***     3.07247    -2.99  .0028   -15.19874  -3.15488
        |Ancillary parameters for survival
   Theta| .24002D-06         .47115      .00 1.0000 -.92344D+00   .92344D+00
   Sigma|      .99166***     .25293     3.92  .0001      .49593   1.48739
--------+--------------------------------------------------------------------
```

E61.3.4 Other Heterogeneity Mixtures

A variety of base models and heterogeneity distributions are contained within *LIMDEP*'s menu of parametric models. The table below lists some of these.

Conditional Density	Heterogeneity	Unconditional Density	Estimator
gamma	gamma	generalized F(M1,M2)	**; Model = F**
gamma	exponential	generalized F(M1,1)	**; Model = F(M1,1)**
exponential	gamma	Pareto	**; Model = E ; Het**
exponential	exponential		**; Model = F(1,1)**
Weibull	gamma	Burr	**; Model = W ; Het**
Weibull	exponential	loglogistic	**; Model = L**

Table E61.1 Models for Heterogeneity

E61.4 Heterogeneity in the Scale Parameter for Loglinear Models

The parametric survival models, Weibull, loglogistic, lognormal, Weibull/gamma, inverse Gaussian and generalized F and their time varying covariates counterparts are specified so that the parameter $\lambda = \exp(-\boldsymbol{\beta}'\mathbf{x})$ makes the location of the survival distribution heterogeneous across individuals. The scale of the distribution, as specified by the parameter σ, is generally fixed for all individuals. The modification described here allows σ to be a function of individual specific covariates as well.

The loglinear specifications are defined in terms of a transformation,

$$w = (\log t - \boldsymbol{\beta}'\mathbf{x}) / \sigma$$

where σ is the scale parameter. You may specify the same sort of multiplicative heteroscedasticity as in the regression, tobit, logit, and probit models,

$$\sigma_i = \sigma \exp(\boldsymbol{\gamma}'\mathbf{h}_i)$$

where $\mathbf{h}_i$ is a vector of covariates. This extension is provided for the following models, all with or without time varying covariates or split population: Weibull, Weibull with gamma heterogeneity, loglogistic, lognormal, inverse Gaussian and generalized *F*. It is provided for the lognormal with or without the optional split population specification. To request this specification, just add

; Hfn = list of variables in $\mathbf{h}_i$

NOTE: The list must not contain *one*. The variance model already contains a constant, σ.

For hypothesis testing and providing starting values, note that with this extension, the parameter vector in the full model become

$$\theta = \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \theta, (M1,M2), \sigma.$$

Some parts are optional. If you do not specify the split population model, α will not be present, while the parameter θ is only present for the Weibull or exponential model with gamma heterogeneity or the generalized gamma model, and $(M1,M2)$ apply only to the generalized F model.

E61.5 Split Population Survival Models

The following describes a modification of the parametric survival models: Weibull (with or without gamma heterogeneity), lognormal, loglogistic, or exponential. (The models developed here are based on Schmidt and Witte (1989). The specification is not available for the Gompertz model.)

For analyzing survival time data with censoring indicator, δ_i (we use *LIMDEP*'s rather than Schmidt and Witte's notation), *LIMDEP*'s parametric survival models are based on the log likelihood

$$\log L = \Sigma_{\delta=1} \log\{(1/\sigma)f[\mu_i/\sigma]\} + \Sigma_{\delta=0} \log S(\mu_i/\sigma),$$

where, $\mu_i = \log t_i - \boldsymbol{\beta}'\mathbf{x}_i$

and, in all these models, $w_i = (\log t_i - \boldsymbol{\beta}'\mathbf{x}_i) / \sigma = \mu_i / \sigma$.

In the log likelihood, $(1/\sigma)f(w_i)$ is the density and $S(w_i)$ is the survival function (equal to one minus the CDF). The model as stated assumes that censored observations will all fail eventually. Schmidt and Witte suggest that the model be modified by allowing for the possibility that a censored observation might, in fact, never fail. Thus, they suggest that we model the probability of eventual failure as

$$\text{Prob}[R_i = 1] = P_i \quad \text{(our notation).}$$

Then, for an observed individual, the appropriate term which appears in the log likelihood is

$$\text{Prob}[R_i = 1] \times g(t_i \mid R_i = 1) + \text{Prob}[\delta_i = 0],$$

where $g(\bullet)$ is the original density above, and the probability attached to a censored observation ($\delta_i = 0$) is

$$\text{Prob}[R_i = 0] \text{ (i.e., never fail)} + \text{Prob}[R_i = 1] \times \text{Prob[fail at time t or later]}.$$

Let the determinants of the probability of eventual failure be $\mathbf{z}_i$ (in the Schmidt and Witte paper, $\mathbf{z}_i = \mathbf{x}_i$, which makes sense, though *LIMDEP* does not require it) and let

$$\text{Prob}[R_i = 1] = P_i = G(\boldsymbol{\alpha}'\mathbf{z}_i).$$

Combining terms, the revised log likelihood is now

$$\begin{aligned}\log L = \;& \Sigma_{\delta=1} \log\{ G(\boldsymbol{\alpha}'\mathbf{z}_i)(1/\sigma)f(w_i)\} \\ & + \Sigma_{\delta=0} \log\{[1 - G(\boldsymbol{\alpha}'\mathbf{z}_i)] + G(\boldsymbol{\alpha}'\mathbf{z}_i)S(w_i)\}.\end{aligned}$$

It remains to model $G(\boldsymbol{\alpha}'\mathbf{z}_i)$. Schmidt and Witte suggest a logistic model,

$$G(\boldsymbol{\alpha}'\mathbf{z}_i) = 1/[1 + \exp(\boldsymbol{\alpha}'\mathbf{z}_i)] = 1 - \Lambda(\boldsymbol{\alpha}'\mathbf{z}_i).$$

LIMDEP allows two models, the preceding logit model and a normal (probit) model,

$$G(\boldsymbol{\alpha}'\mathbf{z}_i) = 1 - \Phi(\boldsymbol{\alpha}'\mathbf{z}_i),$$

where $\Phi(\bullet)$ is the standard normal CDF. (There is no variance parameter, as it would not be identified in the model. The same principle as in the univariate probit model applies here.)

This augmented model is requested simply by adding

; Rh2 = list of variables in $\mathbf{z}_i$

The default model for $G(\bullet)$ is the normal model. Request Schmidt and Witte's logistic model by adding

; Logit

to the command with the **; Rh2** specification. No other changes are required in the command. This applies to **; Model = Weibull**, **; Model = Exponential**, **; Model = Normal**, or **; Model = Loglogistic** for the survival rate.

Schmidt and Witte discuss various models for G and the survival rate with and without individual effects. In your command, either (or both) of the models $S(\bullet)$ or $G(\bullet)$ may be specified to have just a constant term, *one* or may have covariates. As noted, there is no requirement that the Rhs and Rh2 models be the same. For example, to estimate their 'SPLIT' models (their Table 2, page 153), just specify **; Rh2 = one** and the rest of the command as usual. Likewise, by specifying both **; Rhs = one** and **; Rh2 = one**, you would obtain the models described on their page 147.

The additional output for this model consists of a header which displays the specification requested and an additional set of coefficients in the statistical output. In the table which is given after the coefficient estimates, there will now appear an additional row with an estimate of the average value of the probability of eventual failure. This is labeled 'SPLIT' in the table.

WARNING: This model is a bit quirky. If the model does not have much explanatory power, and if the censoring indicator is not explained very well by both the duration variable (*logtime*) and the covariates in the duration model, then the estimated probability of eventual failure will tend to gravitate toward 1.0 (as one might expect). This will show up in the model as extreme values of the coefficients in the equation for $G(\bullet)$. When this happens, the other coefficients will be identical to those which would be estimated if the 'splitting' model were ignored (i.e., as if you had not included the **; Rh2** specification). The model reported will appear to show coefficients in the $G(\bullet)$ equation, but it will not be possible to compute standard errors, and, in fact, the coefficients themselves will not be usable. If this occurs, the coefficients which were computed will be reported, along with zeros for their standard errors. A message will be given that the covariance matrix is singular (which it is). *LIMDEP* then uses a generalized inverse to invert the nonsingular submatrix.

E61.6 Left and Right Truncation

There are cases in which the natural limit point of zero is not actually appropriate for the duration data in hand. Consider, for example, an experiment in which duration measurement did not even begin until a certain amount of time had passed. The actual distribution of observed survival times will logically be constrained to some range other than zero to infinity. Presumably the truncation point will be somewhere above zero. In order to accommodate such a situation, the survival distribution, which is normally defined over $[0,+\infty)$, must be scaled up so that it integrates to one over the appropriate range.

Accounting for truncation can bring drastic changes in the estimated distribution. The relevant theory is exactly that underlying the truncated regression model. To account for truncation, modify the model command to

```
SURVIVAL    ; Lhs = ...
            ; Rhs = ...
            ; Model = ...
            ; other options if any
            ; Limit = limit point $
```

The default is lower (left) truncation. Specify upper truncation, instead, with

```
            ; Limit = limit ; Upper
```

(I.e., your data may be such that the observation is not observed if *T* exceeds a certain value. Consider, for example, observed failure times for a product with a warranty period of a fixed length.) The limit may be a constant or the name of a variable, if the truncation varies by observation. The model is otherwise unchanged. A header at the beginning of the output for the model will echo the specification of a model with truncation. But, there will be no further mention of the fact, since subsequent changes are all internal. The following restrictions apply:

- The model must be one of Weibull, exponential, lognormal, or loglogistic. This option is not available for the gamma, inverse Gaussian, generalized F, Gompertz, or the split population models.

- Newton's method is not available for this model. If you prefer a Newton-like method, you may still use BHHH. Note, though, that this extension makes calculation much more difficult. We have had our best success with BFGS.

E61.7 Sample Selection

The random parameters treatment noted above opens the possibility of a sample selection model for parametric survival models. The structure would be the base case parametric model, as modified above, using the Weibull model as the standard case,

$$h(t_i) = \lambda_i\, p\, (\lambda_i t_i)^{p-1}$$

where $$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma v_i)$$

We accommodate this case by treating the random component as a random constant term in the parametric model. The observation mechanism is now

$$d_i^* = \boldsymbol{\alpha}'\mathbf{z}_i + \varepsilon_i, \quad d_i = 1(d_i^* > 0)$$

where the correlation between v_i and ε_i is ρ. We assume that the data for the duration model are only observed when $d_i = 1$. The model is fit by full information maximum likelihood. (This means that there is no ‘lambda,’ the familiar inverse Mills ratio, added to the duration model. That treatment is only appropriate for the linear model fit by two step least squares.)

This model is requested by the following command set:

```
PROBIT      ; Lhs = d
            ; Rhs = variables in z
            ; Hold $
SURVIVAL    ; Lhs = logt [, and possibly a censoring indicator]
            ; Rhs = variables in x
            ; Model = one of Weibull, Normal, Loglogistic
            ; Selection
            ; RPM ; Fcn = one(n) $
```

(Other controls for the random parameters models, such as the number of replications, Halton draws, and so on, operate as with all other random parameters specifications.)

E62: Stochastic Frontier Models and Efficiency Analysis

E62.1 Introduction

Chapters E62-E65 present *LIMDEP*'s programs for two types of efficiency analysis, stochastic frontier analysis (SFA) and data envelopment analysis (DEA). To a large extent, these are competing methodologies. No formulation has yet been devised that unifies the two in a single analytical framework. Arguably, the former is a fully parameterized model whereas the latter is 'nonparametric,' albeit also atheoretical in nature.

The stochastic frontier model is used in a large literature of studies of production, cost, revenue, profit and other models of goal attainment. The model as it appears in the current literature was originally developed by Aigner, Lovell, and Schmidt (1977). The canonical formulation that serves as the foundation for other variations is their model,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u,$$

where y is the observed outcome (goal attainment), $\boldsymbol{\beta}'\mathbf{x} + v$ is the optimal, frontier goal (e.g., maximal production output or minimum cost) pursued by the individual, $\boldsymbol{\beta}'\mathbf{x}$ is the deterministic part of the frontier and $v \sim N[0,\sigma_v^2]$ is the stochastic part. The two parts together constitute the 'stochastic frontier.' The amount by which the observed individual fails to reach the optimum (the frontier) is u, where

$$u = |U| \text{ and } U \sim N[0,\sigma_u^2]$$

(change to $v + u$ for a stochastic cost frontier or any setting in which the optimum is a minimum). In this context, u is the 'inefficiency.' This is the normal-half normal model which forms the basic form of the stochastic frontier model.

Many varieties of the stochastic frontier model have appeared in the literature. A major survey that presents an extensive catalog of these formulations is Kumbhakar and Lovell (2000). (See, as well, Bauer (1990), Greene (2008) and several other surveys, many of which are cited in Kumbhakar and Lovell and in Greene.) The estimator in *LIMDEP* computes parameter estimates for most single equation cross section and panel data variants of the stochastic frontier model.

A large number of variants of the stochastic frontier model based on different assumptions about the distribution of the 'inefficiency' term, u have been proposed in the received literature. Most of these are available in *LIMDEP*, as suggested in the list below. The bulk of the received technology centers on cross section style modeling. However, recent advances include many extensions that take advantage of the features of panel data. A large array of panel data estimators are also supported by *LIMDEP* as well.

The conventional approach to deterministic frontier estimation is currently data envelopment analysis. This is usually handled with linear programming techniques. The analysis assumes that there is a frontier technology (in the same spirit as the stochastic frontier production model) that can be described by a piecewise linear hull that envelopes the observed outcomes. Some (efficient) observations will be on the frontier while other (inefficient) individuals will be inside. The technique produces a deterministic frontier that is generated by the observed data, so by construction, some individuals are 'efficient.' This is one of the fundamental differences between DEA and SFA. Data envelopment analysis is documented in Chapter E65.

The analysis of production, cost, etc. in the stochastic frontier framework involves two steps. In the first, the frontier model is estimated, usually by maximum likelihood. In the second, the estimated model is used to construct measures of inefficiency or efficiency. Individual specific estimates are computed that provide the basis of comparison of firms either to absolute standards or to each other. The sections of this chapter develop several model forms used in the first step. Efficiency estimation, the second step, appears formally in Section E62.8. The general methodology is then used in the already developed specifications and with several proposed in the sections that follow, as well as in Chapters E63 and E64.

E62.2 Stochastic Frontier Model Specifications

The stochastic frontier model is

$$y = \boldsymbol{\beta}'\mathbf{x} + v\text{-}u,\ u = |U|.$$

In this area of study, unlike most others, estimation of the model parameters is usually not the primary objective. Estimation and analysis of the inefficiency of individuals in the sample and of the aggregated sample are usually of greater interest. This part of the development will present tools for estimation of inefficiency.

Typically, the production or cost model is based on a Cobb-Douglas, translog, or other form of logarithmic model, so that the essential form is

$$\log y = \boldsymbol{\beta}'\mathbf{x} + v - u$$

where the components of **x** are generally logs of inputs for a production model or logs of output and input prices for a cost model, or their squares and/or cross products. In this form, then, at least for relatively small variation, u represents the proportion by which y falls short of the goal, and has a natural interpretation as proportional or percentage inefficiency. The numerous examples below will demonstrate. Users are also referred to the various survey sources listed earlier.

The results one obtains are, of course, critically dependent on the model assumed. Thus, specification and estimation of model parameters, while perhaps of secondary interest, are nonetheless a major first step in the model building process. In nearly all received formulations, the random component, v, is assumed to be normally distributed with zero mean. In some models, v may be heteroscedastic. But, in either form, the large majority of the different frontier models that have been proposed result from variations on the distribution of the inefficiency term, u. The range of specifications examined in this chapter includes the following:

- Distributional assumptions: half normal, exponential, gamma
- Partially nonparametric frontier function
- Sample selection model

The following extensions are presented in Chapters E63:

- Truncated normal with nonzero, heterogeneous mean in the underlying *U*
- Heteroscedasticity in *v* and/or *u*
- Heterogeneity in the parameter of the exponential or gamma distribution
- Alvarez, Amsler, Orea and Schmidt's (2006) 'scaling model'
- Alvarez, Arias and Greene's (2006) model of fixed, latent management

A number of treatments for panel data are presented in Chapter E64.

E62.3 Basic Commands for Stochastic Frontier Models

The command for all specifications of the stochastic frontier model is

FRONTIER ; Lhs = y ; Rhs = one, ... ; … other specifications $

NOTE: *One* must be the first variable in the Rhs list in all model specifications.

The default specification is Aigner, Lovell and Schmidt's canonical *normal-half normal model*. The default form is a production frontier model,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, u = |U|.$$

That is, the right hand side of the equation specifies the *maximum* goal attainable. To specify a cost frontier model or other model in which the frontier represents a *minimum*, so that

$$y = \boldsymbol{\beta}'\mathbf{x} + v + u, u = |U|,$$

use

; Cost

This specification is used in all forms of the stochastic frontier model. As noted below, one additional specification you may find useful is

; Start = values for β, λ, σ.

(The meanings of the parameters are developed below.) ALS also developed the *normal-exponential model*, in which *u* has an exponential distribution rather than a half normal distribution. To request the exponential model, use

; Model = Exponential (or ; Model = E)

in the **FRONTIER** command. For this model, the parameters are (β,θ,σ_v). Further details appear below. There are also several model forms, and numerous modifications such as heteroscedasticity that are developed below.

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps ancillary parameters σ, λ, etc. with main parameter β vector in *b*.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Choice	uses choice based sampling (sandwich with weighting) estimated matrix.
; Cluster = spec	requests computation of the cluster form of corrected covariance estimator.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

E62.3.1 Predictions, Residuals and Partial Effects

Predicted values and ‘residuals’ for the stochastic frontier models are computed as follows: The same forms are used for cross section and panel data forms. The predicted value is $\boldsymbol{\beta}'\mathbf{x}$. (These are rarely useful in this setting.) The ‘residual’ is computed directly as

$$e_i = y_i - \hat{\boldsymbol{\beta}}'\mathbf{x}_i$$

This residual is usually not of interest in itself. It is, however, the crucial ingredient in the efficiency estimator discussed in Section E62.8. The estimator of u_i that we will use is computed by the Jondrow formula $E[u|v\text{-}u]$ or $E[u|v\text{+}u]$ if based on a cost frontier,

$$\hat{E}[u\,|\,\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ \varepsilon = v \pm u\ ,\ w = \varepsilon\lambda/\sigma,$$

$$\sigma = \sqrt{\sigma_v^2 + \sigma_u^2},\ \lambda = \frac{\sigma_u}{\sigma_v}.$$

In the JLMS formula, e_i is the estimator of ε_i. The formulas and computations are discussed in Section E62.8.

The frontier model is, save for its involved disturbance term, a linear regression model. The conditional mean in the model is

$$E[y_i|\mathbf{x}_i] = \boldsymbol{\beta}'\mathbf{x}_i - \mathrm{E}[u_i|\mathbf{x}_i].$$

In most cases, $\mathrm{E}[u_i|\mathbf{x}_i]$is not a function of $\mathbf{x}_i$, so the derivatives of $\mathrm{E}[y_i|\mathbf{x}_i]$ with respect to $\mathbf{x}_i$ are just $\boldsymbol{\beta}$. In other cases, we will consider, the conditional mean of u_i does depend on $\mathbf{x}_i$ or other variables, so the partial effects in the model might be more involved than this. Once again, however, these will usually not be of direct interest in the study. But, in all cases, $\hat{E}[u\,|\,\varepsilon]$ will be an involved function of $\mathbf{x}_i$ and any other variables that appear anywhere else in the model. We will examine the partial effects on the efficiency estimators in Section E62.8.

E62.3.2 Results Saved by the Frontier Estimator

The results saved by the frontier estimator are

Matrices: *b* = regression parameters, $\alpha,\boldsymbol{\beta}$
varb = asymptotic covariance matrix

Scalars: *sy*, *ybar*, *nreg*, *kreg*, and *logl*

Last Function: JLMS estimator of u_i.

Use **; Par** to add the ancillary parameters to these. The ancillary parameters that are estimated for the various models are as follows, including the scalars saved by the estimation program:

Half and truncated normal:	estimates λ, σ, saves *lmda* and $s = \sigma$,
Truncated normal:	same as half normal, estimates μ, saved as *mu*,
Exponential:	estimates θ, σ_v, saves *theta* and $s = \sigma_v$,
Heteroscedastic model:	average value of σ as s, average value of λ as *lmda*
Heterogeneity in mean:	estimates λ, σ, saves *lmda* and $s = \sigma$.

E62.4 Data for the Analysis of Frontier Models

We will use two data sets to illustrate the frontier estimators. The first, the data on U.S. airlines is a panel data set that we will use primarily for illustrating the stochastic frontier model. The second, the famous WHO data on health care attainment, will be used both for the stochastic frontier models and for the later work on data envelopment analysis.

E62.4.1 Data on U.S. Airlines

We will develop several examples in this section using a panel data set on the U.S. airline industry from the pre-deregulation period (airlines.dat). The observations are an unbalanced panel on 25 airlines. The original balanced panel data set contained 15 observations (1970-1984) on each of 25 airlines. Mergers, strikes and other data problems reduced the sample to the unbalanced panel of 256 observations The group sizes (number of firms) are 2 (4), 4(1), 7 (1), 9 (3), 10 (3), 11 (1), 12 (2), 13 (1), 14 (3) and 15 (6). The variables in the data set are

firm	= ID, 1,...,25	*year*	= 1970...1984	t = *year* - 1969 = 1,...,15	
cost	= total cost	*revenue*	= revenue	*output*	= total output
stage	= average stage length	*points*	= number of points served	*loadfct*	= load factor
cmtl	= materials cost	*mtl*	= materials quantity	*pm*	= price of material
cfuel	= fuel cost	*fuel*	= fuel quantity	*pf*	= fuel price
ceqpt	= equipment cost	*eqpt*	= equipment quantity	*pe*	= equipment price
clabor	= labor cost	*labor*	= labor quantity	*pl*	= labor price
cprop	= property cost	*property*	= property quantity	*pp*	= property price
k	= capital index	*pk*	= capital price index		

Transformed variables used in the examples are as follows:

lc	= log(*cost*)	*cn*	= *cost*/*pp*	*lcn*	= log(*cn*)
lpm	= log(*pm*)	*lpf*	= log(*pf*)	*lpe*	= log(*pe*)
lpl	= log(*pl*)	*lpp*	= log(*pp*)	*lpk*	= log(*pk*)
lpmpp	= log(*pm*/*pp*)	*lpfpp*	= log(*pf*/*pp*)	*lpepp*	= log(*pe*/*pp*)
lplpp	= log(*pl*/*pp*)	*lf*	= log(*fuel*)	*lm*	= log(*mtl*)
le	= log(*eqpt*)	*ll*	= log(*labor*)	*lp*	= log(*property*)
lq	= log(*output*)	*lq*2	= lq^2		

E62.4.2 World Health Organization (WHO) Health Attainment Data

The data used by the WHO in their 2000 *World Health Report* assessment of health care attainment by 191 countries have been used by many researchers worldwide both for developing frontier models and for analyzing health outcomes. The data are a panel of five years, 1993-1997, on health outcome data for 191 countries and a number of internal political units, e.g., the states of Mexico. The main outcome variables are *dale* and *comp* (an aggregate of such measures as efficiency and equity of health care delivery in the country). The main input variables are *hexp* and *educ*. A variety of other variables, listed below, were observed only in 1997. The following descriptive statistics apply to the entire data set of 840 observations:

Variable	**Mean**	**Std. Dev.**	**Description**
country	*	*	country number omitting internal units, 1...,191
year	*	*	year (1993-1997)
small	*	*	internal political unit, 0 for countries, else 1,...,6.
comp	75.0062726	12.2051123	composite health care attainment
dale	58.3082712	12.1442590	disability adjusted life expectancy
hexp	548.214857	694.216237	health expenditure per capita, PPP units
educ	6.31753664	2.73370613	educational attainment, years
oecd	.279761905	.449149577	OECD member country, dummy variable
gdpc	8135.10785	7891.20036	per capita GDP in PPP units
popden	953.119353	2871.84294	population density per square KM
gini	.379477914	.090206941	gini coefficient for income distribution
tropics	.463095238	.498933251	dummy variable for tropical location
pubthe	58.1553571	20.2340835	proportion of health spending paid by government
geff	.113293978	.915983955	World Bank government effectiveness measure
voice	.192624849	.952225978	World Bank measure of democratization

(The data were analyzed in Greene (2004a,b). Some of the variables, such as *popden* and *gdpc*, were augmented from other sources in these studies.) Although the data are a five year panel – a few countries were observed for fewer than five years – there is almost no cross year variation in any variable. (The proportion of total variation that is within groups is less than 1% for the four time varying variables.) We have created a cross section from these data as follows: First, we discarded the data on internal political units. We then averaged *comp, dale, hexp* and *educ* across the five years. We retained a sample of 191 cross sectional (country) units. The following command set creates the data set.

```
SAMPLE      ; 1-840 $
REJECT      ; small > 0 $
SETPANEL    ; Group = country ; Pds = ti $
RENAME      ; hc3 = educ $
CREATE      ; lpubthe = log(pubthe) $
CREATE      ; dalebar  = Group Mean(dale, Pds = ti) $
CREATE      ; compbar = Group Mean(comp, Pds  = ti) $
CREATE      ; educbar  = Group Mean(educ, Pds = ti) $
CREATE      ; hexpbar  = Group Mean(hexp, Pds = ti) $
CREATE      ; logdbar  = Log(dalebar) ; logcbar  = Log(compbar) $
CREATE      ; logebar  = Log(educbar) ; loghbar= Log(hexpbar) $
CREATE      ; loghbar2 = loghbar^2 $
REJECT      ; year # 1997 $
```

E62.5 Skewness of the OLS Residuals and Problems Fitting Stochastic Frontier Models

Before maximum likelihood estimation begins, the skewness of the OLS residuals in the regression of *y* on **x** is checked. Waldman (1982) has shown that when the OLS residuals are skewed in the wrong direction, a solution for the maximum likelihood estimator for the stochastic frontier model is simply OLS for the slopes and for σ_v^2 and 0.0 for σ_u^2. If this condition is found, a lengthy warning is issued. We emphasize, this is not a bug in the program, nor is it something to be 'fixed,' beyond changing the specification of the model or rethinking the stochastic frontier as the modeling platform. This is our single most frequently posed question, so we offer an application to demonstrate the effect. Consider the commands

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-500 $
CREATE      ; u = Abs(Rnn(0,2))
            ; v = Rnn(0,1)
            ; x = Rnn(0,1)
            ; y = x + v + u $
REGRESS     ; Lhs = y ; Rhs = one,x
            ; Res = e $
FRONTIER    ; Lhs = y ; Rhs = one,x $
KERNEL      ; Rhs = e $
```

The **CREATE** command generates *y* exactly according to the model, except note that *u* is not subtracted, it is added. Thus, we should expect this model to perform poorly. The estimation results from the **FRONTIER** command are shown below. Note the string of warnings. Estimation is allowed to proceed, but the results are not a 'frontier' as such. The final estimate of λ is essentially zero, with a huge standard error and the reported estimate of σ_u^2 in the box above the results is 0.0000. The other estimates are, in fact, the same as OLS. The kernel density estimator for the OLS residuals is clearly skewed in the positive, that is, the wrong direction. Once again, we emphasize, this is a failure of the data to conform to the model.

```
  Error   315: Stoch. Frontier: OLS residuals have wrong skew. OLS is MLE.
WARNING! OLS residuals have the wrong skewness for SFM
Other forms of the model models may also behave poorly.
In this case, one MLE for the half normal model is OLS
for beta and sigma and zero for the inefficiency term.
Warning   141: Iterations:current or start estimate of sigma nonpositive
Warning   141: Iterations:current or start estimate of sigma nonpositive
Warning   141: Iterations:current or start estimate of sigma nonpositive
Warning   141: Iterations:current or start estimate of sigma nonpositive
Warning   141: Iterations:current or start estimate of sigma nonpositive
Line search at iteration 30 does not improve fn. Exiting optimization.
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                   Y
Log likelihood function      -921.33848
Estimation based on N =    500, K =   4
Inf.Cr.AIC  =   1850.7 AIC/N =    3.701
Variances: Sigma-squared(v)=    2.33375
           Sigma-squared(u)=     .00000
           Sigma(v)        =    1.52766
           Sigma(u)        =     .00000
Sigma = Sqr[(s^2(u)+s^2(v)]=    1.52766
Gamma = sigma(u)^2/sigma^2 =     .00000
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0         -921.33851
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .000
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    1.61107       165.2912      .01  .9922  -322.35365   325.57580
       X|    1.00746***      .07057    14.28  .0000      .86914    1.14578
        |Variance parameters for compound error
  Lambda| .10897D-05       135.6070      .00 1.0000 -.26578D+03  .26578D+03
   Sigma|    1.52766***      .00242   630.99  .0000     1.52292    1.53241
--------+--------------------------------------------------------------------
```

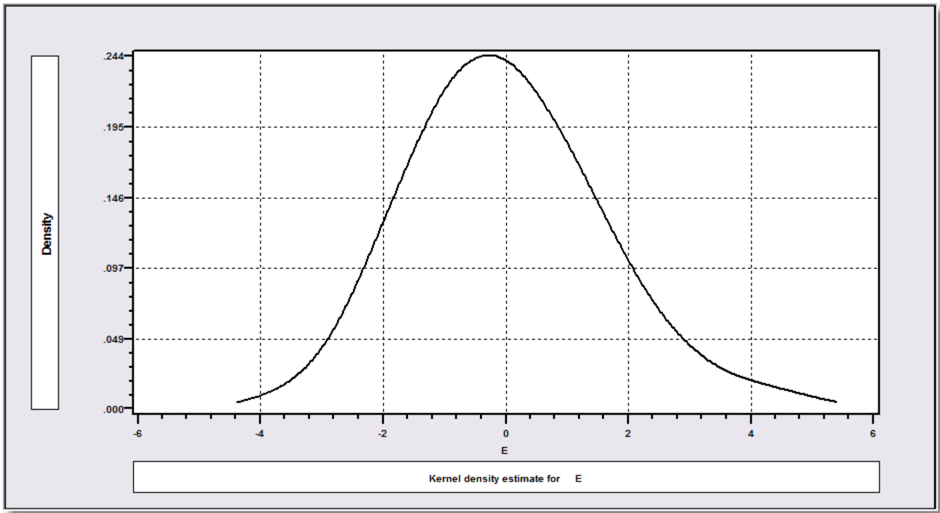


Figure E62.1 Kernel Density for Least Squares Residuals

Unfortunately, the Waldman result is a sufficient condition, not a necessary one. That is, it has been shown that when the OLS residuals have the 'right' skewness, then the MLE for the frontier model is unique, and you will have no trouble in estimation. When they have the 'wrong' skewness, it is only shown that the OLS results are a local stationary point of the log likelihood, not that they are the global maximizers. There may be another point that is yet better than OLS. Our airline data used below provide an example. Consider the following results, where we present both the stochastic frontier estimates and OLS. (The model, itself, is developed later, so we show only the useful results here.) As above, we receive the initial warning about the skewness of the OLS residuals. Then, estimation proceeds and an apparently routine solution emerges that is different from, and better than (has a higher log likelihood) OLS.

```
Error   315: Stoch. Frontier: OLS residuals have wrong skew. OLS is MLE.
WARNING! OLS residuals have the wrong skewness for SFM
Other forms of the model models may also behave poorly.
In this case, one MLE for the half normal model is OLS
for beta and sigma and zero for the inefficiency term.
Normal exit:  11 iterations. Status=0, F=   -105.0617
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                   LQ
Log likelihood function       105.06169
Variances: Sigma-squared(v)=     .02411
           Sigma-squared(u)=     .00457
           Sigma(v)        =     .15527
           Sigma(u)        =     .06757
Stochastic Production Frontier, e = v-u
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -1.05847***      .02333   -45.37  .0000    -1.10419  -1.01274
      LF|     .38355***      .07045     5.44  .0000      .24547    .52163
      LE|     .21961***      .07300     3.01  .0026      .07653    .36270
      LM|     .71667***      .07654     9.36  .0000      .56666    .86668
      LL|    -.41139***      .06382    -6.45  .0000     -.53647   -.28630
      LP|     .18973***      .02960     6.41  .0000      .13171    .24775
        |Variance parameters for compound error
  Lambda|     .43515**       .20117     2.16  .0305      .04086    .82944
   Sigma|     .16933***      .00057   295.74  .0000      .16821    .17045
--------+--------------------------------------------------------------------
Ordinary     least squares regression ............
Diagnostic   Log likelihood        =      105.05876
             Standard error of e   =         .16244
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error       t    |t|>T*        Interval
--------+--------------------------------------------------------------------
Constant|   -1.11237***      .01015  -109.57  .0000    -1.13227  -1.09247
      LF|     .38283***      .07116     5.38  .0000      .24335    .52231
      LE|     .21922***      .07389     2.97  .0033      .07441    .36404
      LM|     .71924***      .07732     9.30  .0000      .56769    .87078
      LL|    -.41015***      .06455    -6.35  .0000     -.53665   -.28364
      LP|     .18802***      .02980     6.31  .0000      .12961    .24643
--------+--------------------------------------------------------------------
```

There is no simple bullet proof strategy for handling this situation. You can try different starting values with **; Start = values for β, λ, σ** that differ from OLS, but it is hard to know where these will come from. Moreover, it is likely that you will end up at OLS anyway. As Waldman points out, this is a potentially ill behaved log likelihood function. We offer the preceding as a caution for the practitioner. For the particular data set used here, we can identify a specific culprit. The 'failure' of the model emerges in the presence of the variable *lm*, and does not occur when *lm* is omitted from the equation. We have no theory, however, for why this should be the case. Simply deleting variables from the model until one which does not have the skewness problem emerges does not seem like an effective strategy.

We do note, the failure might signal a misspecified model. For example, for our airlines example, the specification above omits the capital variable. When *lk* = log(*k*) is added to the model, we obtain the following quite routine results (albeit with the wrong signs on capital and labor inputs).

```
Normal exit:  13 iterations. Status=0, F=   -108.4392
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LQ
Log likelihood function        108.43918
Estimation based on N =    256, K =   9
Inf.Cr.AIC  =   -198.9 AIC/N =    -.777
Variances: Sigma-squared(v)=     .01902
           Sigma-squared(u)=     .01692
           Sigma(v)        =     .13791
           Sigma(u)        =     .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=     .18957
Gamma = sigma(u)^2/sigma^2 =     .47074
Var[u]/{Var[u]+Var[v]}     =     .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        | Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***      .72136    -4.14  .0000    -4.40206  -1.57439
      LF|     .37257***      .07038     5.29  .0000      .23463    .51052
      LE|    2.09473***      .68790     3.05  .0023      .74647   3.44299
      LM|     .69910***      .07580     9.22  .0000      .55054    .84766
      LL|    -.42909***      .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***      .09498     4.69  .0000      .25917    .63149
      LK|   -2.09806***      .76556    -2.74  .0061    -3.59853   -.59759
        | Variance parameters for compound error
  Lambda|     .94309***      .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***      .00064   297.81  .0000      .18832    .19082
--------+--------------------------------------------------------------------
```

We emphasize, the Waldman result, and this particular theoretical outcome, is specific to the normal-half normal model. However, when it occurs, problems of a similar sort will often, *but not always*, show up in other models. Thus, in spite of a warning, your fitted exponential, or panel data model, may be quite satisfactory.

E62.6 The Ordinary Least Squares Estimator

For the simplest specification

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, u = |U|$$

in which $\boldsymbol{\beta}$ contains a constant term and both v and U are homoscedastic and have zero means, i.e., in the original half normal or exponential models, the OLS estimator of all elements of $\boldsymbol{\beta}$ except the constant term are consistent. It is convenient to rewrite the model as

$$y = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}_1 + v - u.$$

Under the assumptions, we can write the model as

$$y = (\beta_0 - E[u]) + \boldsymbol{\beta}_1'\mathbf{x}_1 + v - (u - E[u])$$

or

$$y = \alpha + \boldsymbol{\beta}_1'\mathbf{x}_1 + e$$

in which e has zero mean and constant variance, and is orthogonal to $(1,\mathbf{x}_1)$. Thus, the model as shown can be estimated consistently by OLS. The constant term estimates $\alpha = (\beta_0 - E[u])$. Assuming that $E[u]$ is estimable, therefore, estimation of β by MLE vs. OLS is a question of efficiency, not consistency. (However, we remain interested in estimation of u, so this may be a moot point.)

E62.6.1 Corrected Ordinary Least Squares – COLS

The COLS estimator is obtained by turning the least squares estimator into a deterministic frontier model. This is done by shifting the intercept in the OLS estimator upward (for a production frontier) or downward (for a cost frontier) so that all points lie either below or above the estimated function. Figure E62.2 shows the result for estimation of a simple cost frontier for the airlines data. The function is shifted so that it rests on the single most extreme point (residual) in the data. The COLS estimator is requested with

FRONTIER ; Lhs = goal variable
; Rhs = one, …
; Model = COLS $

Add **; Cost** if the model is a cost frontier.

Efficiency values, as discussed below, are obtained as follows:

; Eff = variable name

saves the residuals from the deterministic frontier. These are the estimates of u_i. Note in Figure E62.2, for a cost frontier, all values of u_i are positive. If you fit a production frontier, then all points will lie below the regression and all residuals will be negative. The estimated inefficiency that is saved will be $-e_i$. Thus, in both cases, the values saved by **; Eff = variable** are the positive estimates of the size of the deviation of the observation from the frontier. The estimator saved by **; Eff = variable name** is the inefficiency estimate, in this model, a direct estimate of u_i. The estimator of technical or cost efficiency is

$$Efficiency = \exp(-\hat{u}_i)$$

If you fit a production frontier, use

; Techeff = variable name

to save this variable. For a cost frontier, use

; Costeff = variable name

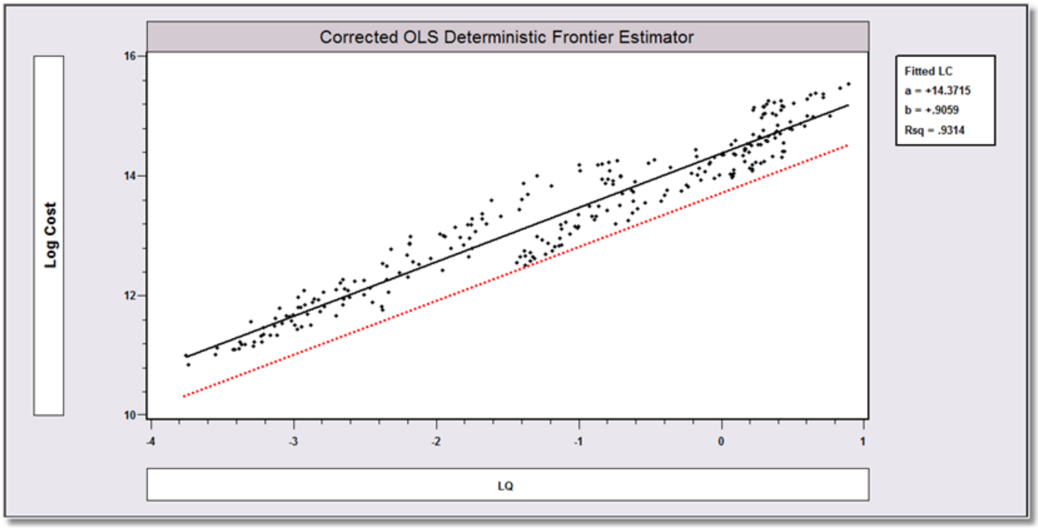


Figure E62.2 COLS Estimator of Cost Frontier Function

The following shows computation of a COLS estimator for the airlines. The **FRONTIER** command requests both the inefficiency estimates, *ui*, and the cost efficiency estimates, *eui_cost*. The kernel density estimate for the cost efficiency is shown in Figure E62.3. The results for the estimator begin with the standard output for least squares regression. The second panel includes some preliminary results for the stochastic frontier model, including the chi squared test for zero skewness (which is rejected); $\chi^2 = (n/6)(m_3/s^3)^2$. The standard normal statistic is the signed (based on m_3) square root of χ^2. The third panel presents descriptive statistics for u_i and $\exp(-u_i)$.

```
CREATE     ; lc   = Log(cost/pp)
           ; lpkp = Log(pk/pp)
           ; lplp = Log(pl/pp)
           ; lpmp= Log(pm/pp)
           ; lpep = Log(pe/pp)
           ; lpfp = Log(pf/pp) $
CREATE     ; lk   = Log(k) $
CREATE     ; ly   = Log(output) ; ly2 = .5*ly*ly $
FRONTIER   ; Lhs = lc ; Rhs = one,ly,ly2,lpkp,lplp,lpmp,lpep,lpfp
           ; Cost ; Model = COLS
           ; Costeff = Eui_cost ; Eff = ui $
KERNEL     ; Rhs = eui_cost
           ; Title = Estimated Cost Efficiency Based on COLS Estimator $
```

```
-----------------------------------------------------------------------------
Corrected OLS Deterministic Frontier Cost Function
LHS=LC       Mean                 =       2.84024
             Standard deviation   =       1.09256
             No. of observations  =           256  Degrees of freedom
Regression   Sum of Squares       =       300.028          7
Residual     Sum of Squares       =       4.36487        248
Total        Sum of Squares       =       304.393        255
             Standard error of e  =        .13267
Fit          R-squared            =        .98566  R-bar squared =    .98526
Model test   F[  7,   248]        =    2435.25310  Prob F > F*   =    .00000
Diagnostic   Log likelihood       =     157.91523  Akaike I.C.   = -4.00909
             Restricted (b=0)     =    -385.41031  Bayes  I.C.   = -3.89830
             Chi squared [  7]    =    1086.65108  Prob C2 > C2* =    .00000
--------------------------------------------------
Skewness test for inefficiency based  on residuals
Normalized skewness = m3/s^3      =        .21340
Chi squared test (1 degree of freedom)    1.94294  Critical value=  3.84000
Standard normal test statistic            1.39389  Test value = +/- 1.96000
Estimated Efficiency Values Based on e(i)+Min e(i)
--------+-----------------------------------------
        |     Mean      Std.Dev.   Minimum  Maximum
CostInef|     .357         .133       .000     .773
Cost Eff|     .706         .091       .462    1.000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LC|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Deterministic COLS Frontier Function
Constant|    19.4363      27.45697     .71  .4790    -34.3783   73.2510
      LY|     .94303***     .01809   52.12  .0000      .90757    .97849
     LY2|     .08248***     .01236    6.67  .0000      .05825    .10671
    LPKP|    1.42385       2.14849     .66  .5075    -2.78711   5.63480
    LPLP|     .01915        .10169     .19  .8506    -.18016     .21847
    LPMP|     .04504       1.41721     .03  .9746    -2.73264   2.82272
    LPEP|    -.57070        .67904    -.84  .4007    -1.90159    .76019
    LPFP|    -.04811**      .01986   -2.42  .0154    -.08704    -.00919
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
```

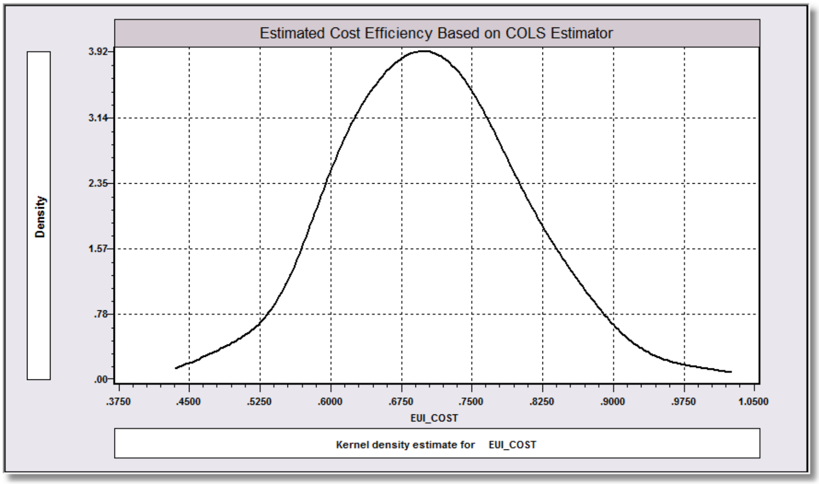


Figure E62.3 Kernel Estimator for Cost Efficiency

E62.6.2 Modified OLS and Starting Values for the MLE

Under the specific distributional assumptions of the half normal and exponential models, we do have method of moments estimators of the underlying parameters. They are based on the moment equations

$$\text{Var}[e] \;=\; \text{Var}[v] \;+\; \text{Var}[u]$$

and
$$\text{Skewness}[e] \;=\; \text{Skewness}[u]$$

since v is symmetric. The left hand sides can be consistently estimated using the OLS residuals:

$$m_2 \;=\; (1/n)\Sigma_i\, e_i^2$$

and
$$m_3 \;=\; (1/n)\Sigma_i e_i^3.$$

Both of the functions on the right hand side are known for the half normal and exponential models. In particular, for the half normal model, the moment equations are

$$m_2 \;=\; \sigma_v^2 \;+\; [1 - 2/\pi]\sigma_u^2\,,$$

$$m_3 \;=\; (2/\pi)^{1/2}[1 - 4/\pi]\sigma_u^3.$$

The solutions are:
$$\hat{\sigma}_u = \left[\frac{m_3\sqrt{\pi/2}}{1-4/\pi}\right]^{1/3} \text{ and } \hat{\sigma}_v = \sqrt{m_2 - (1-2/\pi)\hat{\sigma}_u^2}\,.$$

Note that there is no solution for σ_u if m_3 is not negative, which is the problem discussed in Section E62.5. Assuming that this problem does not arise, the corrected constant term is

$$\hat{\alpha} = a + \text{Est.}E[u] \;=\; a \;+\; \hat{\sigma}_u\sqrt{2/\pi}\,.$$

This is the 'modified least squares' (MOLS) estimator that is discussed in a number of sources, such as Greene (2005). These are the values used for starting values for the MLE, as well. Looking ahead, note that there is no natural method of moments estimator for the mean parameter in the truncated normal model discussed in Section E63.3. For this model, we use

$$\hat{\mu}/\sigma_u \;=\; 0.$$

For the normal-exponential model, the moment equations that correspond to the preceding are

$$m_2 \;=\; \sigma_v^2 \;+\; 1/\theta^2$$

$$m_3 \;=\; -2/\theta^3.$$

Therefore,
$$\hat{\theta} = \left[\frac{-2}{m_3}\right]^{1/3} \text{ and } \hat{\sigma}_v = \sqrt{m_2 - 1/\hat{\theta}^2}$$

and
$$\hat{\alpha} = a \;+\; 1/\hat{\theta}\,.$$

The header information in the results table will display the decomposition of the variance of the composed error in two parts. In the case of the half normal model,

$$\text{Var}[u] = [(\pi\text{-}2)/\pi]\sigma_u^2$$

not σ_u^2. Therefore, the estimated parameters might be a bit misleading as to the relative influence of u on the total variation in the structural disturbance.

We note, these estimators are sometimes quite far from the maximum likelihood estimators, particularly when the sample is small. But, they are generally quite satisfactory as starting values for the MLE. The following demonstrates these results for the airline data, where we use MOLS and MLE to fit a normal-half normal cost frontier. (Note, the signs of the OLS residuals are reversed because we are fitting a cost function.) In the results below, we have imposed the assumption of linear homogeneity in prices in the cost function by normalizing the six input prices, *pk*, *pl*, *pe*, *pp*, *pm*, *pf*, by the property price, *pp*. The model contains $\log(p_j/p_p)$. To complete the constraint, we have also normalized total cost by p_p before taking logs.

```
CREATE      ; lpk = Log(pk) $
CREATE      ; lpmpp = lpm - lpp ; lpfpp = lpf - lpp ; lpepp = lpe - lpp
            ; lplpp = lpl - lpp ; lpkpp = lpk - lpp $
CREATE      ; lcp = lc - lpp $
NAMELIST    ; x = one,ly,ly2,,lpkp,lplp,lpmp,lpep,lpfp $
REGRESS     ; Lhs = lc ; Rhs = x ; Res = e $
CREATE      ; e = -e ; e2 = e*e ; e3 = e2*e $
CALC        ; m2 = Xbr(e2) ; m3 = Xbr(e3) $
CALC        ; List ; su = (m3 * Sqr(pi/2) / (1-4/pi))^(1/3)
            ; sv = Sqr(m2 - (1-2/pi) * su^2)
            ; a = b(1) + su * Sqr(2/pi) ; lambda = su/sv
            ; sgma = Sqr(su^2 + sv^2) $
FRONTIER    ; Lhs = lc ; Rhs = x ; Cost $
```

The first set of results below are the OLS estimates with the correction to the constant term and the method of moments estimators of σ_u and σ_v used to start the MLE. The maximum likelihood estimators are shown next. The estimates for the stochastic frontier model include the log likelihood and the implied estimates of σ_u, σ_v and their squares, based on the estimates of $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$, which are estimated by ML. (The reverse transformations are $\sigma_u^2 = \sigma^2\lambda^2/(1 + \lambda^2)$ and $\sigma_v^2 = \sigma^2/(1 + \lambda^2)$. The MLE is documented further in the next section.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LC       Mean                 =       2.84024
             Standard deviation   =       1.09256
             No. of observations  =           256  Degrees of freedom
Regression   Sum of Squares       =       300.028           7
Residual     Sum of Squares       =       4.36487         248
Total        Sum of Squares       =       304.393         255
             Standard error of e  =        .13267
Fit          R-squared            =        .98566  R-bar squared =   .98526
Model test   F[  7,   248]        =    2435.25310  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     157.91523  Akaike I.C.   = -4.00909
             Restricted (b=0)     =    -385.41031  Bayes  I.C.   = -3.89830
             Chi squared [  7]    =    1086.65108  Prob C2 > C2* =   .00000
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|    19.7932       27.45697      .72  .4717    -34.0214   73.6079
      LY|     .94303***      .01809    52.12  .0000      .90757    .97849
     LY2|     .08248***      .01236     6.67  .0000      .05825    .10671
    LPKP|    1.42385        2.14849      .66  .5081    -2.78711   5.63480
    LPLP|     .01915         .10169      .19  .8508     -.18016    .21847
    LPMP|     .04504        1.41721      .03  .9747    -2.73264   2.82272
    LPEP|    -.57070         .67904     -.84  .4015    -1.90159    .76019
    LPFP|    -.04811**       .01986    -2.42  .0161     -.08704   -.00919
--------+--------------------------------------------------------------------

[CALC] SU      =       .1296481
[CALC] SV      =       .1046056
[CALC] A       =     19.8966785
[CALC] LAMBDA  =      1.2393989
[CALC] SGMA    =       .1665862
Calculator: Computed   5 scalar results
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LCN
Log likelihood function        159.20743
Estimation based on N =    256, K =  10
Inf.Cr.AIC  =   -298.4 AIC/N =   -1.166
Variances: Sigma-squared(v)=      .01021
           Sigma-squared(u)=      .01890
           Sigma(v)        =      .10103
           Sigma(u)        =      .13746
Sigma = Sqr[(s^2(u)+s^2(v)]=      .17059
Gamma = sigma(u)^2/sigma^2 =      .64927
Var[u]/{Var[u]+Var[v]}     =      .40216
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.584
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LCN|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    19.8020       25.91115      .76  .4447    -30.9829   70.5869
      LY|     .95577***      .01781    53.68  .0000      .92088    .99067
     LY2|     .09086***      .01198     7.58  .0000      .06738    .11435
    LPKP|    1.43400        2.02750      .71  .4794    -2.53982   5.40783
    LPLP|     .01242         .09676      .13  .8979     -.17722    .20205
    LPMP|     .05744        1.33747      .04  .9657    -2.56396   2.67883
    LPEP|    -.56860         .64356     -.88  .3770    -1.82995    .69275
    LPFP|    -.06002***      .01993    -3.01  .0026     -.09907   -.02096
        |Variance parameters for compound error
  Lambda|    1.36059***      .20306     6.70  .0000      .96261   1.75857
   Sigma|     .17059***      .00058   294.50  .0000      .16946    .17173
--------+--------------------------------------------------------------------
```

E62.7 Estimating the Normal-Half Normal and Normal-Exponential Models

ALS's canonical form of the model is the *normal-half normal model*,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - Su,\ u = |U|,\ S = +1 \text{ for production, -1 for cost,}$$
$$U \sim N[0,\sigma_u^2],$$
$$v \sim N[0,\sigma_v^2].$$

The command for estimating the stochastic frontier model is

FRONTIER ; Lhs = y ; Rhs = one, ... $

The default form is the normal-half normal model. In this form, model estimates consist of $\boldsymbol{\beta}$, $\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$ and $\lambda = \sigma_u/\sigma_v$, and the usual set of diagnostic statistics for models fit by maximum likelihood. The other basic form in the ALS model is the exponential model,

$$u \sim \theta \exp(-\theta u),\ u \geq 0,$$

which has mean inefficiency $E[u] = 1/\theta$ and standard deviation, $\sigma_u = 1/\theta$. The parameters estimated in the exponential specification are $(\boldsymbol{\beta},\theta,\sigma_v)$. The estimate of σ_u is reported in the results as well.

The following illustrate the estimator, with a normal-half normal cost frontier and a normal-exponential production frontier. The coefficient estimates for the exponential cost frontier are shown as well.

FRONTIER ; Cost ; Lhs = lcn ; Rhs = x $
FRONTIER ; Cost ; Lhs = lcn; Rhs = x; Model = Exponential $

The stochastic frontier results include the standard output for MLEs The derived estimates of σ_u, σ_v, σ_u^2, σ_v^2 and σ are shown as well. The value of $\gamma = \sigma_u^2/\sigma^2$ is given for comparability with other parts of the literature. This ratio, which lies in (0,1) is sometimes reported as a variance decomposition of ε. However, the variance of $u = |U|$ is $(1 - 2/\pi)\sigma_u^2$, so the appropriate decomposition is $(1 - 2/\pi)\sigma_u^2/[\sigma_v^2 + (1 - 2/\pi)\sigma_u^2]$. This is the value shown next under γ in the results.

A likelihood ratio test against the hypothesis of no inefficiency follows the variance estimates. The degrees of freedom for the test are accumulated in the table.. The first is for σ_u in the base case. The second is for the heteroscedasticity terms in Var[u] when they are introduced in the model. Heteroscedasticity is developed in Chapters E63. The third term is for the truncation parameters in the normal-truncated normal model, also developed in the next chapter. The 'degrees of freedom for the inefficiency model' are the sum of these three terms. The likelihood ratio statistic is presented next. This is a nonstandard test because the null value of σ_u is on the boundary of the parameter space. Appropriate tables for the mixed chi squared test used here are given in Kodde and Palm (1986). (A copy of the relevant parts of the table is kept internally by the program. (See, also, Coelli, Rao and Battese (1998) for further details.)

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LCN
Log likelihood function       159.20743
Estimation based on N =    256, K =  10
Inf.Cr.AIC  =   -298.4 AIC/N =   -1.166
Variances: Sigma-squared(v)=     .01021
           Sigma-squared(u)=     .01890
           Sigma(v)        =     .10103
           Sigma(u)        =     .13746
Sigma = Sqr[(s^2(u)+s^2(v)]=     .17059
Gamma = sigma(u)^2/sigma^2 =     .64927
Var[u]/{Var[u]+Var[v]}     =     .40216
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.584
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LCN|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    19.8020      25.91115      .76  .4447    -30.9829   70.5869
      LY|     .95577***     .01781    53.68  .0000      .92088    .99067
     LY2|     .09086***     .01198     7.58  .0000      .06738    .11435
    LPKP|    1.43400       2.02750      .71  .4794    -2.53982   5.40783
    LPLP|     .01242        .09676      .13  .8979     -.17722    .20205
    LPMP|     .05744       1.33747      .04  .9657    -2.56396   2.67883
    LPEP|    -.56860        .64356     -.88  .3770    -1.82995    .69275
    LPFP|    -.06002***     .01993    -3.01  .0026     -.09907   -.02096
        |Variance parameters for compound error
  Lambda|    1.36059***     .20306     6.70  .0000      .96261   1.75857
   Sigma|     .17059***     .00058   294.50  .0000      .16946    .17173
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Results for the normal-exponential model appear below. It is not possible to use a LR test to choose between these two models. The test has zero degrees of freedom – neither model is obtained by a restriction on the other. One possibility might be a Vuong (1989) statistic, which would be computed as

$$V = \frac{\sqrt{n}\ \overline{m}}{s_m},\ m_i = \log(f_i \mid normal) - \log(f_i \mid exponential)\ .$$

Results of the test are shown below the model results. The statistic is well inside the inconclusive region.

```
--------------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LCN
Log likelihood function       159.89917
Estimation based on N =    256, K =  10
Inf.Cr.AIC  =   -299.8 AIC/N =   -1.171
Exponential frontier model
Variances: Sigma-squared(v)=     .01147
           Sigma-squared(u)=     .00568
           Sigma(v)        =     .10709
           Sigma(u)        =     .07539
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =    3.968
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+-----------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     LCN|  Coefficient      Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    22.6569       25.48354      .89  .3740    -27.2899   72.6038
      LY|     .96069***      .01892    50.77  .0000      .92360    .99777
     LY2|     .09281***      .01249     7.43  .0000      .06832    .11729
    LPKP|    1.65439        1.99409      .83  .4067    -2.25395   5.56272
    LPLP|    -.00962         .09785     -.10  .9217     -.20140    .18216
    LPMP|    -.06595        1.31569     -.05  .9600    -2.64465   2.51275
    LPEP|    -.62841         .63243     -.99  .3204    -1.86795    .61114
    LPFP|    -.06397***      .02033    -3.15  .0017     -.10381   -.02412
        |Variance parameters for compound error
   Theta|    13.2651***     2.90719     4.56  .0000      7.5671   18.9630
  Sigmav|     .10709***      .00980    10.93  .0000      .08788    .12629
--------+-----------------------------------------------------------------------
```

```
FRONTIER    ; ... half normal model $
CREATE      ; fn = logl_obs $
FRONTIER    ; ...  Model = Exponential $
CREATE      ; fe = logl_obs
            ; mi = fn - fe $
CALC        ; List
            ; vuong = Sqr(n) * Xbr(mi)/Sdv(mi) $
```

```
[CALC] VUONG   =      -.9047927
```

E62.7.1 Log Likelihoods for the Half Normal and Exponential Models

As will be evident below, different formulations of the log likelihood are most convenient for estimation of the different forms of the frontier models. (And, different authors sometimes parameterize the models differently.) The base case is the normal-half normal model. In this form, $v_i \sim \mathrm{N}[0,\sigma_v^2]$ and $u_i = |U_i|$ where $U_i \sim \mathrm{N}[0,\sigma_u^2]$. It follows that $f(u_i) = 2\phi(u_i/\sigma_u)$, $u_i \geq 0$. The density of $\varepsilon_i = v_i - u_i$ has been shown to be

$$f(\varepsilon_i) = (2/\sigma)\phi(\varepsilon_i/\sigma)\Phi(-\varepsilon_i\lambda/\sigma).$$

The most common form of the individual term in the log likelihood function (and the one used in *LIMDEP*) is

$$\log L_i = \tfrac{1}{2}\log(2/\pi) - \log\sigma - \tfrac{1}{2}(\varepsilon_i/\sigma)^2 + \log\Phi[-S\varepsilon_i\lambda/\sigma]$$

where
$$\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i$$
$$\lambda = \sigma_u / \sigma_v,$$
$$\sigma^2 = \sigma_u^2 + \sigma_v^2,\ \sigma_v^2 = \sigma^2/(1+\lambda^2),\ \sigma_u^2 = \sigma^2\lambda^2/(1+\lambda^2)$$
$$S = +1 \text{ for production frontier, } -1 \text{ for cost frontier}$$

Olsen's transformation is used for maximizing the log likelihood. We reparameterize the function in terms of $\eta = 1/\sigma$ and $\boldsymbol{\gamma} = (1/\sigma)\boldsymbol{\beta}$. Then,

$$\log L_i = \tfrac{1}{2}\log(2/\pi) + \log\eta + \tfrac{1}{2}\omega_i^2 + \log\Phi(-S\omega_i\lambda)$$

where
$$\omega_i = \eta y_i - \gamma'x_i.$$

Define the functions
$$a_i = -S\omega_i\lambda$$
$$\delta_i = \phi(a_i)/\Phi(a_i)$$
$$\Delta_i = -a_i\delta_i = \delta_i^2.$$

Then, the gradient and Hessian are

$$\partial\log L_i / \partial\begin{pmatrix}\boldsymbol{\gamma}\\ \eta\\ \lambda\end{pmatrix} = \omega_i\begin{pmatrix}\mathbf{x}_i\\ -y_i\\ 0\end{pmatrix} + \delta_i S\begin{pmatrix}\lambda\mathbf{x}_i\\ -\lambda y_i\\ \omega_i\end{pmatrix} + \begin{pmatrix}0\\ 1/\eta\\ 0\end{pmatrix}$$

$$\partial^2\log L_i / \partial\begin{pmatrix}\boldsymbol{\gamma}\\ \eta\\ \lambda\end{pmatrix}\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \eta\\ \lambda\end{pmatrix}' = -\begin{pmatrix}\mathbf{x}_i\mathbf{x}_i' & \mathbf{0} & 0\\ -y_i\mathbf{x}_i' & y_i^2 & 0\\ \mathbf{0}' & 0 & 0\end{pmatrix} +$$

$$\Delta_i\begin{pmatrix}\lambda^2\mathbf{x}_i\mathbf{x}_i' & -\lambda^2 y_i\mathbf{x}_i & -\lambda\omega_i\mathbf{x}_i\\ -\lambda^2 y_i\mathbf{x}_i' & \lambda^2 y_i^2 & \lambda\omega_i y_i\\ -\lambda\omega_i\mathbf{x}' & \lambda\omega_i y_i & \omega_i^2\end{pmatrix} - \begin{pmatrix}\mathbf{0} & \mathbf{0} & -\delta_i S\mathbf{x}_i\\ \mathbf{0}' & 1/\eta^2 & \delta_i S y_i\\ -\delta_i S\mathbf{x}_i' & \delta_i S y_i & 0\end{pmatrix}$$

The log likelihood for the exponential model is

$$\log L_i = \log\theta + ½\theta^2\sigma_v^2 + \theta S\varepsilon_i + \log\Phi[-S\varepsilon_i/\sigma_v - \theta\sigma_v].$$

The parameter θ in the exponential model is $1/\sigma_u$. The Olsen transformation is not useful for this model. Define $c_i = -S\varepsilon_i/\sigma_v - \theta\sigma_v$, $\delta_i = \phi(c_i)$, $\Delta_i = -c_i\delta_i - \delta_i^2$ and $a_i = S\varepsilon_i/\sigma_v - \theta$. The gradient and Hessian for the exponential model are

$$\partial \log L_i / \partial \begin{pmatrix} \boldsymbol{\beta} \\ \theta \\ \sigma_v \end{pmatrix} = \delta_i \begin{pmatrix} S\mathbf{x}_i / \sigma_v \\ -\sigma_v \\ S\varepsilon_i / \sigma_v^2 - \theta \end{pmatrix} + \begin{pmatrix} -\theta S\mathbf{x}_i \\ 1/\theta + \theta\sigma_v^2 + S\varepsilon_i \\ \theta^2\sigma_v \end{pmatrix}$$

$$\partial^2 \log L_i / \partial \begin{pmatrix} \boldsymbol{\beta} \\ \theta \\ \sigma_v \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \theta \\ \sigma_v \end{pmatrix}' = \Delta_i \begin{pmatrix} \mathbf{x}_i\mathbf{x}_i' / \sigma_v^2 & -S\mathbf{x}_i & a_i S\mathbf{x}_i / \sigma_v \\ -S\mathbf{x}_i' & \sigma_v^2 & -a\sigma_v \\ a_i S\mathbf{x}_i' / \sigma_v & -a_i\sigma_v & a_i^2 \end{pmatrix}$$
$$+ \begin{pmatrix} 0 & -S\mathbf{x}_i & -\delta_i S\mathbf{x}_i \\ -S\mathbf{x}_i' & -1/\theta^2 + \sigma_v^2 & 2\theta\sigma_v - \delta_i \\ -\delta_i S\mathbf{x}_i' & 2\theta\sigma_v - \delta_i & \theta^2 - 2\delta_i S\varepsilon_i / \sigma_v^3 \end{pmatrix}$$

E62.7.2 Alternative Parameterization

Some treatments of the normal-half normal model (e.g., Coelli (1996b)) use the alternative parameterization $\gamma = \sigma_u^2 / \sigma^2$ in the formulation of the log likelihood. This does not change the model, since it is a one to one transformation of the parameters;

$$\lambda = \sqrt{\frac{\gamma}{1-\gamma}}\,.$$

The parameterization in terms of λ is more convenient but does not produce different results.

E62.7.3 Variance Estimator in Frontier 4.1

A number of researchers have used Tim Coelli's (1996b) Frontier 4.1 program for estimation of stochastic frontier models. Frontier 4.1 and *LIMDEP* use different methods for computing estimators of the asymptotic covariance matrix of the ML estimator. *LIMDEP* uses either the BHHH estimator or the negative inverse of the Hessian. Frontier 4.1 used the weighting matrix used by the DFP algorithm to approximate the inverse Hessian during the iterations. As a general proposition, we recommend against this 'estimator,' and never use it. There is no theoretical assurance of its accuracy if convergence is reached in a finite number of iterations. Nonetheless, we have been asked about this many times. In the interest of methodological advance, *LIMDEP* provides a command switch,

; F41

that will invoke this estimator. (This is only provided for the stochastic frontier estimators.) No indication is given in the output that this option has been used.

E62.8 Estimating Inefficiency and Efficiency Measures

The main objectives of fitting the frontier models is to estimate the inefficiency terms in the stochastic model, u_i, by observation. The Jondrow estimator of $E[u|v\text{-}u]$ is the standard estimator. This is

$$\hat{E}[u\,|\,\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ \varepsilon = v \pm u\ ,\ w = S\varepsilon\lambda/\sigma.$$

(This is an indirect estimator of u. Unfortunately, it is not possible to estimate u_i directly from any observed sample information. The various surveys noted earlier discuss the computation of and properties of this estimator.) The counterpart for the normal-exponential model is

$$\hat{E}[u\,|\,\varepsilon] = \sigma_v\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ w = (S\varepsilon/\sigma_v + \theta\sigma_v).$$

These are computed and saved as new variables in your data set with

; Eff = variable name

The **; List** specification will also request a listing of this variable. This form is used for all distributions and all variations of the stochastic frontier model.

By adding **; Eff = u** to the frontier command, then

KERNEL ; Rhs = u $

we obtain the results below. (We also added the title to the command with **; Title = …**) Note an important element of the estimation. The 'Standard Deviation' reported below is 0.054895, whereas the estimate of σ_u is 0.13746. The difference arises because the 0.054895 is an estimate of the standard deviation of $E[u|\varepsilon]$, not the standard deviation of u.

```
+--------------------------------------+
| Kernel Density Estimator for U       |
| Observations       =           256   |
| Points plotted     =           256   |
| Bandwidth          =       .016298   |
| Statistics for abscissa values----   |
| Mean               =       .109394   |
| Standard Deviation =       .054895   |
| Minimum            =       .030722   |
| Maximum            =       .350422   |
| ----------------------------------   |
| Kernel Function    =      Logistic   |
| Cross val. M.S.E.  =       .000000   |
| Results matrix     =        KERNEL   |
+--------------------------------------+
```

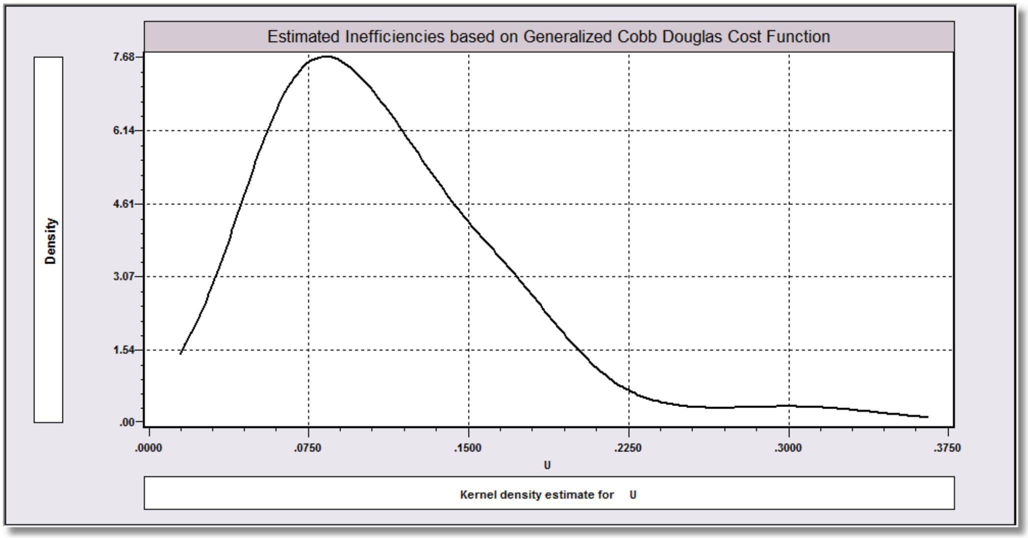


Figure E62.4 Analysis of Estimated Inefficiencies

E62.8.1 Estimating Technical or Cost Efficiency

One might be interested in estimating the 'efficiency' of the individuals in the sample. The model is usually specified in logs, of the form

$$\log y = \boldsymbol{\beta}'\mathbf{x} + v - u.$$

Under this assumption, the efficiency of the individual would be

$$EFF = \frac{y}{Optimal\ y} \approx \text{Exp}(-u)$$

This can be obtained with

; Techeff = the variable name

or **; Costeff = the variable name**

if you estimate a cost frontier instead. You may compute both inefficiencies and efficiency measures in the same command. Figure E62.5 was obtained by adding

; Costeff = ecu

to the **FRONTIER** command, then requesting the kernel density estimator as before (with the title changed accordingly).

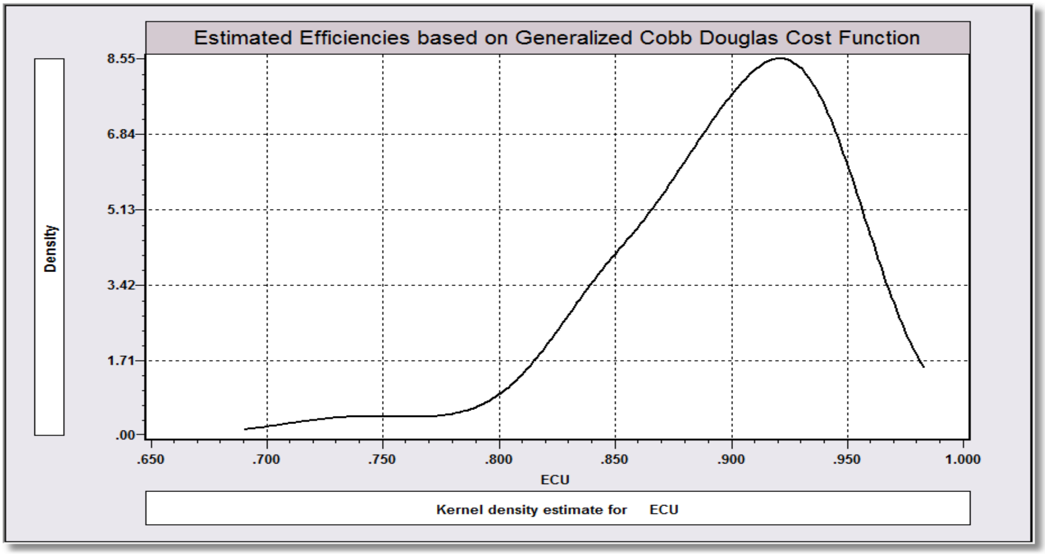


Figure E62.5 Estimated Cost Efficiencies

E62.8.2 Confidence Intervals for Inefficiency and Efficiency Estimates

Horrace and Schmidt (1996, 2000) suggest a useful extension of the Jondrow result. JLMS have shown that the distribution of $u_i|\varepsilon_i$ is that of a $N[\mu_i^*,\sigma^*]$ random variable, truncated from the left at zero, where $\mu_i^* = -\varepsilon_i\lambda^2/(1+\lambda^2)$ and $\sigma^* = \sigma\lambda/(1+\lambda^2)$. This result and standard results for the truncated normal distribution (see, e.g., Greene (2011)) can be used to obtain the conditional mean and variance of $u_i|\varepsilon_i$. With these in hand, one can construct some of the features of the distribution of $u_i|\varepsilon_i$ or $E[TE_i|\varepsilon_i] = E[\exp(-u_i|\varepsilon_i]$. The literature on this subject, including the important contributions of Bera and Sharma (1999) and Kim and Schmidt (2000) refer generally to ‘confidence intervals’ for $u_i|\varepsilon_i$. For reasons that will be clear shortly, we will not use that term – at least not yet, until we have made more precise what we are estimating.

For locating 100(1-α)% of the conditional distribution of $u_i|\varepsilon_i$, we use the following system of equations

$$\sigma^2 \;=\; \sigma_v^{\,2} + \sigma_u^{\,2}$$

$$\lambda \;=\; \sigma_u/\sigma_v$$

$$\mu_i^* \;=\; -\varepsilon_i\sigma_u^{\,2}/\sigma^2 = -\varepsilon_i\lambda^2/(1+\lambda^2)$$

$$\sigma^* \;=\; \sigma_u\sigma_v/\sigma = \sigma\lambda/(1+\lambda^2)$$

$$LB_i = \mu_i^* + \sigma^*\Phi^{-1}\left[1-(1-\tfrac{\alpha}{2})\Phi\left(\mu_i^*/\sigma^*\right)\right]$$

$$UB_i = \mu_i^* + \sigma^*\Phi^{-1}\left[1-\tfrac{\alpha}{2}\Phi\left(\mu_i^*/\sigma^*\right)\right]$$

Then, if the elements were the true parameters, the region $[LB_i,UB_i]$ would encompass 100(1-α)% of the distribution of $u_i|\varepsilon_i$. For constructing ‘confidence intervals’ for technical efficiency, $TE_i|\varepsilon_i$, it is necessary only to compute $TEUB_i = \exp(-LB_i)$ and $TELB_i = \exp(-UB_i)$.

We note two caveats about the estimator. First, the received papers based on classical methods have labeled this a *confidence interval* for u_i. However, it is a range that encompasses 100(1-α)% of the probability in the *conditional* distribution of $u_i|\varepsilon_i$. based on $E[u_i|\varepsilon_i]$, not u_i, itself. The interval is 'centered' at the estimator of the conditional mean, $E[u_i|\varepsilon_i]$, not the estimator of u_i, itself, as a conventional 'confidence interval' would be. The estimator is actually characterizing the conditional distribution of $u_i|\varepsilon_i$, not constructing any kind of interval that brackets a particular u_i – that is not possible. Second, these limits are conditioned on known values of the parameters, so they ignore any variation in the parameter estimates used to construct them. Thus, we regard this as a minimal width interval.

You can request computation of these lower and upper bounds by adding

; CI(100(1 - α)) = lower, upper

where 100(1-α) is one of 90, 95, or 99 and *lower*, *upper* are names for two variables that will be created. You may use this feature with **; Eff = variable** or **; Techeff = variable** (or **; Costeff = variable** for a cost frontier). If you have both **; Eff** and **; Techeff** in the command, the confidence intervals are computed for **; Techeff**. (You can obtain the interval for **; Eff** in this case by computing the negatives of the logs with **CREATE**.)

We obtained these bounds for our cost function with

; Costeff = euc ; CI(95) = eucl,eucu

We followed the estimation with

PLOT ; Rhs = eucl,ecu,eucu
; Title = Upper and Lower Bound Estimates of Cost Efficiency
; Vaxis = Cost Efficiency$

to obtain Figure E62.6.

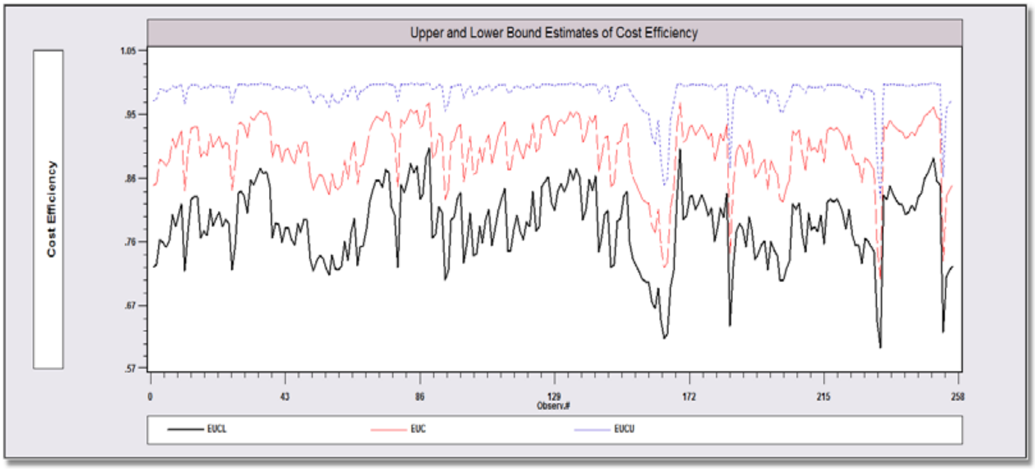


Figure E62.6 Lower and Upper Bound Estimates of Cost Efficiency

The centipede plot is also a useful device in this context. The following redraws Figure E62.6 using a different view for the lower and upper bounds

```
CREATE     ; Firm_i = Trn(1,1) $
PLOT       ; Lhs = firm_i ; Rhs = eucl,eucu
           ; Centipede ; Endpoints = 0,260 ; Grid
           ; Title = Confidence Limits for Cost Efficiency $
```

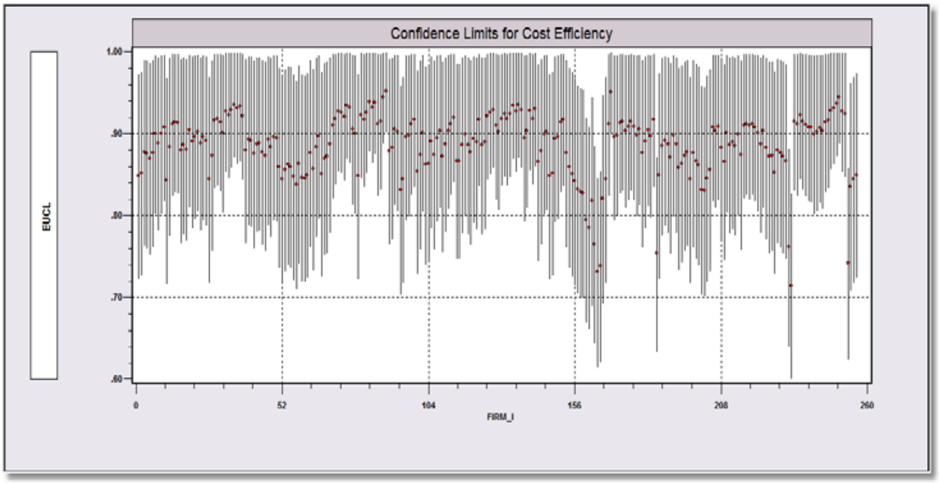


Figure E62.7 Centipede Plot of Efficiency Bounds

E62.8.3 Partial Effects on Efficiencies

The variables in the production or cost frontier function begin with either the inputs for the production model or input prices and outputs in the cost model. Analyses of how these variables affect technical or cost efficiency are not likely to be particularly revealing. However, if the function includes environmental variables (we call these z_i), it might be of interest to examine how variation in these impacts efficiency. For our example, we consider

$$\text{Log}(Cost/P_p) = \alpha + \beta_q \log Q + \beta_{qq} \log^2 Q + \Sigma_k\beta_k \log(P_k/P_p)$$
$$+ \delta_L \textit{load factor} + \delta_N \textit{nodes} + \delta_S \textit{Log stage length} + v + u$$

In this case, it might be interesting to examine how increased load factor, route complexity, or stage length impact efficiency.

Expressions for the technical inefficiency values appear at the beginning of Section E62.8. In those expressions, we will use

$$\textit{Efficiency} = \exp\{-\hat{E}[u\,|\,\varepsilon]\}.$$

The two expressions for the normal and exponential models are functions of a $w(\varepsilon)$ that is specific to the model. Each may be written as

$$\textit{Efficiency} = \exp\{-\tau_m A[w_m(\varepsilon)]\}$$

Where m = half normal or exponential, $\tau_m = \sigma\lambda/(1+\lambda^2)$ for the half normal and $1/\sigma_v$ for the exponential, and w_m is defined earlier. We now suppose that

$$\varepsilon = y - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\mathbf{z}$$

where **x** is the theoretical inputs to the goal and **z** are the environmental variables. We require the derivatives with respect to **z**. For convenience, let $W = -w$ and exploit the symmetry of the normal density. Then, $A[w_m(\varepsilon)] = [\phi(W)/\Phi(W) + W]$. The derivative is

$$\partial Efficiency/\partial\mathbf{z} = Efficiency \times -\tau_m \times dA(W)/dW \times -1 \times \partial w_m/\partial\varepsilon \times -\boldsymbol{\delta}.$$

The two terms that we need to complete the derivation are $\partial w_m/\partial\varepsilon = S\lambda/\sigma$ for the half normal model and S/σ_v for the exponential model and

$$\frac{dA(W)}{dW} = \left[1 - \frac{W\phi(W)}{\Phi(W)} - \left(\frac{\phi(W)}{\Phi(W)}\right)^2\right] = D(W).$$

Collecting terms,

$$\frac{\partial Efficiency}{\partial z} = Efficiency \times D(W) \times \begin{pmatrix} \lambda^2/(1+\lambda^2) \\ or \\ 1 \end{pmatrix} \times S \times (-\boldsymbol{\delta})$$

We can sign this result, though the magnitude will be empirical. The first three terms are all between zero and one, as is their product. S is either +1 for a production frontier or -1 for a cost frontier. Thus, in total, the derivative is a fraction of the corresponding coefficient, which takes the same sign for a cost frontier and the opposite sign for a production frontier.

Partial derivatives and simulations are computed with **PARTIALS** and **SIMULATE**. The general approach would be

```
FRONTIER    ; Cost (optional)
            ; Lhs = goal variable
            ; Rhs = one, x variables, z variables $
```

The command might also contain **; Eff = variable**, **; Techeff = variable** or **; Costeff = variable**. Then, you may follow it with

```
     PARTIALS    ; Effects: variables desired ; other options $
or   SIMULATE    ; Scenario … all options $
```

The function analyzed in these two commands is the technical or cost efficiency,

$$Efficiency = \exp\{-\hat{E}[u \mid \varepsilon]\}.$$

The following demonstrates using the cost frontier, with variables *z* = (*load factor*, *log stage length*, *points served*). Data on *z* are missing for one of the firms.

```
CREATE      ; logstage = Log(stage) $
NAMELIST    ; x = one,ly,ly2,,lpkp,lplp,lpmp,lpep,lpfp
            ; z = loadfctr,logstage,points $
FRONTIER    ; Cost ; Lhs = lc ; Rhs = x,z
            ; Eff = u ; Costeff = euc ; CI(95) = eucl,eucu $
SIMULATE    ; Scenario: & loadfctr = .4(.025)1 ; Plot(ci) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LC
Log likelihood function          215.15699
Estimation based on N =    256, K =  13
Inf.Cr.AIC  =   -404.3 AIC/N =   -1.579
Variances: Sigma-squared(v)=      .00820
           Sigma-squared(u)=      .00753
           Sigma(v)        =      .09054
           Sigma(u)        =      .08676
Sigma = Sqr[(s^2(u)+s^2(v)]=      .12539
Gamma = sigma(u)^2/sigma^2 =      .47870
Var[u]/{Var[u]+Var[v]}     =      .25020
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0             214.75424
Chi-sq=2*[LogL(SF)-LogL(LS)] =      .806
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    9.19939      21.64273      .43  .6708   -33.21957  51.61835
      LY|     .97398***     .01751    55.63  .0000     .93966   1.00829
     LY2|     .05123***     .01029     4.98  .0000     .03106    .07140
    LPKP|     .49455       1.69257      .29  .7701   -2.82283   3.81193
    LPLP|     .13721*        .08121     1.69  .0911    -.02195    .29637
    LPMP|     .45863       1.11624      .41  .6812   -1.72915   2.64642
    LPEP|    -.10302         .53634     -.19  .8477   -1.15422    .94818
    LPFP|    -.02090         .01794    -1.16  .2441    -.05607    .01427
LOADFCTR|    -.99466***      .17446    -5.70  .0000   -1.33660   -.65273
LOGSTAGE|    -.17940***      .02531    -7.09  .0000    -.22902   -.12979
  POINTS|     .00164***      .00031     5.20  .0000     .00102    .00225
        |Variance parameters for compound error
  Lambda|     .95827***      .16869     5.68  .0000     .62763   1.28890
   Sigma|     .12539***      .00039   321.29  .0000     .12463    .12616
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
Model Simulation Analysis for JLMS efficiency estimator in SF model
-----------------------------------------------------------------------
Simulations are computed by average over sample observations
-----------------------------------------------------------------------
User Function       Function   Standard
(Delta method)        Value      Error      |t|  95% Confidence Interval
-----------------------------------------------------------------------
Avrg. Function        .93354     .00635  147.07       .92110       .94598
LOADFCTR=    .40      .95844     .00346  277.19       .95166       .96522
LOADFCTR=    .43      .95502     .00344  277.54       .94827       .96176
LOADFCTR=    .45      .95123     .00357  266.70       .94424       .95822
LOADFCTR=    .48      .94706     .00392  241.56       .93937       .95474
LOADFCTR=    .50      .94247     .00456  206.48       .93353       .95142
LOADFCTR=    .53      .93746     .00552  169.87       .92664       .94828
(some rows omitted)
LOADFCTR=    .83      .84622     .03145   26.91       .78458       .90786
LOADFCTR=    .85      .83696     .03384   24.73       .77063       .90329
LOADFCTR=    .88      .82763     .03616   22.89       .75676       .89850
LOADFCTR=    .90      .81827     .03839   21.32       .74303       .89352
LOADFCTR=    .93      .80892     .04053   19.96       .72947       .88836
LOADFCTR=    .95      .79958     .04259   18.78       .71611       .88305
LOADFCTR=    .98      .79029     .04455   17.74       .70296       .87761
```

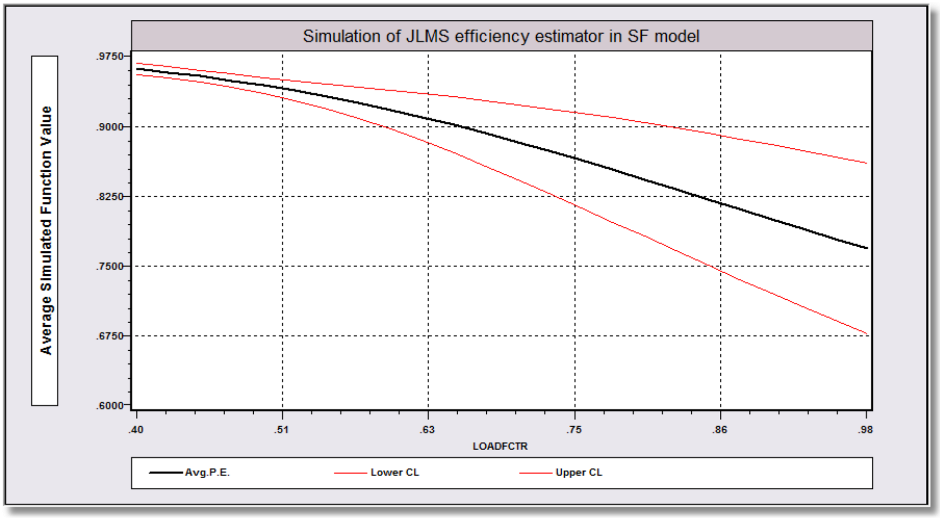


Figure E62.8 Simulated Cost Efficiency Values

We have also analyzed the partial effects.

FRONTIER ; Cost ; Lhs = lcp ; Rhs = x,z $
PARTIALS ; Effects: loadfctr & loadfctr = .4(.025)1 ; Plot(ci) $
PARTIALS ; Effects: z ; Summary $

```
-----------------------------------------------------------------------
Partial Effects  Analysis for JLMS efficiency estimator in SF model
-----------------------------------------------------------------------
Effects on function with respect to LOADFCTR
Results are computed by average over sample observations
Partial effects for continuous LOADFCTR computed by differentiation
Effect is computed as derivative     = df(.)/dx
-----------------------------------------------------------------------
df/dLOADFCTR         Partial    Standard
(Delta method)       Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
APE. Function        -.22444     .06690    3.35      -.35557      -.09331
LOADFCTR=    .40     -.13020     .02575    5.06      -.18067      -.07973
LOADFCTR=    .43     -.14405     .03134    4.60      -.20547      -.08263
LOADFCTR=    .45     -.15900     .03766    4.22      -.23281      -.08519
LOADFCTR=    .48     -.17497     .04464    3.92      -.26246      -.08748
(Some rows omitted)
LOADFCTR=    .85     -.37205     .09615    3.87      -.56051      -.18359
LOADFCTR=    .88     -.37392     .09265    4.04      -.55551      -.19234
LOADFCTR=    .90     -.37452     .08896    4.21      -.54887      -.20017
LOADFCTR=    .93     -.37403     .08524    4.39      -.54109      -.20697
LOADFCTR=    .95     -.37265     .08160    4.57      -.53259      -.21271
LOADFCTR=    .98     -.37054     .07813    4.74      -.52368      -.21739
```

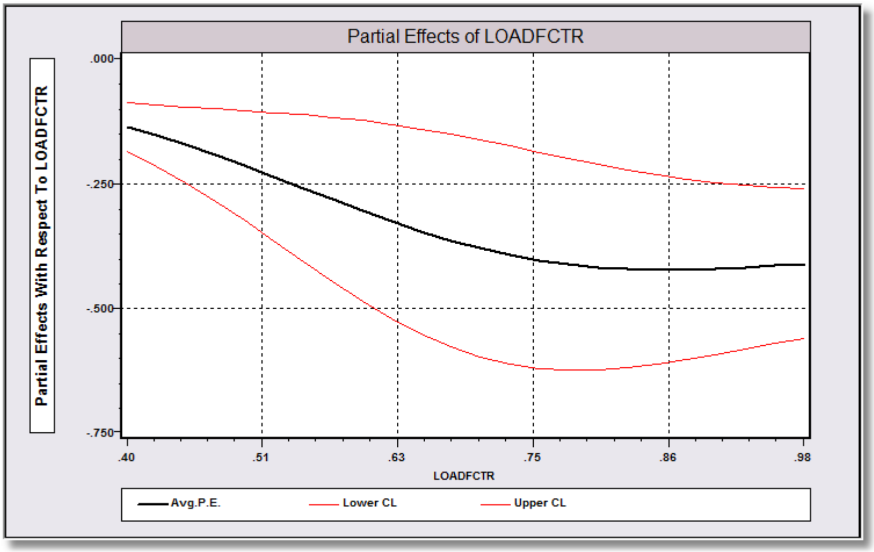


Figure E62.9 Partial Effects of Load Factor

```
-----------------------------------------------------------------------
Partial Effects for JLMS efficiency estimator in SF model
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                     Partial    Standard
(Delta method)       Effect      Error     |t|  95% Confidence Interval
-----------------------------------------------------------------------
      LOADFCTR       -.25723     .07389    3.48      -.40205      -.11240
      LOGSTAGE       -.04620     .01292    3.58      -.07153      -.02088
      POINTS          .00035     .00012    2.95       .00012       .00058
-----------------------------------------------------------------------
```

E62.8.4 Partial Effects of Model Variables on Efficiencies

The preceding has examined the partial effects with respect to **z** in the model

$$y = \boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\delta}'\mathbf{z} + v\text{-}Su.$$

It was noted that partial effects with respect to **x** are not likely to be particularly interesting. Nonetheless, they could be computed.

NOTE: Partial effects of variables in the stochastic frontier efficiency models may be computed with respect to any variable in any model, regardless of where those variables appear in the model. That includes **x** in the original frontier model, **z** in the means of the truncated regression formats, and **z** in the variances of the heteroscedasticity models.

To continue the earlier example, the partial effect of Log*Q* could be computed in the cost function using

```
NAMELIST    ; x = one,lq,lq^2,lpmpp,lpfpp,lpepp,lplpp,lpkpp $
NAMELIST    ; z = loadfctr,logstage,points $
FRONTIER    ; Cost ; Lhs = lcp ; Rhs = x,z  $
PARTIALS    ; Effects : lq ; summary $
```

Note that the specification will correctly account for the fact that the square of Log*Q* appears in the cost function when it computes the partial effects.

E62.8.5 Examining Ranks of Inefficiencies

Researchers often analyze outcome data in which the absolute values of the inefficiencies are not necessarily of interest. Rather, it is the ranking of observations that they wish to analyze. The WHO analysis of health care attainment (see Section E62.4.2) is a prominent example. *LIMDEP* provides several tools for examining ranks of inefficiencies.

First, to rank the raw observations on efficiency or inefficiency, use

```
CREATE      ; rank variable = Rnk(variable) $
```

The Rnk function sorts the data for you and creates the ranking variable. The observation with the highest value gets the rank of one. The lowest gets a rank of *n*. Note, tied observations do not get the same rank. Tied observations are ranked in the order in which they appear in the data. For example, in a sample of 100, if 10 observations are tied for third place, they will receive ranks 3 through 12.

Two **CALC** functions provide descriptive measures for ranks. For two sets of ranks, the Spearman rank correlation coefficient is computed as

$$\rho = 1 - 6\ \Sigma_i d_i^2\ /n(n^2 - 1),$$

$$d_i = variable1_i - variable2_i$$

The function for computing this is

```
CALC        ; List ; Rkc(variable1,variable2) $
```

The rank correlation is a correlation coefficient, so it has a natural range of measurement. (See the application below.) For more than two sets of ranks, a useful statistic is Kendall's coefficient of concordance,

$$W = 12 \sum_{i=1}^{n} (S_i - \bar{S})^2/[nK^2(n^2 - 1)]$$

where $S_i = \Sigma_k rank_{k,i}$.

To compute this measure, use

```
CALC        ; List ; Cnc(ranks1,...,ranksK) $
```

The concordance coefficient is not a correlation coefficient, so its magnitude is ambiguous. It can be used for a large sample test of discordance. Under the null hypothesis that the sets of ranks are independent, the statistic has a large sample chi squared distribution. In particular,

$$K(n\text{-}1)W \rightarrow \chi^2[K(n\text{-}1)].$$

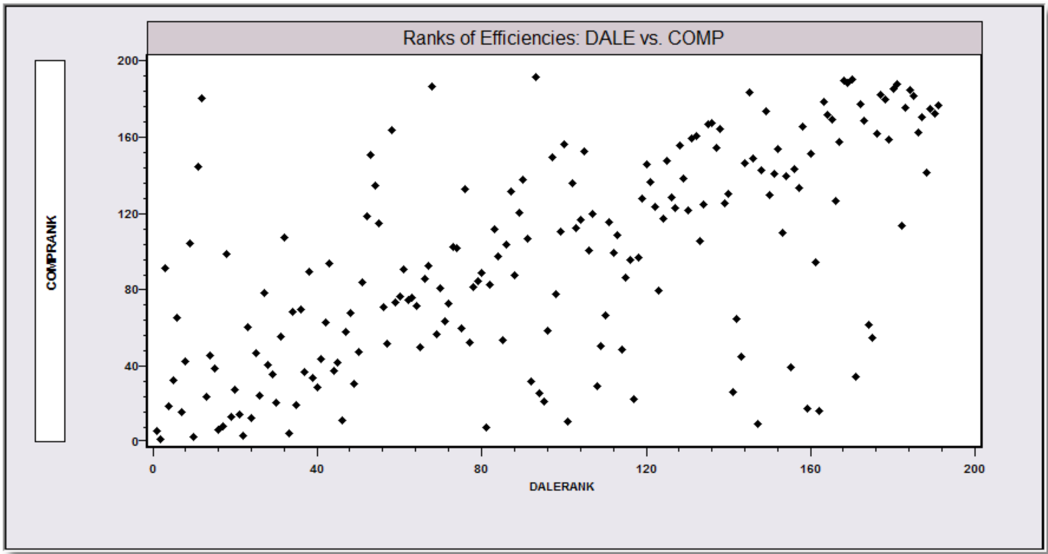
To illustrate these computations, we have analyzed the WHO data described in Section E62.4.2. We have fit identical stochastic frontier models for the two attainment variables, *lcomp*, the log of the composite measure, and *ldale*, the log of disability adjusted life expectancy. We then computed the ranks for the 191 countries and plotted the ranks for the two measures as well as the raw efficiency measures. The simple correlation for the efficiency measures and the rank correlation for the ranks are displayed. The commands are as follows:

```
NAMELIST    ; x = one,logebar,loghbar,loghbar2 $
NAMELIST    ; z = gini,lpopden,lgdpc,geff,voice,oecd,lpubthe,tropics $
FRONTIER    ; Lhs = logdbar ; Rhs = x,z
            ; Eff  = udale ; Techeff = edale $
FRONTIER    ; Lhs = logcbar ; Rhs = x,z
            ; Eff  = ucomp ; Techeff = ecomp $
CREATE      ; dalerank = 192 - Rnk(edale) $
CREATE      ; comprank = 192 - Rnk(ecomp) $
PLOT        ; Lhs = dalerank ; Rhs = comprank
            ; Endpoints = 0,200 ; Limits = 0,200
            ; Title = Ranks of Efficiencies: DALE vs. COMP $
PLOT        ; Lhs = edale ; Rhs = ecomp ; Endpoints = .8,1 ; Grid
            ; Title = Efficiencies: DALE vs. COMP $
CALC        ; List ; Rkc(dalerank,comprank) $
CALC        ; List ; Cor(edale,ecomp) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                LOGDBAR
Log likelihood function         155.83849
Estimation based on N =    191, K =  14
Inf.Cr.AIC  =   -283.7 AIC/N =   -1.485
Variances: Sigma-squared(v)=      .00145
           Sigma-squared(u)=      .03288
           Sigma(v)        =      .03808
           Sigma(u)        =      .18134
Sigma = Sqr[(s^2(u)+s^2(v)]=      .18529
Gamma = sigma(u)^2/sigma^2 =      .95777
Var[u]/{Var[u]+Var[v]}     =      .89180
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0            141.59006
Chi-sq=2*[LogL(SF)-LogL(LS)] =   28.497
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 LOGDBAR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    2.60812***      .18255    14.29  .0000     2.25034   2.96590
 LOGEBAR|     .11227***      .01869     6.01  .0000      .07564    .14891
 LOGHBAR|     .30118***      .05072     5.94  .0000      .20177    .40059
LOGHBAR2|    -.02710***      .00455    -5.96  .0000     -.03601   -.01818
    GINI|    -.30417***      .10600    -2.87  .0041     -.51192   -.09642
 LPOPDEN|     .00213         .00402      .53  .5955     -.00574    .01001
   LGDPC|     .07541***      .02424     3.11  .0019      .02789    .12293
    GEFF|    -.00673         .01551     -.43  .6642     -.03714    .02367
   VOICE|     .02093*        .01113     1.88  .0601     -.00089    .04275
    OECD|     .01608         .03055      .53  .5987     -.04381    .07596
 LPUBTHE|     .00974         .01497      .65  .5150     -.01959    .03908
 TROPICS|    -.03703**       .01714    -2.16  .0307     -.07063   -.00344
        |Variance parameters for compound error
  Lambda|    4.76248***     1.22054     3.90  .0001     2.37026   7.15470
   Sigma|     .18529***      .00086   214.30  .0000      .18360    .18698
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable               LOGCBAR
Log likelihood function        248.18065
Estimation based on N =    191, K =  14
Inf.Cr.AIC  =   -468.4 AIC/N =   -2.452
Variances: Sigma-squared(v)=      .00142
           Sigma-squared(u)=      .00888
           Sigma(v)        =      .03768
           Sigma(u)        =      .09421
Sigma = Sqr[(s^2(u)+s^2(v)]=      .10147
Gamma = sigma(u)^2/sigma^2 =      .86207
Var[u]/{Var[u]+Var[v]}     =      .69429
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           241.57767
Chi-sq=2*[LogL(SF)-LogL(LS)] =   13.206
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 LOGCBAR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    3.21081***      .10704    30.00  .0000     3.00101   3.42060
 LOGEBAR|     .06590***      .01319     4.99  .0000      .04004    .09177
 LOGHBAR|     .18617***      .03763     4.95  .0000      .11240    .25993
LOGHBAR2|    -.01509***      .00328    -4.61  .0000     -.02151   -.00867
    GINI|    -.25334***      .07579    -3.34  .0008     -.40189   -.10478
 LPOPDEN|     .00523*        .00281     1.86  .0628     -.00028    .01073
   LGDPC|     .05747***      .01681     3.42  .0006      .02453    .09040
    GEFF|     .00290         .01068      .27  .7858     -.01803    .02384
   VOICE|     .02082**       .00872     2.39  .0170      .00373    .03791
    OECD|     .01699         .01946      .87  .3827     -.02115    .05513
 LPUBTHE|     .01798**       .00903     1.99  .0466      .00027    .03568
 TROPICS|    -.02365**       .01191    -1.99  .0471     -.04700   -.00031
        |Variance parameters for compound error
  Lambda|    2.50000***      .41784     5.98  .0000     1.68104   3.31896
   Sigma|     .10147***      .00045   224.53  .0000      .10058    .10235
--------+--------------------------------------------------------------------

[CALC] *Result*=       .6353076
[CALC] *Result*=       .6062125
```



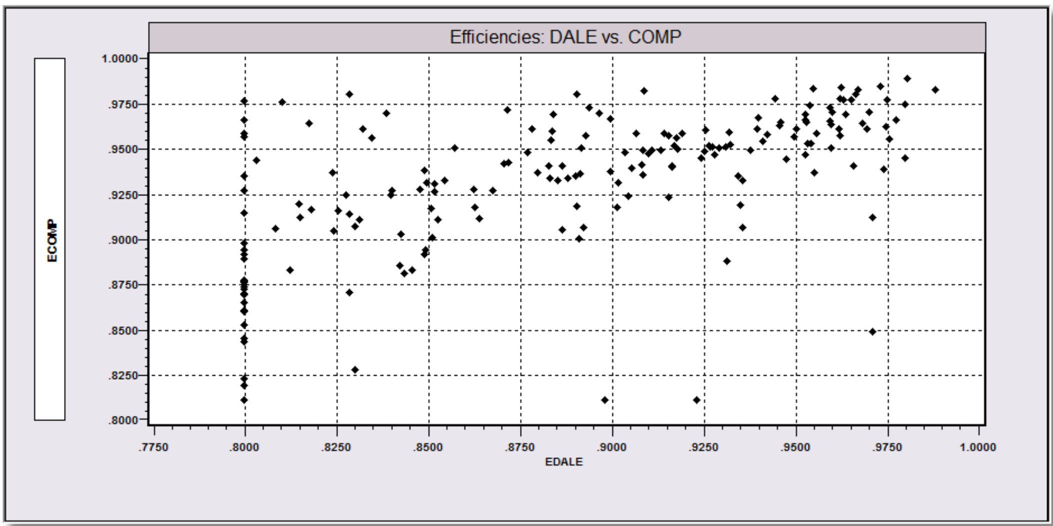


Figure E62.10 Ranks and Estimates of Efficiency

E62.9 Partially Nonparametric Stochastic Frontier Model

The stochastic frontier is fully parametric in both the deterministic part of the frontier and the distribution of the components of ε_i. This section examines a partially nonparametric model of the form

$$y = g(\mathbf{x},\mathbf{z}) + v - Su.$$

The estimator is based on the locally linear regression in Section E9.5. The underlying logic is the result that in the stochastic frontier model, apart from the constant term, OLS consistently estimates the slope parameters of the model and estimates the constant term with a known bias. For the constant, a, the bias is E[u], the unconditional mean, which in the stochastic frontier model is

$$\text{E}[u] = \sigma_u\sqrt{2/\pi}\,.$$

Continuing this approach, then, the least squares residuals estimate ε_i + E[u]. In addition, the least squares residual variance, $\mathbf{e}'\mathbf{e}/n$, consistently estimates Var[ε_i] = $\theta^2 = \sigma_v^2 + [(1 - 2/\pi)\sigma_u^2]$. The implication is that the only parameter remaining to estimate is σ_u^2. In Section E62.6.2, we used the third moment of the OLS residuals and the method of moments to estimate σ_u, then used this estimate to estimate α, the constant term in the frontier function.

The approach proposed here uses this same method with three differences.

1. The residuals used to compute the variance estimator are based on a locally linear, nonparametric estimator of the deterministic function.

2. The remaining parameter to be estimated in this case is λ rather than σ_u. We will base the estimation on the result $\sigma_u^2 = \sigma^2\lambda^2/(1+\lambda^2)$.

3. The approach will be based on a maximum likelihood estimator rather than the method of moments.

Estimation uses the following steps: We begin with estimation of the conventional normal-half normal frontier model with a linear frontier function in order to obtain an initial estimator of λ and of θ^2. The LOWESS estimator developed in Section E9.5 is then employed to estimate g(x,z) for each point in the sample. The residuals from the estimated functions are used with the estimate of θ^2 for estimation of λ. With θ^2 and λ in hand, we can compute the constant term, a set of residuals, and the JLMS estimators of technical or cost efficiency. Technical details appear in Section E62.9.2.

E62.9.1 Application

We have reestimated the airlines cost frontier with the semiparametric estimator. The frontier functions differ noticeably, primarily in the parameter estimates that are statistically insignificant. The kernel estimators suggest, however, that the difference in the estimates of inefficiency are quite modest. The descriptive statistics suggest the same pattern. The final plot shows more graphically how the nonparametric function has changed the estimates. The fact that most of the estimates from the nonparametric estimator lie below the 45 degree line is consistent with the appearance that generally, they are smaller than the parametric values. The last set of results are the ordinary (Pearson) correlation and Kendall's tau.

```
FRONTIER    ; Cost ; Lhs = lc ; Rhs = x,z ; Costeff = eup $
FRONTIER    ; Cost ; Lhs = lc ; Rhs = x,z ; Lowess ; Costeff = eunp$
KERNEL      ; Rhs = eunp,eup
            ; Title = Estimated Inefficiencies from Parametric and Nonparametric
                      Frontiers $
DSTAT       ; Rhs = eup,eunp $
PLOT        ; Lhs = eup ; Rhs = eunp ; Rh2 = eup ; Fill ; Grid ; Vaxis = EUNP
            ; Title = Nonparametric vs. Parametric Estimates $
CALC        ; List; Cor(eup,eunp) ; Ktr(eup,eunp) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LC
Log likelihood function       215.15699
Estimation based on N =    256, K =  13
Variances: Sigma-squared(v)=     .00820
           Sigma-squared(u)=     .00753
           Sigma(v)        =     .09054
           Sigma(u)        =     .08676
Sigma = Sqr[(s^2(u)+s^2(v)]=     .12539
Gamma = sigma(u)^2/sigma^2 =     .47870
Var[u]/{Var[u]+Var[v]}     =     .25020
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          214.75424
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .806
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    9.19939       21.64273       .43  .6708   -33.21957  51.61835
      LY|     .97398***      .01751     55.63  .0000      .93966   1.00829
     LY2|     .05123***      .01029      4.98  .0000      .03106    .07140
    LPKP|     .49455        1.69257       .29  .7701    -2.82283   3.81193
    LPLP|     .13721*        .08121      1.69  .0911     -.02195    .29637
    LPMP|     .45863        1.11624       .41  .6812    -1.72915   2.64642
    LPEP|    -.10302         .53634      -.19  .8477    -1.15422    .94818
    LPFP|    -.02090         .01794     -1.16  .2441     -.05607    .01427
LOADFCTR|    -.99466***      .17446     -5.70  .0000    -1.33660   -.65273
LOGSTAGE|    -.17940***      .02531     -7.09  .0000     -.22902   -.12979
  POINTS|     .00164***      .00031      5.20  .0000      .00102    .00225
        |Variance parameters for compound error
  Lambda|     .95827***      .16869      5.68  .0000      .62763   1.28890
   Sigma|     .12539***      .00039    321.29  .0000      .12463    .12616
--------+--------------------------------------------------------------------
+--------------------------------------------------+
| Locally linear weighted regression estimation |
| Sample size                256                  |
| Model size                  11                  |
| Band width              .500000                 |
| LOESS Sum of Squared Residuals        1.69637 |
| OLS   Sum of Squared Residuals        2.79975 |
| Derivatives Matrix   LOCLBETA                   |
+--------------------------------------------------+
Reestimating lambda using residuals based on LOWESS regression
Normal exit:   3 iterations. Status=0, F=   -337.3385
-----------------------------------------------------------------------------
Partially Nonparametric Stochastic Frontier Fit by LOWESS
Dependent variable                      LC
Estmation based on N =    256, K =   11
Variances: Sigma-squared(u)=     .00438  Sigma(u) =   .06616
           Sigma-squared(v)=     .00504  Sigma(v) =   .07096
Sigma = Sqr[(s^2(u)+s^2(v)]=     .09702  Lambda   =   .93233
Stochastic Cost Frontier Model, e = v+u
-----------------------------------------------------------------------------
Statistical results are for the sample means of the LOWESS estimated betas.
They are not moments of an asymptotic distribution.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    34.8551       23.42958      1.49  .1368    -11.0661   80.7762
      LY|     .98897***      .05040     19.62  .0000      .89018   1.08775
     LY2|     .04598***      .01677      2.74  .0061      .01310    .07885
    LPKP|    2.48149        1.78813      1.39  .1652    -1.02319   5.98616
    LPLP|     .09976         .10851       .92  .3579     -.11292    .31244
    LPMP|    -.85374        1.34656      -.63  .5261    -3.49295   1.78547
    LPEP|    -.71103         .43514     -1.63  .1023    -1.56389    .14183
    LPFP|    -.02183         .03324      -.66  .5114     -.08698    .04332
LOADFCTR|    -.78691         .65061     -1.21  .2265    -2.06208    .48826
LOGSTAGE|    -.20490*        .11308     -1.81  .0700     -.42653    .01672
  POINTS|     .00225         .00205      1.10  .2710     -.00176    .00627
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|      Mean        Std.Dev.       Minimum       Maximum      Cases Missing
--------+---------------------------------------------------------------------
     EUP|    .933537      .025027      .812486      .975689       256        0
    EUNP|    .948487      .019528      .844732      .983878       256        0
--------+---------------------------------------------------------------------

[CALC] *Result*=       .8690148
[CALC] *Result*=       .6339461
Calculator: Computed   2 scalar results
```

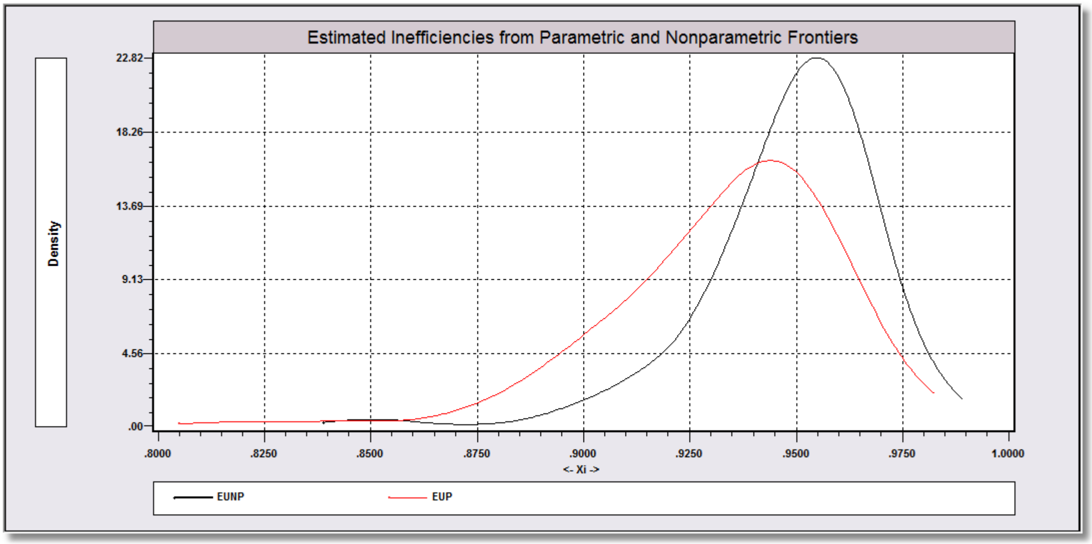


Figure E62.11 Kernel Estimators of Inefficiency Distributions

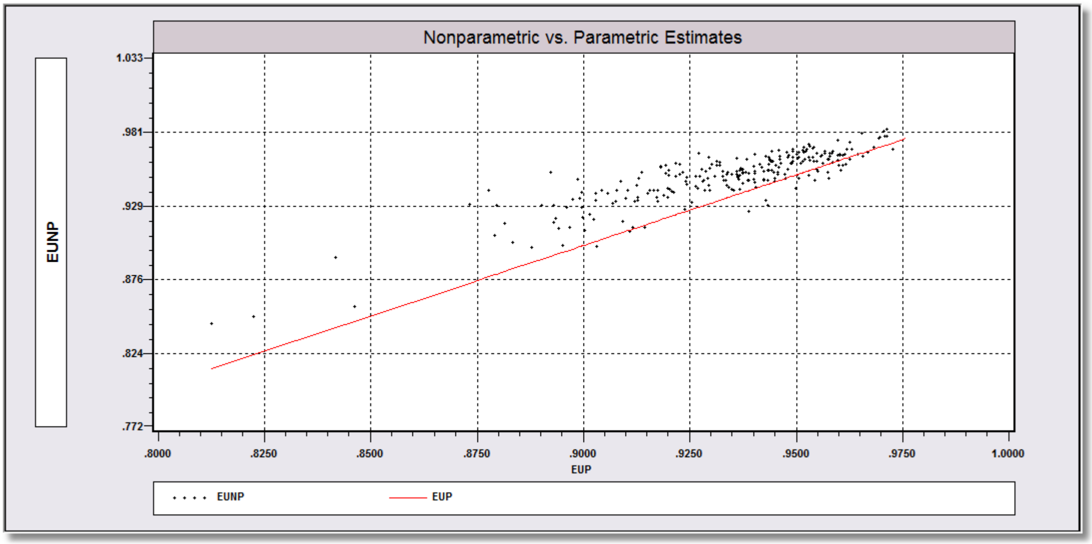


Figure E62.12 Plot of Nonparametric Estimates vs. Parametric Estimates

E62.9.2 Technical Details

The log likelihood function for the normal-half normal model is the sum of

$$\log L_i = \tfrac{1}{2}\log(2/\pi) - \log\sigma - \tfrac{1}{2}(\varepsilon_i/\sigma)^2 + \log\Phi[-S\varepsilon_i\lambda/\sigma].$$

The value of $\theta^2 = \sigma_v^2 + [(1 - 2/\pi)\sigma_u^2]$is estimated using the squared LOWESS residuals; it is the sample variance = q^2. The LOWESS residuals, themselves, are estimates of ε_i + E[u_i]. With q^2 and the residuals in hand, the log likelihood is a function only of λ. During the iteration, we compute

$$a = \lambda/(1+\lambda^2)^{1/2},$$

$$s^2 = q^2 / (1 - (2/\pi)a^2), \text{ then } s$$

$$m = as\sqrt{2/\pi}$$

$$e_i = \text{residual}_i - m.$$

These residuals and s are used to compute logL_i and the derivative with respect to λ. This estimation step provides the estimator of λ that we need to compute the efficiencies. After estimation of λ, computation of the JLMS estimates of inefficiency is done the same as in the parametric form of the model, using the LOWESS residuals.

E62.10 The Normal-Gamma Model

The normal-gamma model is the remaining distributional form of the stochastic frontier model. Under this specification,

$$u_i \sim \frac{\theta^P \exp(-\theta u_i) u_i^{P-1}}{\Gamma(P)}, \; u_i \geq 0, P > 0, \theta > 0.$$

This model is more flexible than the half normal or exponential model in that with two parameters, it allows the both the shape and location to vary independently. (The truncation model does likewise, but it is considerably more difficult to estimate.) To specify the gamma model, use

; Model = Gamma (or ; Model = G)

The normal-gamma model is estimated by the method of simulated maximum likelihood. (See Greene (2000b) and the details in Section E62.10.2.) The counterpart to the JLMS estimator of the inefficiency, $E[u|\varepsilon]$ must also be estimated by simulation.

E62.10.1 Application of the Normal-Gamma Model

We illustrate the gamma model by fitting a cost frontier model with normal-gamma inefficiency. For comparison, we have also fit the exponential model, which results when *P* is constrained to equal one. (The exponential model is fit directly by its own log likelihood, not by constraining *P* to equal one in the gamma model.) We have also computed the inefficiencies for the two models, and plotted kernel density estimators to compare them. The commands are

```
FRONTIER   ; Lhs   = lc ; Rhs = x ; Cost ; Model = Gamma ; Costeff = eucg
           ; Pts   = 50 ; Halton $
FRONTIER   ; Lhs   = lc ; Rhs = x ; Cost ; Model = Exponential ; Costeff = euce $
KERNEL     ; Rhs   = eucg,euce
           ; Title = Kernel Density Estimates for E[u|e,exponential and gamma] $
```

We note by the Wald and likelihood ratio tests, we cannot reject the hypothesis of the exponential model (*P* is close to one). The similarity of the kernel density estimators is consistent with this finding.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LC
Log likelihood function         159.94270
Estimation based on N =    256, K =   11
Inf.Cr.AIC  =   -297.9 AIC/N =   -1.164
Model estimated: Aug 22, 2011, 22:09:16
Normal-Gamma frontier model
Variances: Sigma-squared(v)=     .01169
           Sigma-squared(u)=     .00547
           Sigma(v)        =     .10814
           Sigma(u)        =     .07399
Stochastic Cost Frontier Model, e = v+u
Half Normal:u(i)=|U(i)|; frontier model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0            157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =    4.055
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    22.9007       27.13658      .84  .3987    -30.2860   76.0874
      LY|     .96086***     .02028     47.38  .0000      .92112   1.00061
     LY2|     .09283***     .01327      7.00  .0000      .06682    .11883
    LPKP|    1.67283       2.12387       .79  .4309    -2.48987   5.83553
    LPLP|    -.01112        .06724      -.17  .8687     -.14290    .12066
    LPMP|    -.07676       1.37564      -.06  .9555    -2.77297   2.61944
    LPEP|    -.63376        .68533      -.92  .3551    -1.97698    .70946
    LPFP|    -.06405***     .02311     -2.77  .0056     -.10934   -.01876
        |Variance parameters for compound error
   Theta|    12.4180**     5.05037      2.46  .0139      2.5194   22.3165
       P|     .84426        .69128      1.22  .2220     -.51062   2.19913
  Sigmav|     .10814***     .01148      9.42  .0000      .08563    .13064
--------+--------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
Log likelihood function       159.89917
Exponential frontier model
Variances: Sigma-squared(v)=     .01147
           Sigma-squared(u)=     .00568
           Sigma(v)        =     .10709
           Sigma(u)        =     .07539
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =    3.968
Kodde-Palm C*: 95%: 2.706,   99%:  5.412
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LC|  Coefficient      Error       z    |z|>Z*        Interval
--------+-------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    22.6569      25.48354      .89  .3740     -27.2899    72.6038
      LY|     .96069***     .01892    50.77  .0000       .92360     .99777
     LY2|     .09281***     .01249     7.43  .0000       .06832     .11729
    LPKP|    1.65439       1.99409      .83  .4067     -2.25395    5.56272
    LPLP|    -.00962        .09785     -.10  .9217      -.20140     .18216
    LPMP|    -.06595       1.31569     -.05  .9600     -2.64465    2.51275
    LPEP|    -.62841        .63243     -.99  .3204     -1.86795     .61114
    LPFP|    -.06397***     .02033    -3.15  .0017      -.10381    -.02412
        |Variance parameters for compound error
   Theta|    13.2651***    2.90719     4.56  .0000      7.5671     18.9630
  Sigmav|     .10709***     .00980    10.93  .0000       .08788     .12629
--------+-------------------------------------------------------------------
```

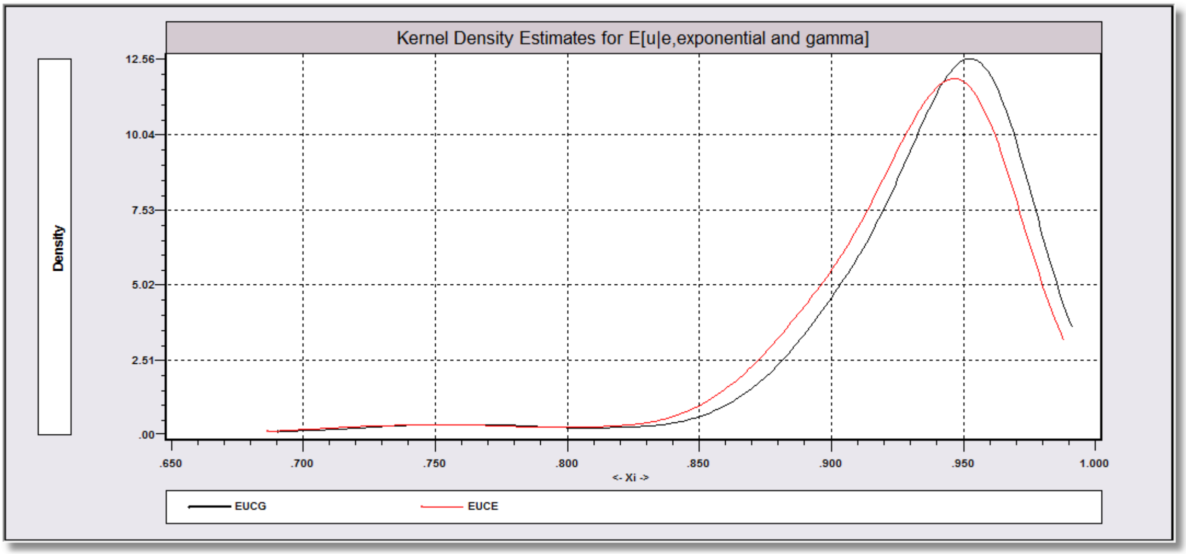


Figure E62.13 Kernel Density Estimates for Gamma and Exponential Inefficiencies

E62.10.2 Technical Details on Normal-Gamma Model

The log likelihood for this model is equal to the log likelihood for the normal-exponential model plus a term that is produced by the difference between the exponential and the gamma distributions;

$$\text{Log } L = \text{Log } L(\textit{exponential}) + n[(P\text{-}1)\log\theta - \log\Gamma(P)] + \Sigma_i \log h(P\text{-}1,\varepsilon_i)$$

where

$$h(r,\varepsilon_i) = \frac{\int_0^\infty z^r (1/\sigma_v)\phi((z-\eta_i)/\sigma_v)dz}{\int_0^\infty (1/\sigma_v)\phi((z-\eta_i)/\sigma_v)dz}, \ \eta_i = -\varepsilon_i - \theta\sigma_v^2.$$

The normal-exponential model results if $P = 1$. Computation of the function $h(r,\varepsilon_i)$ is the obstacle to estimation. Beckers and Hammond (1987) derived a closed form expression, but the result has never been operationalized – it is complex in the extreme. Greene (1990) attempted estimation by using a crude approximation with Simpson's rule, but failed to obtain reasonable results. (See Ritter and Simar (1997).)

A satisfactory solution is produced by the technique of maximum simulated likelihood. The integral and its derivatives can be estimated consistently by Monte Carlo simulation. The crucial result is that $h(r,\varepsilon_i)$ is the expectation of a random variable;

$$h(r,\varepsilon_i) = E[z^r \mid z \geq 0]$$

where

$$z \sim N[\eta_i, \sigma_v^2]$$

$$\eta_i = -\varepsilon_i - \theta\sigma_v^2$$

Therefore, $h(r,\varepsilon_i)$ is the expected value of z^r where z has a truncated at zero normal distribution. Thus, we estimate $h(r,\varepsilon_i)$ by using the mean of a sample of draws from this distribution. For given values of ε_i and η_i (i.e., y_i, $\mathbf{x}_i$, $\boldsymbol{\beta}$, σ_v, θ, r), $h(r,\varepsilon_i)$ is consistently estimated by

$$\hat{h}_i = \frac{1}{Q}\sum_{q-1}^{Q} z_{iq}^r$$

where z_{iq} is a random draw from the truncated normal distribution with mean parameter η_i and variance parameter σ_v. This produces the simulated log likelihood function

$$\text{Log } L_S = \text{Log } L(\textit{exponential}) + n[(P\text{-}1)\log\theta - \log\Gamma(P)] + \Sigma_i \log \hat{h}(P\text{-}1,\varepsilon_i)$$

which for a given set of draws is a smooth and continuous function of the parameters.

Random draws from the truncated distribution are obtained using Geweke's method as follows: Let

L = truncation point = 0 for this application

μ = the mean of untruncated distribution = $-\varepsilon_i - \theta\sigma_v^2$

σ = the standard deviation of the untruncated distribution = σ_v

$P_L = \Phi[(L - \mu) / \sigma]$

F = one draw from U[0,1]

$z = \mu + \sigma\Phi^{-1}[P_L + F\times(1 - P_L)]$

Then, z = the draw from the truncated distribution.

Collecting all terms, then, this produces the simulated log likelihood function:

$$\begin{aligned}\text{Log } L \;&=\; n\{\log\theta + \tfrac{1}{2}\sigma_v^2\theta^2\} + \Sigma_i\{\theta d\varepsilon_i + \log\Phi[-(d\varepsilon_i/\sigma_v + \theta\sigma_v)]\}\\ &+\; n[(P\text{-}1)\log\theta - \log\Gamma(P)]\\ &+\; \Sigma_i \log\left\{\frac{1}{Q}\sum_{q=1}^{Q}\left[\mu_i + \sigma_v\Phi^{-1}\left(F_{iq} + (1-F_{iq})\Phi\left(\frac{-\mu_i}{\sigma_v}\right)\right)\right]^{P-1}\right\}\end{aligned}$$

$$\varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i$$

$$\mu_i = -\varepsilon_i - \theta\sigma_v^2$$

and F_{iq} is a fixed set of Q draws from U[0,1] specific to the individual. Derivatives of $h(r,\varepsilon_i)$ and log $h(r,\varepsilon_i)$ are also estimated by simulation. The JLMS efficiency measure has the simple form

$$E[u|\varepsilon] = h(P,\varepsilon_i) / h(P\text{-}1,\varepsilon_i).$$

The final consideration is the method of obtaining the draws. The default method is to use the random number generators. Since this is a very computation intensive model, it is usually more efficient to use Halton draws – you can use many fewer Halton draws than random draws to obtain the same quality results. Halton draws are discussed in Section R24.7. To use Halton draws with this estimator, add

; Halton

to the command. The number of points for either method is specified with

; Pts = the desired number of draws

We have used this feature in the example in the previous section.

E62.11 Sample Selection in a Stochastic Frontier Model

This model is a counterpart to familiar models of sample selection. See Greene (2010) for details on the methodology. Additional results appear in Terza (2010). The model is a familiar sample selection form

$$\begin{aligned}
d^* &= \boldsymbol{\alpha}'\mathbf{z} + w, \; d = 1(d^* > 0) \\
y &= \boldsymbol{\beta}'\mathbf{x} + v - u \\
u &= |U| \text{ with } U \sim N[0,\sigma_u^2] \\
(v,w) &\sim \text{bivariate normal with } [(0,0),(\sigma_v^2, \rho\sigma_v, 1)] \\
(y,\mathbf{x}) & \quad \text{only observed when } d = 1.
\end{aligned}$$

Thus, the selection operates through the heterogeneity component of the production model, not the inefficiency. (Thus, observation is not viewed as a function of the level of inefficiency.)

The model is fit by maximum simulated likelihood. To request it, use *LIMDEP*'s usual format for sample selection models,

```
PROBIT     ; Lhs = d ; Rhs = variables in w ; Hold $
FRONTIER   ; Lhs = y ; Rhs = variables in x; Selection $
```

The model must be the base case, half normal, with no panel data application, no truncation, or heteroscedasticity, etc. You may control the simulations with **; Halton** and **; Pts** for the simulation. Efficiency and inefficiency estimates are saved as with other models with **; Eff** and **; Techeff**. However, observations in the nonselected part of the sample are given missing values (-999) for any of these computations. The **PARTIALS** and **SIMULATE** commands do not inherit the selection model – these commands are not available after fitting this model.

E62.11.1 Application

The following creates a data set that conforms exactly to the assumptions of the model.

```
CALC       ; Ran(123457) $
SAMPLE     ; 1-2000 $
CREATE     ; z1 = Rnn(0,1) ; z2 = Rnn(0,1) $
CREATE     ; v1 = Rnn(0,1) ; v2 = Rnn(0,1) $
CREATE     ; e1 = v1 ; e2 = .7071 * (v1+v2) $
CREATE     ; ds = z1 + z2 + e1 ; d = ds > 0 $
CREATE     ; u = Abs(Rnn(0,1)) ; x1 = Rnn(0,1) ; x2 = Rnn(0,1) $
CREATE     ; y = x1 + x2 + e2 - u $
PROBIT     ; Lhs = d ; Rhs = one,z1,z2 ; Hold $
FRONTIER   ; Lhs = y ; Rhs = one,x1,x2 ; Selection $
```

```
----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                     D
Log likelihood function      -825.27526
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       D|  Coefficient      Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
        |Index function for probability
Constant|     .03616        .03525     1.03  .3051     -.03294    .10525
      Z1|     .96314***     .04604    20.92  .0000      .87291   1.05338
      Z2|    1.01534***     .04702    21.59  .0000      .92318   1.10750
--------+-------------------------------------------------------------------
Warning   141: Iterations:current or start estimate of sigma nonpositive
Normal exit:  14 iterations. Status=0, F=    1916.202
----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     Y
Log likelihood function     -1916.20216
Estimation based on N =   2000, K =   6
Inf.Cr.AIC  =   3844.4 AIC/N =    1.922
Variances: Sigma-squared(v)=    1.00545
           Sigma-squared(u)=    1.07396
           Sigma(u)        =    1.03632
           Sigma(v)        =    1.00272
           Sigma           =    1.44202
           Lambda          =    1.03351
Sample Selection/Frontier Model
Murphy/Topel Corrected VC Matrix
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0        -1662.32532
Chi-sq=2*[LogL(SF)-LogL(LS)] = -507.754
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
----------------------------------------------------------------------------

--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient      Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    -.04492        .10971     -.41  .6822     -.25994    .17011
      X1|    1.00102***     .03357    29.82  .0000      .93522   1.06682
      X2|     .95627***     .03195    29.93  .0000      .89364   1.01890
Sigma(u)|    1.03632***     .13217     7.84  .0000      .77728   1.29537
Sigma(v)|    1.00272***     .05471    18.33  .0000      .89549   1.10995
Rho(w,v)|     .77553***     .06187    12.54  .0000      .65427    .89679
--------+-------------------------------------------------------------------
```

E62.11.2 Log Likelihood and Estimation Method

Write the model structure as

$$d^* = \boldsymbol{\alpha}'\mathbf{z} + w,\ w \sim N[0,1],\ d = 1(d^* > 0)$$

$$y = \boldsymbol{\beta}'\mathbf{x} + \sigma_v v - \sigma_u u$$

$$u = |U| \text{ with } U \sim N[0,1]$$

$$(v,w) \sim \text{bivariate normal with } [(0,0),(1,\rho,1)]$$

$$(y,\mathbf{x}) \quad \text{only observed when } d = 1.$$

(Note for convenience later, we have moved the scale parameters into the structural model.) To set up the estimator, we now write w in its conditional on v form,

$$w|v = \rho v + h \text{ where } h \sim N[0, (1 - \rho^2)] \text{ and } h \text{ is independent of } v.$$

Therefore, $d^*|v = \boldsymbol{\alpha}'\mathbf{z} + \rho v + h,\ d = 1(d^* > 0|v)$

Then, $\text{Prob}[d = 1 \text{ or } 0 \mid \mathbf{z},v] = \Phi\left[(2d-1)\left(\dfrac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right)\right]$

For the selected observations, $d = 1$, conditioned on v, the joint density for y and d is the product of the marginals since conditioned on v, y and d are independent;

$$f(y, d = 1|\mathbf{x},\mathbf{z},v) = f(y|\mathbf{x},v)\ \text{Prob}(d = 1|\mathbf{z},v).$$

We have the second part above. For the first part,

$$y|\mathbf{x},v = (\boldsymbol{\beta}'\mathbf{x} + \sigma_v v) - \sigma_u u$$

where u is the truncation at zero of a standard normal variable, so $f(u) = 2\phi(u)$, $u \geq 0$. The Jacobian of the transformation from u to y is $1/\sigma_u$, so by the change of variable, the conditional density is

$$f(y \mid \mathbf{x},v) = \frac{2}{\sigma_u}\phi\left(\frac{(\boldsymbol{\beta}'\mathbf{x}+\sigma_v v)-y}{\sigma_u}\right), (\boldsymbol{\beta}'\mathbf{x}+\sigma_v v)-y \geq 0.$$

Therefore, the joint conditional density is

$$f(y, d=1 \mid \mathbf{x},\mathbf{z},v) = \frac{2}{\sigma_u}\phi\left(\frac{(\boldsymbol{\beta}'\mathbf{x}+\sigma_v v)-y}{\sigma_u}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right).$$

To obtain the unconditional density, it is necessary to integrate v out of the conditional density. Thus,

$$f(y,d=1\,|\,\mathbf{x},\mathbf{z}) = \int_v \frac{2}{\sigma_u}\phi\left(\frac{\sigma_v v-(y-\boldsymbol{\beta}'\mathbf{x}))}{\sigma_u}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right)f(v)dv.$$

The relevant term in the log likelihood is log $f(y,d=1|\mathbf{x},\mathbf{z})$. For the nonselected observations, the contribution to the log likelihood is the log of the unconditional probability of nonselection, which is

$$\text{Prob}(d = 0|\mathbf{z}) = \int_v \Phi\left[-\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right)\right]f(v)dv.$$

The integrals do not exist in closed form, so these terms cannot be evaluated as is. Before proceeding, we note the additional complication, $\boldsymbol{\beta}'\mathbf{x} + \sigma_v v - y = \sigma_u u > 0$, so the density $f(v)$ is not the standard normal that intuition might suggest; it is a truncated normal.

The integrals can be computed by simulation. By construction,

$$\int_v \frac{2}{\sigma_u}\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}+\sigma_v v-y)}{\sigma_u}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right)f(v)dv = E_v\left[\frac{2}{\sigma_u}\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}+\sigma_v v-y)}{\sigma_u}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v}{\sqrt{1-\rho^2}}\right)\right]$$

so by sampling from the distribution of v, we can compute the function of v and average to obtain the integrals. In order to sample the draws on v, we note the implied truncation,

$$v \geq (y - \boldsymbol{\beta}'\mathbf{x})/\sigma_v \text{ or } v \geq \varepsilon/\sigma_v.$$

Draws from the truncated normal can be obtained using result (E-1) in Greene (2011). Let A equal a draw from the uniform (0,1) population. The desired draw from the truncated normal distribution will be

$$v_r = \Phi^{-1}\left[\Phi(\varepsilon/\sigma_v) + A_r\Phi(-\varepsilon/\sigma_v)\right].$$

Collecting all terms, then, the simulated log likelihood will be

$$\log L_S = \sum_i \log\frac{1}{R}\sum_{r=1}^{R}\left\{d_i\left[\frac{2}{\sigma_u}\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}+\sigma_v v_{ir}-y)}{\sigma_u}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\mathbf{z}+\rho v_{ir}}{\sqrt{1-\rho^2}}\right)\right]+(1-d_i)\left[\Phi\left(\frac{-\boldsymbol{\alpha}'\mathbf{z}-\rho v_{ir}}{\sqrt{1-\rho^2}}\right)\right]\right\}$$

where the draws on v_{ir} are as shown above. Derivatives of this simulated log likelihood are obtained numerically using finite differences.

E63: Heteroscedasticity and Truncation in Stochastic Frontier Models

E63.1 Introduction

This chapter develops several extensions of the stochastic frontier model presented in Chapter E62. The four models considered here are as follows:

- Heteroscedasticity in v and/or u
- Truncated normal with nonzero, heterogeneous mean in the underlying U
- Heterogeneity in the parameter of the exponential or gamma distribution
- Amsler et al.'s 'scaling model'

E63.2 Heteroscedasticity and Heterogeneity

In the development of the frontier model, an important question concerns how to introduce observed heterogeneity into the specification. Suppose the vector of variables $\mathbf{z}_i$ contains the information. For example, in the airline data, we have data on load factor, stage length and number of points in the route map, that may also impact production, cost and efficiency. In the model proposed thus far, the only point at which one might introduce $\mathbf{z}_i$ appears to be in the goal function itself, which would become

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \boldsymbol{\alpha}'\mathbf{z}_i + v_i - u_i.$$

This is a common approach. (See, e.g., Greene (2004a,b).) In this chapter, we present two other methods of introducing observed heterogeneity in the frontier model, in the variance parameters and in the mean of the underlying inefficiency.

E63.2.1 Heterogeneity in the Scale Parameters

A natural departure point is to allow observable variation in σ_v^2 and/or σ_u^2. For the first of these, the term heteroscedasticity is appropriate. (The papers by Hadri et al. (1999, 2003a,b) develop heteroscedasticity models for frontier specifications.) For the second of these, a result which seems routinely to be overlooked in the literature is that allowing σ_u^2 to vary over observations, call it $\sigma_{u,i}^2$, induces more than just heteroscedasticity. Unavoidably in all model specifications, when this parameter varies over individuals, then both the variance and the mean of u_i do also. For the half normal model, regardless of how $\sigma_{u,i}$ varies,

$$E[u_i] = \sigma_{u,i}\phi(0)/\Phi(0) = 0.79788\sigma_{u,i}.$$

A like result emerges in the truncated normal model. In the exponential model, the mean of u_i equals its standard deviation, while in the gamma model, it is a multiple, $P^{1/2}$, of it. Thus, in all cases, as regards u_i, the term heteroscedasticity, while not inappropriate, is nonetheless ambiguous. These models cannot be heteroscedastic without also having a heterogeneous mean. In what follows, therefore, we continue to use the familiar terminology, but we emphasize the nature of the model as well.

The models of scale heterogeneity may extend either variance parameter with the specification of the variance functions

$\mathrm{Var}[U|\mathbf{z}_i] = \sigma_{ui}^2 = \sigma_u^2 \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$ (heteroscedastic)

$\mathrm{Var}[v|\mathbf{z}_i] = \sigma_v^2 = \sigma_v^2 \exp(\boldsymbol{\delta}'\mathbf{w}_i)$ (heteroscedastic)

$\mathrm{Var}[u|\mathbf{z}_i] = \sigma_u^2 \exp(\boldsymbol{\gamma}'\mathbf{z})$ and $\mathrm{Var}[v|\mathbf{z}_i] = \sigma_v^2 \exp(\boldsymbol{\delta}'\mathbf{w}_i)$ (doubly heteroscedastic)

There is no requirement that the same variables enter the two functions, and either or both may be heterogeneous. The model specification is

; Heteroscedasticity or ; Het

and either or both of

; Hfv = variables in the variance of v
; Hfu = variables in the variance of u

If either variance is not given, it is assumed to be constant. The variance function is the exponential format used throughout *LIMDEP* If either variance is unspecified, the implied model is $\sigma_{ji}^2 = \exp(\delta$ or $\gamma)$ which is the same as

; Hfv = one or ; Hfu = one

If both are unspecified, then the implied model

; Het ; Hfv = one ; Hfu = one

is the default, normal-half normal stochastic frontier model. It provides identical estimates. (Try it.) A constant (*one*) is automatically inserted into both lists if you do not include it. This form may be used with the normal-half normal and normal-truncated normal models.

E63.2.2 Exponential and Gamma Models with Heterogeneity

The one sided component of the normal-exponential and normal-gamma models is parameterized with a scale parameter, θ, which is thus far taken to be a constant. In these models,

$$E[u_i] = P/\theta = P\times\sigma_u$$

where $P = 1$ in the exponential model. The exponential heteroscedasticity model for u_i is extended to these two models by using

$$\theta_i = \theta \exp(-\boldsymbol{\delta}'\mathbf{z}_i).$$

With this parameterization, the estimates from this model will be comparable to those for the half normal and truncated normal models. (See the examples below.) To request this form, use

; Het ; Hfu = the list of variables.

The list should not contain a constant term, *one*. This may be used in all implementations of the exponential gamma model. Note, however, that in the panel data settings, the parameter is assumed to be time invariant. The values for $\mathbf{z}_i$ are taken from the data record for the last period for firm *i*. We will return to this subject below. The symmetric component, *v*, may also be heteroscedastic, as in the other models, with

; Hfv = list of variables.

E63.2.3 Efficiency Estimation with Heteroscedasticity

This extension does not change the computation of measures of efficiency or inefficiency. The central results are the JLMS estimators,

$$\hat{E}[u\,|\,\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ \varepsilon = v - u,\ w = S\varepsilon\lambda/\sigma$$

for the half normal models and

$$\hat{E}[u\,|\,\varepsilon] = \sigma_v\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ w = (S\varepsilon/\sigma_v + \theta\sigma_v)$$

for the exponential models. These functions are evaluated for each observation at

$$\lambda_i = \sigma_{u,i} / \sigma_{v,i}$$

and

$$\sigma_i^2 = \sigma_{u,i}^2 + \sigma_{v,i}^2$$

for the half normal model and $\sigma_{v,i}$ and θ_i likewise in the exponential and gamma models.

E63.2.4 Application

The estimates below show a production frontier based on the six inputs. The second set of results presents the heteroscedastic model, with the variance of *v* a function of the log of the average stage length and the variance of *u* depending on the load factor and the log of the number of points served. We examine the efficiency results, then compute the average partial effects of the environmental variables on technical efficiency.

```
FRONTIER    ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = eu $
FRONTIER    ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = euhet
            ; Het ; Hfv = lstage ; Hfu = loadfctr,points $
PARTIALS    ; Effects: lstage / loadfctr / points ; Summary $
KERNEL      ; Rhs = eu,euhet
            ; Title = Kernel Estimators for Technical Efficiency $
PLOT        ; Lhs = eu ; Rhs = euhet ; Rh2 = eu ; Fill ; Grid
            ; Title = Estimates of Technical Efficiency
            ; Vaxis = exp(-E[u|e]) for Heteroscedastic Model $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                       LQ
Log likelihood function       108.43918
Estimation based on N =    256, K =   9
Variances: Sigma-squared(v)=      .01902
           Sigma-squared(u)=      .01692
           Sigma(v)        =      .13791
           Sigma(u)        =      .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=      .18957
Gamma = sigma(u)^2/sigma^2 =      .47074
Var[u]/{Var[u]+Var[v]}     =      .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =      .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***      .72136    -4.14  .0000    -4.40206  -1.57439
      LL|    -.42909***      .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***      .09498     4.69  .0000      .25917    .63149
      LF|     .37257***      .07038     5.29  .0000      .23463    .51052
      LE|    2.09473***      .68790     3.05  .0023      .74647   3.44299
      LM|     .69910***      .07580     9.22  .0000      .55054    .84766
      LK|   -2.09806***      .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***      .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***      .00064   297.81  .0000      .18832    .19082
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LQ
Log likelihood function        149.30854
Estimation based on N =    256, K =  12
Inf.Cr.AIC  =   -274.6 AIC/N =   -1.073
Variances: Sigma-squared(v)=      .01292
           Sigma-squared(u)=      .03575
           Sigma(v)        =      .11367
           Sigma(u)        =      .18907
Sigma = Sqr[(s^2(u)+s^2(v)]=      .22061
Gamma = sigma(u)^2/sigma^2 =      .73450
Var[u]/{Var[u]+Var[v]}     =      .50132
Variances averaged over observations
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  3
LogL when sigma(u)=0           108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =   82.468
Kodde-Palm C*: 95%: 8.761,  99%: 12.483
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.29243***      .72664    -4.53  .0000    -4.71662  -1.86824
      LL|    -.47507***      .08890    -5.34  .0000     -.64932   -.30083
      LP|     .50435***      .10452     4.83  .0000      .29950    .70920
      LF|     .53204***      .07550     7.05  .0000      .38406    .68003
      LE|    2.36654***      .69245     3.42  .0006     1.00936   3.72372
      LM|     .53413***      .08670     6.16  .0000      .36419    .70406
      LK|   -2.43136***      .77258    -3.15  .0016    -3.94558   -.91713
        |Parameters in variance of v (symmetric)
Constant|   -3.97891***      .86601    -4.59  .0000    -5.67626  -2.28155
  LSTAGE|    -.06406         .13359     -.48  .6315     -.32590    .19777
        |Parameters in variance of u (one sided)
Constant|    9.96191**      4.51238     2.21  .0273     1.11781  18.80600
LOADFCTR|   -25.9711***     9.37571    -2.77  .0056    -44.3471   -7.5950
  POINTS|    -.00353         .01288     -.27  .7840     -.02877    .02171
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The figure below displays the kernel density estimators for the two sets of estimated inefficiencies. The upper one is for the heteroscedastic model. The figure shows clearly the influence of the heterogeneity. The means of the two distributions are virtually the same, but the variance in the heteroscedastic model is considerably higher.

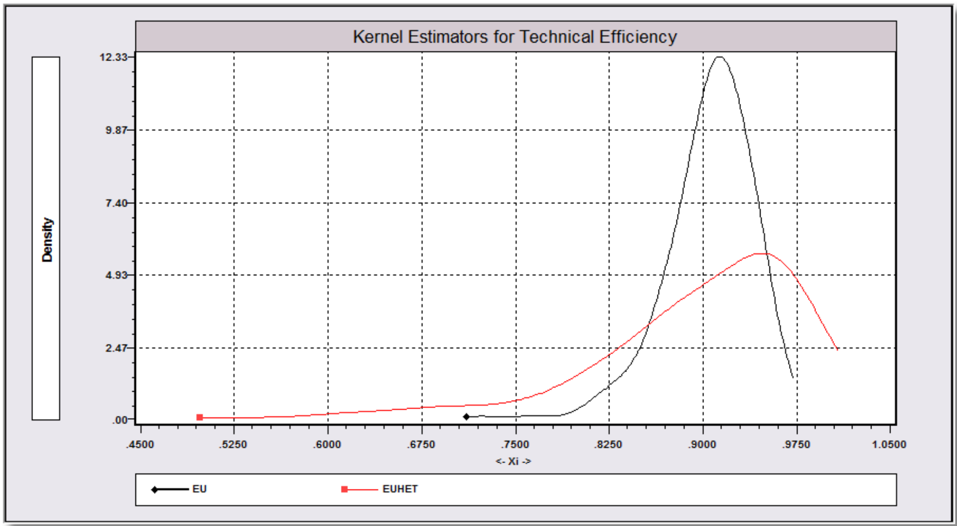


Figure E63.1 Kernel Estimators for Density of E[u|ε] with and without Heteroscedasticity

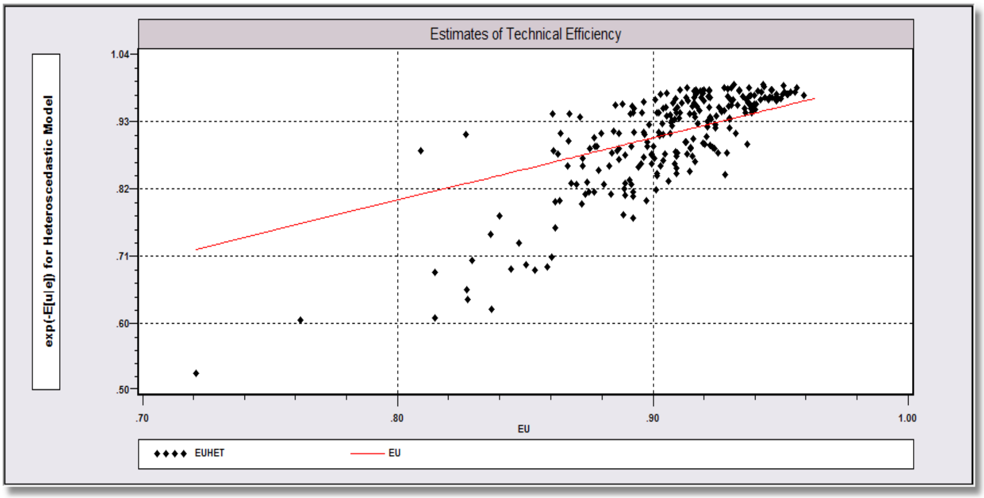


Figure E63.2 Plot of Estimated Inefficiencies, Heteroscedastic vs. Homoscedastic

```
----------------------------------------------------------------------
Partial Effects for JLMS Estimator in Normal/het SF Model
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
----------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect       Error       |t|  95% Confidence Interval
----------------------------------------------------------------------
      LSTAGE      -.00034      .00071      .48      -.00174      .00105
      LOADFCTR     .62934      .17576     3.58       .28485      .97382
      POINTS       .00009      .00031      .28      -.00052      .00069
----------------------------------------------------------------------
```

E63.2.5 Technical Details

For the models with heteroscedasticity, we revert to the original structural form of the model to form the log likelihoods. For the normal-half normal model, for example, we use

$$\log L_i = -\log(2/\pi) - \log\sigma_i - ½(\varepsilon_i/\sigma_i)^2 + \log\Phi[-S\varepsilon_i\lambda_i/\sigma_i]$$

where

$$\sigma_i = \sqrt{\sigma_{ui}^2 + \sigma_{ui}^2}$$

$$\lambda_i = \sigma_{ui} / \sigma_{vi}$$

$$\sigma_{ui}^2 = \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$$

$$\sigma_{vi}^2 = \exp(\boldsymbol{\delta}'\mathbf{w}_i),$$

where S = +1 for a production frontier and -1 for a cost frontier. Likewise, for the truncation model,

$$\log L_i = -½\log 2\pi - \log\sigma_i - ½[(S\varepsilon_i + \mu)/\sigma_i]^2$$

$$+ \log\Phi[(\mu/\lambda_i - S\varepsilon_i\lambda_i)/\sigma_i] - \log\Phi(\mu/\sigma_{u.i}).$$

We build the structure of the model with two freely varying variance parameters, $\sigma_{u,i}$ and $\sigma_{v,i}$, rather than the reduced form parameters λ and σ. The use of λ_i as a free parameter would not be appropriate because the numerator and denominator of λ_i must be allowed to vary freely and independently. A like consideration rules out the composed parameter σ_i. The formulation of the log likelihood and its derivatives follows the results given earlier for the homogeneous cases. Where the derivatives with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ emerge, we use the chain rule to differentiate with respect to $\sigma_{u,i}$ and $\sigma_{v,i}$ first. Note that the independent parameter σ_u and σ_v have been absorbed into the exponential functions. Thus, σ_v is $\exp(\gamma_0)$. This ensures that the variances are always positive.

The normal-gamma and normal-exponential models are not reparameterized. The log likelihood for the exponential model with variance heterogeneity is

$$\log L_i = \log\theta_i + ½\theta_i^2\sigma_{i,v}^2 + \theta_i S\varepsilon_i + \log\Phi[-S\varepsilon_i/\sigma_{i,v} - \theta_i\sigma_{i,v}]$$

where

$$\theta_i = \theta\exp(-\boldsymbol{\gamma}'\mathbf{z}_i)$$

and

$$\sigma_{i,v} = \sigma_v\exp(\boldsymbol{\delta}'\mathbf{w}_i).$$

The sign change in θ_i is used to make the normal-exponential model comparable to the normal-half normal model, since $\text{Var}[u_i] = 1/\theta_i^2$.

E63.3 The Normal-Truncated Normal Model

The normal-truncated normal model relaxes an implicit restriction in the normal-half normal model, that the mean of the underlying inefficiency variable is zero. The extended model is obtained by allowing μ, the mean of U, to be nonzero;

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, u = |U|$$
$$U \sim N[\mu,\sigma_u^2]$$
$$v \sim N[0,\sigma_v^2]$$

(With a constant term in the model, no similar parameter can be introduced into the distribution of v.) The command for estimating this model is

```
FRONTIER    ; Lhs = dependent variable
            ; Rhs = one, other independent variables
            ; Model = Truncated Normal $ (or ; Model = T)
```

The specification of the cost frontier and the estimator of technical inefficiency are requested in the same fashion,

; Cost

and **; Eff = variable name**

Other optional parts of the command are the same as that for the normal-half normal model.

We note, this model is extremely volatile, owing to the rather weak identification of the parameter μ. It is difficult to distinguish the mean from the variance parameter in this model. In the truncation model,

$$E[u_i] = \mu + \sigma_u\phi(\mu/\sigma_u)/\Phi(\mu/\sigma_u).$$

This implies that σ_u and μ can covary so as to produce little or no variation in the expectation of u_i. The likelihood is not a function of the square of u_i, so this mean is the only source of information about these two parameters. (By totally differentiating the expected value, one can solve for the implicit relationship, $d\mu/d\sigma_u$ that produces $dE[u_i] = 0$.) The example below suggests how this aspect of the model influences (or fails to) the estimates of inefficiency. For purposes of the JLMS estimator for the half normal model, when the mean of U is a nonzero μ, the argument to the function is replaced with

$$w = S\varepsilon\lambda/\sigma - \mu/(\sigma\lambda).$$

The remaining part of the computation is the same.

E63.3.1 Application

The results below show estimates of a stochastic cost frontier with the half normal then the truncated normal specifications. The additional parameterization appears to have had a large impact on the results; the estimates are noticeably different. The plot of the two sets of inefficiency estimates suggest that the effect of the new specification has been little more than to double the estimated values from the model – the dashed line in the figure shows the function $u_{TN} = 2u_{HN}$. The extremely large estimates of μ and the standard error do suggest that something is amiss with the model, however.

The commands are:

FRONTIER ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = u $
FRONTIER ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = ut ; Model = T $
PLOT ; Lhs = u ; Rhs = ut ; Rh2 = u ; Fill ; Grid
; Title = Truncated Normal Inefficiencies vs. Half Normal $
DSTAT ; Rhs = u,ut $

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       108.43918
Variances: Sigma-squared(v)=     .01902
           Sigma-squared(u)=     .01692
           Sigma(v)        =     .13791
           Sigma(u)        =     .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=     .18957
Gamma = sigma(u)^2/sigma^2 =     .47074
Var[u]/{Var[u]+Var[v]}     =     .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***       .72136    -4.14  .0000    -4.40206  -1.57439
      LL|    -.42909***       .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***       .09498     4.69  .0000      .25917    .63149
      LF|     .37257***       .07038     5.29  .0000      .23463    .51052
      LE|    2.09473***       .68790     3.05  .0023      .74647   3.44299
      LM|     .69910***       .07580     9.22  .0000      .55054    .84766
      LK|   -2.09806***       .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***       .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***       .00064   297.81  .0000      .18832    .19082
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       109.49695
Estimation based on N =    256, K =  10
Variances: Sigma-squared(v)=     .01896
           Sigma-squared(u)=    2.48813
           Sigma(v)        =     .13771
           Sigma(u)        =    1.57738
Sigma = Sqr[(s^2(u)+s^2(v)]=    1.58338
Gamma = sigma(u)^2/sigma^2 =     .99244
Var[u]/{Var[u]+Var[v]}     =     .97946
Stochastic Production Frontier, e = v-u
Half Normal:u(i)=|U(i)|; frontier model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.845
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.11541***      .77143    -4.04  .0001    -4.62739  -1.60343
      LL|    -.44532***      .07797    -5.71  .0000     -.59814   -.29249
      LP|     .46908***      .11368     4.13  .0000      .24628    .69188
      LF|     .37437***      .07465     5.02  .0000      .22807    .52068
      LE|    2.20830***      .73883     2.99  .0028      .76023   3.65637
      LM|     .67741***      .09341     7.25  .0000      .49433    .86048
      LK|   -2.20620***      .82402    -2.68  .0074    -3.82126   -.59115
        |Offset [mean=mu(i)] parameters in one sided error
      Mu|   -31.5468       5061.203     -.01  .9950  -9951.3228  9888.2292
        |Variance parameters for compound error
  Lambda|    11.4545       907.8501      .01  .9899  -1767.8991  1790.8081
   Sigma|    1.58338       124.7546      .01  .9899  -242.93113  246.09790
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|       Mean        Std.Dev.     Minimum      Maximum     Cases Missing
--------+---------------------------------------------------------------------
       U|     .902312      .035500      .703534      .963108       256        0
      UT|     .925474      .039335      .608274      .972355       256        0
--------+---------------------------------------------------------------------
```

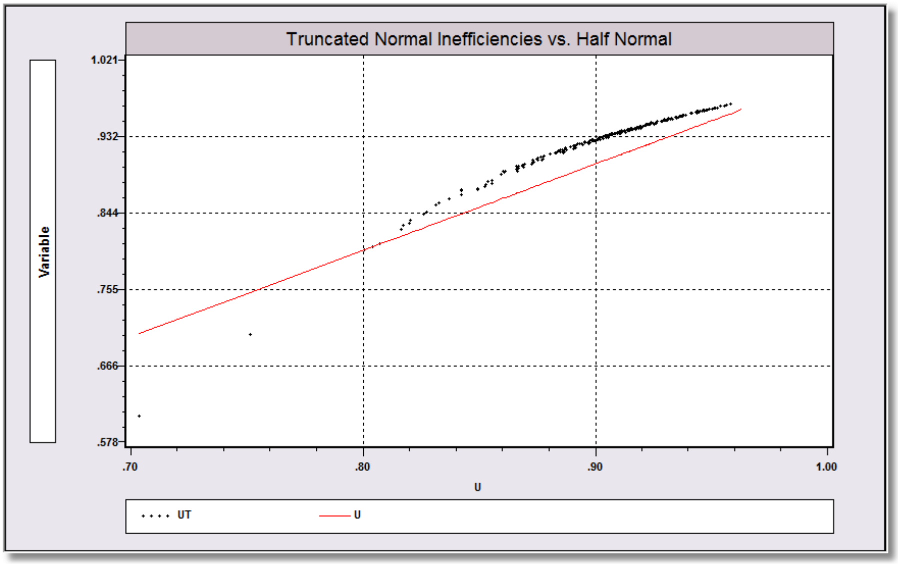


Figure E63.3 Inefficiency Estimates from Truncated Normal Model

E63.3.2 Battese and Coelli (1995) Formulation

There are (apparently) two formulations of the normal – truncated normal model in the literature. The formulated above,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, u = |U|$$

$$U \sim N[\mu,\sigma_u^2] \leftarrow$$

$$v \sim N[0,\sigma_v^2]$$

is due to Stevenson (1980). Note that the inefficiency term is the absolute value of a normally distributed variable with a nonzero mean. Battese and Coelli proposed an apparently different formulation of the truncation model;

$$u = \mu + w$$

where w is a truncated normal, such that

$$w \geq -\mu.$$

This is actually the same model. You can obtain the estimates using this alternative formulation with

; Model = BC95

in place of **; Model = T**. The log likelihood for this formulation involves a one to one reparameterization of the Stevenson model, which has slightly different numerical properties. You can see this in the application below. The estimated inefficiency and efficiency values produced by the two models are the same to five or six digits, however.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       109.48819
Variances: Sigma-squared(v)=     .01918
           Sigma-squared(u)=    2.25705
           Sigma(v)        =     .13850
           Sigma(u)        =    1.50235
Sigma = Sqr[(s^2(u)+s^2(v)]=    1.50872
Gamma = sigma(u)^2/sigma^2 =     .99157
Var[u]/{Var[u]+Var[v]}     =     .97715
Stochastic Production Frontier, e = v-u
Battese/Coelli 1995 truncated normal model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     1
Deg. freedom for inefficiency model:  2
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.828
Kodde-Palm C*: 95%: 5.138,  99%:  8.273
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.09929***      .76919    -4.03  .0001    -4.60687  -1.59172
      LL|    -.44370***      .07771    -5.71  .0000     -.59600   -.29140
      LP|     .46535***      .11351     4.10  .0000      .24288    .68781
      LF|     .37430***      .07432     5.04  .0000      .22863    .51997
      LE|    2.18991***      .73664     2.97  .0030      .74613   3.63369
      LM|     .67921***      .09322     7.29  .0000      .49651    .86191
      LK|   -2.18647***      .82171    -2.66  .0078    -3.79700   -.57594
        |Offset [mean=z(i)*delta] parameters in one sided error
Constant|   -29.6062       4821.053     -.01  .9951  -9478.6972  9419.4848
        |Variance parameters for compound error
   Gamma|     .99157        1.34377      .74  .4606    -1.64216   3.62531
SigmaSqd|    2.27624       363.5754      .01  .9950  -710.31839  714.87086
--------+--------------------------------------------------------------------
 (Stevenson formulation)
Log likelihood function        94.86417
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.11541***      .77143    -4.04  .0001    -4.62739  -1.60343
      LL|    -.44532***      .07797    -5.71  .0000     -.59814   -.29249
      LP|     .46908***      .11368     4.13  .0000      .24628    .69188
      LF|     .37437***      .07465     5.02  .0000      .22807    .52068
      LE|    2.20830***      .73883     2.99  .0028      .76023   3.65637
      LM|     .67741***      .09341     7.25  .0000      .49433    .86048
      LK|   -2.20620***      .82402    -2.68  .0074    -3.82126   -.59115
        |Offset [mean=mu(i)] parameters in one sided error
      Mu|   -31.5468       5061.203     -.01  .9950  -9951.3228  9888.2292
        |Variance parameters for compound error
  Lambda|    11.4545       907.8501      .01  .9899  -1767.8991  1790.8081
   Sigma|    1.58338       124.7546      .01  .9899  -242.93113  246.09790
```

E63.3.3 Technical Details on the Truncated Normal Model

The individual term in the log likelihood for the normal-truncated normal model is

$$\log L_i = -\tfrac{1}{2}\log 2\pi - \log\sigma - \tfrac{1}{2}[(S\varepsilon_i + \mu)/\sigma]^2 - \log\Phi(\mu/\sigma_u) + \log\Phi[(\mu/\lambda - S\varepsilon_i\lambda)/\sigma].$$

The definitions above imply that

$$\sigma_u = \sigma\lambda/\sqrt{1+\lambda^2}\ .$$

Using this and the reparameterization

$$\alpha = \mu/(\lambda\sigma)$$

produces the log likelihood for this model,

$$\text{Log } L_i = -\tfrac{1}{2}\log 2\pi - \log\sigma - \tfrac{1}{2}(d\varepsilon_i/\sigma + \alpha\lambda)^2 - \log\Phi\left(\alpha\sqrt{1+\lambda^2}\right) + \log\Phi(\alpha - d\varepsilon_i\lambda/\sigma).$$

The function is then maximized with respect to β, σ, λ and α. After optimization, the structural parameter μ is recovered from the result $\mu = \alpha\sigma\lambda$. For the model with heterogeneity in the mean presented in Section E63.3.4,

$$\mu_i = \boldsymbol{\theta}'\mathbf{z}_i$$

we simply replace α with $\alpha_i = \boldsymbol{\alpha}'\mathbf{z}_i$, then recover the parameter vector $\boldsymbol{\theta}$ from the same transformation as before, $\boldsymbol{\theta} = \sigma\lambda\boldsymbol{\alpha}$.

For purposes of the JLMS estimator for the half normal model, when the mean of U is a nonzero μ, the argument to the function is replaced with

$$w = S\varepsilon\lambda/\sigma - \mu/(\sigma\lambda).$$

The remaining part of the computation is the same.

E63.3.4 Heterogeneity in the Mean in the Truncation Model

The models listed above are all ‘homogeneous.’ Both the means and the variances of the underlying disturbance distributions are constant. There are several models of heterogeneity available as well. Use

; Model = T ; Rh2 = list of variables that enter the mean

to specify the heterogeneity in mean model, $U_i \sim N[\boldsymbol{\alpha}'\mathbf{z}_i, \sigma_u^2]$. In formulating this model, though it is not required, you should include a constant in $\mathbf{z}_i$ (the Rh2 variables) so that the homogeneous model becomes a special case. Also, if you are fitting a panel data version of this, note that the assumption underlying the model is that the same u_i occurs in every period. Therefore, the $\boldsymbol{\alpha}'\mathbf{z}_i$ should be the same in every period. *LIMDEP* will assume this is the case, and only use the Rh2 variables provided for the first period.

E63.3.5 Truncation and Heteroscedasticity

The doubly heteroscedastic model is also available for the truncated normal stochastic frontier model. In

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + v_i - u_i$$

you may specify **; Model = Truncated Normal; Rh2 = list of variables**

and $\text{Var}[u_i] = \sigma_u^2 \exp(\boldsymbol{\delta}'\mathbf{z}_i)$ with

; Het ; Hfu = list of variables in $\mathbf{z}_i$

and/or $\text{Var}[v_i] = \sigma_v^2 \exp(\boldsymbol{\gamma}'\mathbf{w}_i)$ with

; Het ; Hfv = list of variables in $\mathbf{w}_i$

Note that since both variance functions have a free multiplicative constant, you should not include *one* in either variable list.

In the absence of the Rh2 list, the mean of the underlying truncated variable is taken to be a constant to be estimated. This formulation encompasses all of Stevenson (1980), Reifschneider and Stevenson (1991), Huang and Liu (1994), and Battese and Coelli (1995). (Notwithstanding the assertion in the Battese and Coelli paper, the latter is not a panel data treatment as observations are still assumed to be independent.)

To illustrate the truncated normal estimator, we have refit the stochastic frontier production function with a complete set of firm dummy variables (less the last one) and the load factor variable in the mean of the underlying distribution. In the second model below, we have made the variance of *v* a function of the log of the average stage length. The command set begins with a small repair to the data set. One of the firms has no observations for the load factor, stage length or points served variables – they are coded as zero in the data. These observations are bypassed, then the firm dummies for the fixed effects model are assembled.

```
SAMPLE     ; All $
REJECT     ; loadfctr = 0 $
CREATE     ; i = Seq(firm) $
CREATE     ; Expand(i,0) $
CREATE     ; lk = Log(k) $
NAMELIST   ; xp = one,lf,lm,le,ll,lp,lk $
FRONTIER   ; Lhs = lq ; Rhs = xp ; Model = T ; Rh2 = loadfctr,_i_ $
FRONTIER   ; Lhs = lq ; Rhs = xp ; Model = T ; Rh2 = loadfctr,_i_
           ; Het ; Hfv = lstage $
```

(These are ‘true fixed effects’ models.)

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       196.20748
Estimation based on N =    256, K =  34
Inf.Cr.AIC  =   -324.4 AIC/N =   -1.267
Model estimated: Aug 22, 2011, 22:29:09
Variances: Sigma-squared(v)=     .00960
           Sigma-squared(u)=     .00389
           Sigma(v)        =     .09799
           Sigma(u)        =     .06241
Sigma = Sqr[(s^2(u)+s^2(v)]=     .11618
Gamma = sigma(u)^2/sigma^2 =     .28856
Var[u]/{Var[u]+Var[v]}     =     .12845
Stochastic Production Frontier, e = v-u
Half Normal:u(i)=|U(i)|; frontier model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:    25
Deg. freedom for inefficiency model: 26
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =  176.266
Kodde-Palm C*: 95%:38.301,  99%: 45.026
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.92400***      .68225    -4.29  .0000    -4.26118  -1.58682
      LF|     .31938***      .09026     3.54  .0004      .14246    .49629
      LM|     .81647***      .08387     9.73  .0000      .65209    .98086
      LE|    1.99934***      .64368     3.11  .0019      .73776   3.26092
      LL|    -.42790***      .10954    -3.91  .0001     -.64260   -.21321
      LP|     .42291***      .10529     4.02  .0001      .21654    .62929
      LK|   -2.07145***      .72267    -2.87  .0042    -3.48786   -.65503
        |Offset [mean=mu(i)] parameters in one sided error
LOADFCTR|    -.83124        6.87337     -.12  .9037   -14.30280  12.64031
     I01|     .63250        4.90139      .13  .8973    -8.97405  10.23904
     I02|     .58118        4.27763      .14  .8919    -7.80282   8.96519
(Firms 3-21 omitted)
     I22|     .45249        4.00889      .11  .9101    -7.40480   8.30977
     I23|     .64687       99.45841      .01  .9948  -194.28803  195.58176
     I24|    -.19804        7.26011     -.03  .9782   -14.42760  14.03152
        |Variance parameters for compound error
  Lambda|     .63686**       .28984     2.20  .0280      .06879   1.20494
   Sigma|     .11618***      .01008    11.53  .0000      .09643    .13593
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LQ
Log likelihood function        215.58601
Estimation based on N =    256, K =  35
Variances: Sigma-squared(v)=      .00634
           Sigma-squared(u)=      .01037
           Sigma(u)        =      .10183
           Sigma(v)        =      .07961
Sigma = Sqr[(s^2(u)+s^2(v)]=      .12926
Variances averaged over observations
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     25
Deg. freedom for inefficiency model: 26
LogL when sigma(u)=0           108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =  215.023
Kodde-Palm C*: 95%:38.301,  99%: 45.026
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -1.98442*       1.05055    -1.89  .0589    -4.04346    .07463
      LF|     .45669***      .11002     4.15  .0000      .24105    .67233
      LM|     .59013***      .10421     5.66  .0000      .38589    .79437
      LE|    1.11856        1.00928     1.11  .2677     -.85959   3.09671
      LL|    -.29237***      .10923    -2.68  .0074     -.50646   -.07827
      LP|     .31311**       .14333     2.18  .0289      .03220    .59402
      LK|   -1.14743        1.10875    -1.03  .3007    -3.32054   1.02568
        |Mean of underlying truncated distribution
LOADFCTR|   -2.20067***      .42161    -5.22  .0000    -3.02701  -1.37433
     I01|    1.44767***      .25736     5.63  .0000      .94326   1.95208
     I02|    1.39624***      .22401     6.23  .0000      .95718   1.83529
(Firms 3-22 omitted)
     I24|    1.29355***      .24998     5.17  .0000      .80360   1.78349
        |Scale parms. for random components of e(i)
ln_sgmaU|   -2.28443***      .02100  -108.79  .0000    -2.32559  -2.24328
ln_sgmaV|   -3.22203***     1.20573    -2.67  .0075    -5.58522   -.85884
        |Heteroscedasticity in variance of symmetric v(i)
  LSTAGE|     .11855         .19755      .60  .5485     -.26865    .50574
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E63.4 Alvarez et al. – Equality Constrained Scaling Model

Alvarez, Amsler, Orea and Schmidt (2006) have suggested a form of the truncation model which encompasses a number of ideas in stochastic frontier modeling. Their formulation is a normal-truncated normal frontier model with

$$\mu_i = \mu\times\boldsymbol{\delta}'\mathbf{z}_i \text{ and } \sigma_{u,i} = \sigma_u\times\boldsymbol{\delta}'\mathbf{z}_i.$$

The mean and standard deviation of the underlying truncated normal variable u_i are scaled by the same linear function of the data. We are skeptical of the linear scaling of the variance, and propose our usual exponential form instead. The linear form may be natural for the mean, but it allows the variance to be negative, which is unacceptable. The model used here is

$$\mu_i = \mu\times\exp(\boldsymbol{\delta}'\mathbf{z}_i) \text{ and } \sigma_{u,i} = \sigma_u\times \exp(\boldsymbol{\gamma}'\mathbf{z}_i).$$

The Alvarez model results if $\boldsymbol{\delta} = \boldsymbol{\gamma}$. Otherwise, we allow these to be free and to produce another variant of the frontier model. Note that as stated, this model is now merely a change of the normal-truncated normal model with heteroscedasticity in which the variables enter the truncation mean function in the exponential function rather than linearly.

The equality constrained scaling model is requested with

```
FRONTIER   ; Lhs  = y ; Rhs = one, x...
           ; Model = Scaling
           ; Heteroscedasticity
           ; Rh2 = variables in mean of truncated distribution
           ; Hfu  = the same list of variables $
```

Note in this case, Rh2 and Hfu give the same list. To obtain the scaling model without forcing the equality of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, use

```
FRONTIER   ; Lhs  = y ; Rhs = one, x...
           ; Model = S
           ; Heteroscedasticity
           ; Rh2 = variables in mean of truncated distribution
           ; Hfu  = the same list of variables $
```

Note, **; Model = Scaling** in the equality constrained case and **; Model = S** when the equality constraint is relaxed. (In this formulation, the variable lists could differ.) To constrain $\boldsymbol{\delta} = \mathbf{0}$, which just produces the heteroscedasticity model, use

```
FRONTIER   ; Lhs  = y ; Rhs = one, x...
           ; Model = T
           ; Heteroscedasticity
           ; Hfu  = list of variables $
```

To constrain γ = **0**, you would use the available setup for the truncated normal form, but **; Model = S** rather than **; Model = T** to obtain the exponential scaling of the mean.

```
FRONTIER    ; Lhs = y ; Rhs = one, x...
            ; Model = S
            ; Rh2 = variables in mean of truncated distribution $
```

Finally, with both δ = **0** and γ = **0**, this is just the standard normal-truncated normal model.

Technical Details

The implementation of the scaling model in *LIMDEP* is just a version of the truncation model with heteroscedasticity. The modifications of that model are:

- The constant terms in the mean and variance are enforced by the program.
- The mean function is exponential.
- In the first form of the model, a constraint is imposed that the coefficients in the mean and variance functions are the same.

As Alvarez et al. note in their paper, this model is not supported by any particular theory of the frontier framework. They suggest it as a natural extension of the familiar model with truncation. Rather, they argue that the unnatural form of the model would be the one with different scaling factors in the mean and variance functions.

Application

To illustrate the scaling model, we use the airlines cost data. The cost function is fit with truncation mean and variance functions that depend on the load factor and (log of) the average stage length. The equality constraint is imposed in the first model and relaxed in the second.

```
FRONTIER    ; Lhs  = lc ; Cost ; Rhs  = x
            ; Model = Scaling ; Het
            ; Rh2 = loadfctr,lstage
            ; Hfu  = loadfctr,lstage $
FRONTIER    ; Lhs  = lc ; Cost ; Rhs  = x
            ; Model = S ; Het
            ; Rh2 = loadfctr,lstage
            ; Hfu  = loadfctr,lstage $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                       LC
Log likelihood function       172.27160
Estimation based on N =    256, K =  13
Variances: Sigma-squared(v)=      .01528
           Sigma-squared(u)=      .00000
           Sigma(v)        =      .12361
           Sigma(u)        =      .00169
Sigma = Sqr[(s^2(u)+s^2(v)]=      .12363
Stochastic Frontier Scaling Model
Mean scale factor for E[u] =         .6996
Mean scale factor for V[u] =         .6996
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  5
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =   28.713
Kodde-Palm C*: 95%:10.371,  99%: 14.325
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    18.9477       27.00668       .70  .4829    -33.9844   71.8798
      LY|     .95234***      .02117     44.98  .0000      .91084    .99383
     LY2|     .07740***      .01534      5.04  .0000      .04733    .10747
    LPKP|    1.50434        1.86479       .81  .4198    -2.15058   5.15926
    LPLP|     .12682         .08328      1.52  .1278     -.03640    .29003
    LPMP|    -.16640        1.21907      -.14  .8914    -2.55574   2.22294
    LPEP|    -.52809         .60356      -.87  .3816    -1.71105    .65488
    LPFP|     .00151         .02141       .07  .9436     -.04045    .04348
        |Mean of Truncated Distribution, Mu then scale
    Mu_0|    2.50985       11.12070       .23  .8214   -19.28633  24.30603
LOADFCTR|    -.56559        3.85231      -.15  .8833    -8.11597   6.98479
  LSTAGE|    -.00823         .05624      -.15  .8837     -.11845    .10200
        |Standard Deviation of u: Sigma(u) then scale
Sigmau_0|     .00241        9.18604       .00  .9998   -18.00191  18.00673
LOADFCTR|    -.56559        3.85231      -.15  .8833    -8.11597   6.98479
  LSTAGE|    -.00823         .05624      -.15  .8837     -.11845    .10200
        |Standard deviation of v
Sigma(v)|     .12361         .08711      1.42  .1559     -.04713    .29435
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LC
Log likelihood function        173.52520
Estimation based on N =    256, K =  15
Variances: Sigma-squared(v)=      .01334
           Sigma-squared(u)=      .00121
           Sigma(v)        =      .11551
           Sigma(u)        =      .03476
Sigma = Sqr[(s^2(u)+s^2(v)]=      .19230
Stochastic Frontier Scaling Model
Mean scale factor for E[u] =        .3459
Mean scale factor for V[u] =        .2261
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  5
LogL when sigma(u)=0           157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =   31.220
Kodde-Palm C*: 95%:10.371,  99%: 14.325
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.6452       24.94703      .47  .6406    -37.2501   60.5405
      LY|     .94078***      .02140    43.97  .0000      .89884    .98272
     LY2|     .06680***      .01579     4.23  .0000      .03585    .09776
    LPKP|     .85146        1.94378      .44  .6614    -2.95828   4.66120
    LPLP|     .16345**       .07956     2.05  .0399      .00751    .31939
    LPMP|     .25417        1.26886      .20  .8412    -2.23275   2.74109
    LPEP|    -.34167         .62932     -.54  .5872    -1.57511    .89178
    LPFP|     .00164         .02164      .08  .9395     -.04078    .04406
        |Mean of Truncated Distribution, Mu then scale
    Mu_0|    1.92288***      .44030     4.37  .0000     1.05991   2.78584
LOADFCTR|   -1.74305        4.08382     -.43  .6695    -9.74720   6.26110
  LSTAGE|    -.01930         .04649     -.42  .6781     -.11042    .07182
        |Standard Deviation of u: Sigma(u) then scale
Sigmau_0|     .15374        1.11571      .14  .8904    -2.03301   2.34049
LOADFCTR|   -14.5014       10.21457    -1.42  .1557    -34.5216    5.5188
  LSTAGE|    1.02454        1.26499      .81  .4180    -1.45479   3.50388
        |Standard deviation of v
Sigma(v)|     .11551***      .00793    14.56  .0000      .09996    .13106
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E64: Panel Data Stochastic Frontier Models

E64.1 Introduction

The stochastic frontier model as it appears in the current literature was originally developed by Aigner, Lovell, and Schmidt (1977). The canonical formulation that serves as the foundation for other variations is their model,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u,$$

where y is the observed outcome (goal attainment), $\boldsymbol{\beta}'\mathbf{x} + v$ is the optimal, frontier goal (e.g., maximal production output or minimum cost) pursued by the individual, $\boldsymbol{\beta}'\mathbf{x}$ is the deterministic part of the frontier and $v \sim N[0,\sigma_v^2]$ is the stochastic part. The two parts together constitute the 'stochastic frontier.' The amount by which the observed individual fails to reach the optimum (the frontier) is u, where

$$u = |U| \text{ and } U \sim N[0,\sigma_u^2]$$

(change to $v + u$ for a stochastic cost frontier or any setting in which the optimum is a minimum). In this context, u is the 'inefficiency.' This is the normal-half normal model which forms the basic form of the stochastic frontier model. Chapters E62 and E63 developed several versions of the stochastic frontier model suitable for cross section and pooled data sets. This chapter will develop versions of the model constructed specifically for panel data.

E64.2 Panel Data Estimators for Stochastic Frontier Models

The stochastic frontiers literature has steadily evolved since the developments of basic random and fixed effects models by Pitt and Lee (1981) and by Cornwell, Schmidt and Sickles (1990). All of the generally used forms of panel data models are supported in *LIMDEP*. The following will document them in detail. These sections are arranged as follows:

- Pitt and Lee – Time Invariant Inefficiency, Random Effects,
- Cornwell, Schmidt and Sickles – Time Invariant Inefficiency, Fixed Effects,
- Battese and Coelli – Time Dependent Inefficiency Models,
- True Fixed Effects Models with Time Varying Inefficiency,
- True Random Effects Models with Time Varying Inefficiency,
- Random Parameters Stochastic Frontier Models,
- Alvarez et al. – Fixed Management (Random Parameters) Model,
- Latent Class Stochastic Frontier Models.

The panel models developed here will share features with other panel models in *LIMDEP*, as presented in Chapters R22-R25. As in other settings, panels in all models may be unbalanced. Panels are identified by

SETPANEL ; … $

then **; Panel**
in the command, or **; Pds = group count**

Nearly all of the models to be presented here actually require panel data, but a few will work, albeit not as well as otherwise, with **; Pds = 1**, i.e., with a cross section. This will be specifically noted below when it is the case. Second, in all models, the cost form as opposed to the production form is requested with

; Cost

This and other model specifications are generally the same as the cross sectional cases.

E64.3 Pitt and Lee – Time Invariant Inefficiency, Random Effects

The panel data, random effects specifications based on the model of Pitt and Lee (1981) are

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_i$$

with S = +1 for a production model and -1 for a cost model. The inefficiency component is assumed to be time invariant. The base case is the normal-half normal model

$$u_i = |U_i|,\ U_i \sim \text{N}[0,\sigma^2].$$

This is a direct extension of the cross section variant discussed earlier. Several model formulations are grouped in this class. The command for the Pitt and Lee group of models is given by changing the base case specifications to

FRONTIER ; Lhs = y ; Rhs = one, ... ; Panel $

Pitt and Lee is the default panel data model. The only necessary change for the default case is specification of the panel with **; Panel**. As in the cross section case, the normal-exponential case is requested with

; Model = Exponential

while the normal-truncated normal is requested with

; Rh2 = one or ; Rh2 = one, additional variables

(The **; Model = T** is not needed.) The truncation model may not be combined with the exponential specification; it is only supported for the normal-truncated normal form.

NOTE: The gamma model does not have a random effects (panel data) version. The model extensions, such as the scaling model and sample selection described in Chapter E63 likewise do not support a Pitt and Lee style random effects version.

There is an important consideration for the truncation version with heterogeneous mean. If you are fitting a panel data version of this model, note that the assumption underlying the model is that the same u_i occurs in every period. Therefore, the $\boldsymbol{\alpha}'\mathbf{z}_i$ must be the same in every period. *LIMDEP* will assume this is the case, and only use the Rh2 variables provided for the first period.

When the random effects model is estimated, maximum likelihood estimates of the cross section models are always computed first to obtain the starting values. This will produce a full set of results which will ignore the panel nature of the data set. A second full set of results will then follow for the random effects model.

The model estimates retained for all cases are

b = regression parameters, α,$\boldsymbol{\beta}$
varb = asymptotic covariance matrix.

Use **; Par** to retain the additional parameters in *b* and *varb*. As seen in the applications below, the parameters estimated in each case will differ depending on the model formulation. The ancillary parameters that are estimated for the various models are the same ones saved by the cross section versions. All models save *sy*, *ybar*, *nreg*, *kreg*, and *logl* as well as *s*, *b*, *varb*, etc.

WARNING: Numerous experiments and applications have suggested that the normal-truncated normal model is a difficult one to estimate. Identification appears to be highly variable, and small variations in the data can produce large variation in the results. The model often fails to converge even when convergence of the restricted model with zero underlying mean is routine.

E64.3.1 Model Specifications

There are many different combinations of the components of the random effects model listed above. The following shows the different possibilities for the Pitt and Lee model. (There are also many combinations of these that do not use the panel data random effects form.):

NAMELIST ; x = one, … $
CREATE ; y = the outcome variable $
SETPANEL ; … $

Model 1 = pooled
FRONTIER ; Lhs = y ; Rhs = x $

Model 2 = random effects half normal
FRONTIER ; Lhs = y ; Rhs = x ; Panel $

Model 3 = random effects exponential
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Model = Exponential $

Model 4 = random effects normal heteroscedastic in *u* or *v* only
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Het ; Hfv = … $
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Het ; Hfu = … $

Model 5 = random effects normal doubly heteroscedastic
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Het ; Hfv = … ; Hfu = … $

Model 6 = random effects truncated normal
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Rh2 = one, … $

Model 7 = random effects truncated normal, singly or doubly heteroscedastic
FRONTIER ; Lhs = y ; Rhs = x ; Panel ; Rh2 = one, …
; Het ; Hfv = … ; Hfu = … $

The Pitt and Lee model forms assume that the inefficiency is time invariant. Thus, the estimate of u_i is repeated for each observation in the group. An example below illustrates.

E64.3.2 Applications

The following illustrates a few of the numerous formats of the random effects frontiers. The data set used is the Swiss railroad data used in Greene (2011, Table F19.1). These data are provided with the program as Swiss-railroads.lpj. The variables used here are

ct	= total cost
pk	= capital price
pe	= electricity price
pl	= labor price
q2	= passenger output – passenger km
q3	= freight output – ton km
rack	= dummy variable for 'rack rail' in network
tunnel	= dummy variable for network with tunnels over 300 meters on average
virage	= dummy variable for networks with narrow radius curvature
narrow_t	= dummy variable for narrow track (1m as opposed to standard 1.435m).

Preparing the data set includes bypassing one firm for which there is only a single year of data. For the remaining 49 firms, T_i is a mixture 3, 7, 10, 12 or 13. Figure E64.1 details the distribution of group sizes.

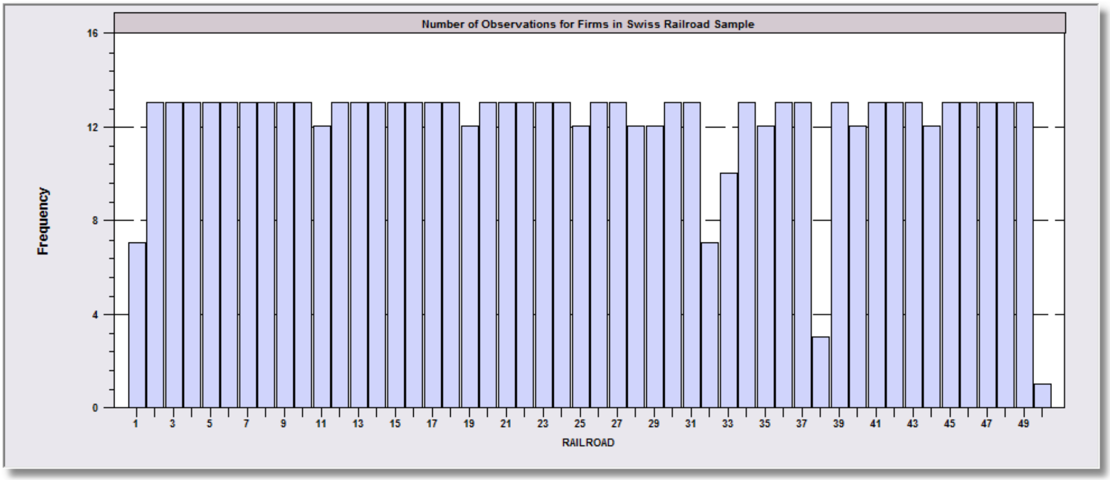


Figure E64.1 Groups Sizes for Swiss Railroad Sample

Descriptive statistics for the data are shown below. Variables with names beginning with 'M' are firm means, repeated for each year for the firm.

We fit four models to illustrate the estimator, the pooled normal-half normal, pooled normal-truncated (heterogeneous), basic Pitt and Lee and a full model with time invariant inefficiency, truncation (heterogeneous) and double heteroscedasticity.

The commands are as follows:

```
SETPANEL    ; Group = id ; Pds = ti $
REJECT      ; ti = 1 $
CREATE      ; lple = Log(pl/pe) ; lpke = Log(pk/pe) ; lnc = Log(ct/pe)$
NAMELIST    ; x = one,lnq2,lnq3,lple,lpke $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; Costeff = eusfpool $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; Panel ; Costeff = eusfp_l $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; Rh2 = rack,tunnel
            ; Het ; Hfu = virage ; Hfv = virage ; Costeff = eushet_t $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; panel ; Rh2 = rack,tunnel
            ; Het ; Hfu = virage ; Hfv = virage ; Costeff = fullmodl $
```

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
ID	25.48760	14.60037	1.0	51.0	605	0
YEAR	90.91570	3.692372	85.0	97.0	605	0
NI	12.58347	1.305259	1.0	13.0	605	0
STOPS	20.42479	18.48285	4.0	121.0	605	0
NETWORK	39431.66	56642.38	3898.0	376997.0	605	0
LABOREXP	12801.95	26232.69	951.0	173549.0	605	0
STAFF	170.3810	333.0317	11.0	1934.0	605	0
ELECEXP	968.1521	1944.830	14.0	14737.0	605	0
KWH	7602.221	15608.39	82.0	104923.0	605	0
TOTCOST	22470.44	42283.57	1534.0	280871.0	605	0
NARROW_T	.676033	.468375	0.0	1.0	605	0
RACK	.234711	.424169	0.0	1.0	605	0
TUNNEL	.188430	.391379	0.0	1.0	605	0
T	5.915702	3.692372	0.0	12.0	605	0
Q1	813914.0	1083923	61000.0	6409000	605	0
Q2	.308145D+08	.550599D+08	409000.0	.311000D+09	605	0
Q3	.101934D+08	.527303D+08	150.0	.477000D+09	605	0
CT	26728.37	49883.51	2120.968	307433.4	605	0
PL	86051.77	6484.535	60932.91	104930.4	605	0
PE	.157485	.022766	.076344	.265182	605	0
PK	4534.491	2128.307	1040.323	14466.06	605	0
VIRAGE	.715702	.451452	0.0	1.0	605	0
LABOR	52.40245	9.598136	20.03025	73.11581	605	0
ELEC	4.044504	1.422098	.568412	9.311660	605	0
CAPITAL	43.55305	9.461303	23.88916	77.33154	605	0
LNCT	11.30622	1.101691	9.462956	14.57019	605	0
LNQ1	13.06322	1.010039	11.01863	15.67321	605	0
LNQ2	16.31759	1.339167	12.92147	19.55500	605	0
LNQ3	12.49439	2.716709	5.010635	19.98343	605	0
LNNET	3.200860	.908512	1.360464	5.932237	605	0
LNPL	13.21935	.163565	12.60449	13.77599	605	0
LNPE	-1.859557	.152870	-2.572503	-1.327338	605	0
LNPK	10.17950	.438886	8.740266	11.37466	605	0

```
 LNSTOP|     2.775052      .655071     1.386294     4.795791      605        0
  LNCAP|     3.137572      .328311     2.123893     3.850147      604        1
  MLNQ1|     13.06322     1.005089     11.16747     15.59433      605        0
  MLNQ2|     16.31759     1.333346     13.20185     19.45679      605        0
  MLNQ3|     12.49439     2.648475     7.734539     19.68075      605        0
 MLNNET|     3.200860      .906363     1.360464     5.927817      605        0
  MLNPL|     13.21935      .126548     12.89796     13.61620      605        0
  MLNPK|     10.17950      .396797     8.938699     11.03543      605        0
MLNSTOP|     2.775052      .651059     1.386294     4.789402      605        0
   LPLE|     13.21943      .163692     12.60449     13.77599      604        1
  LPKPE|     10.16419      .576094          1.0     11.37466      605        0
    LNC|     11.30305     1.099836     9.462957     14.57019      604        1
--------+-------------------------------------------------------------------
```

This is the pooled normal-half normal model.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                  LNC
Log likelihood function      -209.42340
Estimation based on N =    604, K =   7
Inf.Cr.AIC  =    432.8 AIC/N =     .717
Variances: Sigma-squared(v)=     .07332
           Sigma-squared(u)=     .12333
           Sigma(v)        =     .27077
           Sigma(u)        =     .35119
Sigma = Sqr[(s^2(u)+s^2(v)]=     .44345
Gamma = sigma(u)^2/sigma^2 =     .62716
Var[u]/{Var[u]+Var[v]}     =     .37937
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0         -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.060
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -10.0907***     1.14284    -8.83  .0000    -12.3306   -7.8507
    LNQ2|     .64179***      .01371    46.80  .0000      .61491    .66867
    LNQ3|     .06855***      .00655    10.46  .0000      .05570    .08139
    LPLE|     .53971***      .08858     6.09  .0000      .36610    .71333
    LPKE|     .26045***      .03260     7.99  .0000      .19655    .32435
        |Variance parameters for compound error
  Lambda|    1.29697***      .13854     9.36  .0000     1.02545   1.56850
   Sigma|     .44345***      .00056   789.05  .0000      .44235    .44455
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the original Pitt and Lee normal-half normal model with time invariant inefficiency. In comparison to the pooled model above, σ_u has tripled and σ_v has decreased by two thirds. The assumption of time invariance of the inefficiency produces a large reallocation of the random components between noise and inefficiency. This is evident in the kernel estimate below as well.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LNC
Log likelihood function       527.11659
Estimation based on N =    604, K =   7
Inf.Cr.AIC  =  -1040.2 AIC/N =   -1.722
Stochastic frontier based on panel data
Estimation based on      49 individuals
Variances: Sigma-squared(v)=     .00621
           Sigma-squared(u)=     .92297
           Sigma(v)        =     .07879
           Sigma(u)        =     .96071
Sigma = Sqr[(s^2(u)+s^2(v)]=     .96394
Gamma = sigma(u)^2/sigma^2 =     .99332
Var[u]/{Var[u]+Var[v]}     =     .98183
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0         -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] = 1475.140
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -7.25643***      .24767   -29.30  .0000    -7.74185   -6.77101
    LNQ2|     .36259***      .01503    24.12  .0000      .33312     .39205
    LNQ3|     .01902***      .00240     7.94  .0000      .01432     .02372
    LPLE|     .64148***      .02112    30.38  .0000      .60009     .68287
    LPKE|     .30842***      .00700    44.08  .0000      .29471     .32214
        |Variance parameters for compound error
  Lambda|    12.1932**      5.55909     2.19  .0283      1.2975    23.0888
Sigma(u)|     .96071***      .13303     7.22  .0000      .69998    1.22145
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the pooled normal-truncated and doubly heteroscedastic normal model.

```
----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LNC
Log likelihood function         -63.43402
Estimation based on N =    604, K =  11
Inf.Cr.AIC  =    148.9 AIC/N =     .246
Variances: Sigma-squared(v)=     .07144
           Sigma-squared(u)=     .00074
           Sigma(u)        =     .02720
           Sigma(v)        =     .26729
Sigma = Sqr[(s^2(u)+s^2(v)]=     .26867
Variances averaged over observations
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  1
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  4
LogL when sigma(u)=0           -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] =  294.039
Kodde-Palm C*: 95%: 8.761,  99%: 12.483
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -13.4218***     1.01232   -13.26  .0000    -15.4059  -11.4377
    LNQ2|     .62859***      .01404    44.79  .0000      .60108    .65610
    LNQ3|     .09670***      .00669    14.46  .0000      .08359    .10981
    LPLE|     .68419***      .07646     8.95  .0000      .53433    .83405
    LPKE|     .39946***      .03301    12.10  .0000      .33476    .46415
        |Mean of underlying truncated distribution
    RACK|     .62333***      .05632    11.07  .0000      .51293    .73372
  TUNNEL|    -.35607***      .05500    -6.47  .0000     -.46387   -.24828
        |Scale parms. for random components of e(i)
ln_sgmaU|   -2.54850***      .96756    -2.63  .0084    -4.44488   -.65212
ln_sgmaV|   -1.36799***      .06507   -21.02  .0000    -1.49551  -1.24046
        |Heteroscedasticity in variance of truncated u(i)
  VIRAGE|   -1.47329        2.86559     -.51  .6072    -7.08975   4.14316
        |Heteroscedasticity in variance of symmetric v(i)
  VIRAGE|     .06774         .08094      .84  .4026     -.09090    .22638
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

This is the same model as immediately above, with the additional assumption that the inefficiency is time invariant. Compared to the previous specification, σ_u has now increased by a factor of 30 while σ_v has nearly vanished, falling from 0.27 to 0.005, that is, by a factor of 50.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LNC
Log likelihood function         532.94237
Estimation based on N =    604, K =  11
Inf.Cr.AIC  =  -1043.9 AIC/N =   -1.728
Variances: Sigma-squared(v)=      .00003
           Sigma-squared(u)=      .76238
           Sigma(u)        =      .87314
           Sigma(v)        =      .00543
Sigma = Sqr[(s^2(u)+s^2(v)]=      .87316
Variances averaged over observations
Stochastic frontier based on panel data
Estimation based on      49 individuals
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  1
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  4
LogL when sigma(u)=0         -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] = 1486.792
Kodde-Palm C*: 95%: 8.761,  99%: 12.483
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -7.26117***      .25317   -28.68  .0000    -7.75738  -6.76496
    LNQ2|     .36162***      .01558    23.20  .0000      .33107    .39216
    LNQ3|     .01947***      .00257     7.58  .0000      .01444    .02451
    LPLE|     .64342***      .02165    29.72  .0000      .60099    .68584
    LPKE|     .30730***      .00727    42.24  .0000      .29305    .32156
        |Mean of underlying truncated distribution
    RACK|     .81356         .52427     1.55  .1207     -.21399   1.84112
  TUNNEL|    1.46353***      .47072     3.11  .0019      .54094   2.38613
        |Scale parms. for random components of e(i)
ln_sgmaU|    -.17921         .21781     -.82  .4106     -.60611    .24769
ln_sgmaV|   -4.94678***      .20426   -24.22  .0000    -5.34711  -4.54644
        |Heteroscedasticity in variance of truncated u(i)
  VIRAGE|     .06076         .04703     1.29  .1964     -.03142    .15294
        |Heteroscedasticity in variance of symmetric v(i)
  VIRAGE|    -.37544         .44206     -.85  .3957    -1.24185    .49097
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

The kernel estimator compares the estimated cost efficiency distributions for the pooled and basic Pitt and Lee model. The pattern suggested earlier is clearly evident. The same comparison appears for the truncated normal/heteroscedasticity models. (The estimated cost efficiency results for the basic Pitt and Lee model and the expanded one are the same to three or four digits.) The partial listing below shows the estimates for the four models, noting the time invariance of the Pitt and Lee estimates.

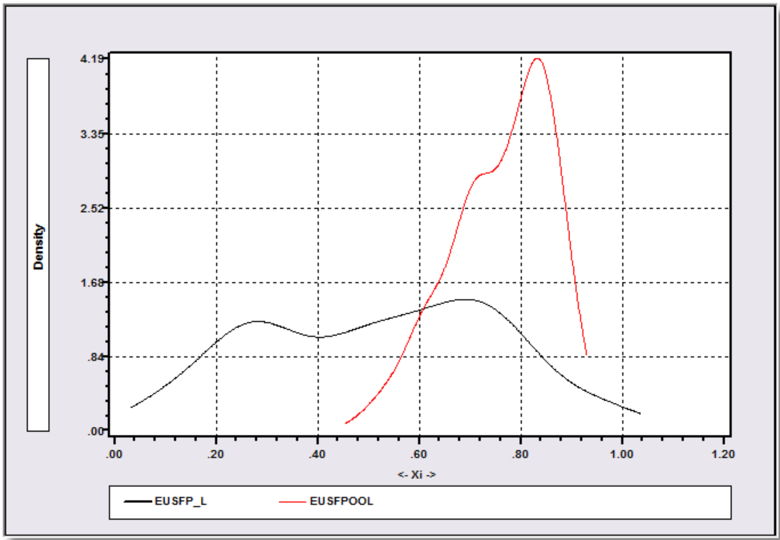


Figure E64.2 Kernel Estimators for Cost Efficiency

Data Editor

56/900 Vars; 33333 Rows: 604 Obs Cell: 0.57818

	EUSFPOOL	EUSFP_L	EUSHET_T	FULLMODL
1 »	0.673518	0.913409	0.820174	0.913804
2 »	0.841946	0.913409	0.877011	0.913804
3 »	0.825565	0.913409	0.795131	0.913804
4 »	0.834643	0.913409	0.855689	0.913804
5 »	0.837169	0.913409	0.867229	0.913804
6 »	0.829983	0.913409	0.852395	0.913804
7 »	0.811368	0.913409	0.770535	0.913804
8 »	0.74011	0.626534	0.675233	0.6258
9 »	0.770612	0.626534	0.711839	0.6258
10 »	0.775549	0.626534	0.720038	0.6258
11 »	0.779228	0.626534	0.725343	0.6258
12 »	0.773121	0.626534	0.718287	0.6258
13 »	0.793122	0.626534	0.746099	0.6258
14 »	0.789952	0.626534	0.740968	0.6258
15 »	0.782502	0.626534	0.73035	0.6258
16 »	0.817268	0.626534	0.785789	0.6258
17 »	0.820948	0.626534	0.791773	0.6258
18 »	0.810805	0.626534	0.771794	0.6258
19 »	0.820409	0.626534	0.790504	0.6258
20 »	0.815996	0.626534	0.785679	0.6258
21 »	[illegible]	[illegible]	[illegible]	[illegible]

Figure E64.3 Estimated Cost Efficiency

E64.3.3 Technical Details

For the three forms of the normal mixture models, we use the following: Let

$\gamma = \sigma_u^2 / \sigma_v^2$

$\tau_i = \mu_i/\sigma_u$

$\mu_i = \boldsymbol{\theta}'\mathbf{z}_i$ for the heterogeneous mean model

$\mu,$ = a constant (0) for the simple truncated (half) normal model

$A_i = 1 + \gamma T_i$

$h_i = \tau_i / A_i - S\gamma T_i \bar{\varepsilon}_i /(\sigma_u A_i)$

$\bar{\varepsilon}_i = (1/T_i)\sum_{t=1}^{T_i}(y_{it} - \boldsymbol{\beta}' x_{it})$.

Then, the contribution of individual i to the log likelihood function for the normal-half normal model is

$$\log L_i = -(Ti/2)\log 2\pi - T_i \log\sigma_u - ½ \log A_i - (T_i/2) \log \gamma$$
$$- ½(\gamma / \sigma_u^2)\sum_{t=1}^{T_i} \varepsilon_{it}^2 + ½ A_i h_i^2 + ½ \log\Phi\left(h_i\sqrt{A_i}\right) - ½ \tau_i^2 - \log\Phi(\tau_i)$$

For the normal-exponential model, let

$h_i = -(\theta\sigma_v/T_i + \mathrm{d}\,\bar{\varepsilon}_i /\sigma_v)$

Then, $\log L_i = -½ \log T_i - (T_i - 1)\log 2\pi + \log\theta - (T_i - 1)\log\sigma_v$

$$- ½(1/\sigma_v^2) \sum_{t=1}^{T_i} \varepsilon_{it}^2 + ½ T_i h_i^2 + \log\Phi\left(h_i\sqrt{T_i}\right)$$

The Jondrow estimator, as formulated in Battese and Coelli (1988) in as follows: Let

$\gamma_i = 1 / (1 + \lambda^2 T_i)$,

$\psi_i^2 = \sigma_u^2\gamma_i$,

$E_i = \gamma_i\mu + (1 - \gamma_i)(-\bar{\varepsilon}_i)$,

and $\bar{\varepsilon}_i = (1/\mathrm{T_i})\Sigma_t\varepsilon_{it}$.

Then, $E[u_i|\varepsilon_{i1},\varepsilon_{i2},...] = E_i + \psi_i[\phi(E_i/\psi_i) / \Phi(E_i/\psi_i)]$.

For the exponential model, replace ψ_i with σ_v and E_i with $\sqrt{T_i}\,(-\bar{\varepsilon}_i - \theta\sigma_v^2/T_i)$.

E64.4 Cornwell, Schmidt and Sickles – Time Invariant Inefficiency, Fixed Effects

Cornwell, Schmidt and Sickles (1990) suggested a modification of the familiar fixed effects linear regression,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it}.$$

The estimated model is

$$\begin{aligned} y_{it} &= a_i + \mathbf{b}'\mathbf{x}_{it} + v_{it} \\ &= \max(a_i) + \mathbf{b}'\mathbf{x}_{it} + v_{it} + [a_i - \max(a_i)] \\ &= a + \mathbf{b}'\mathbf{x}_{it} + v_{it} - u_i \end{aligned}$$

where $$u_i = \max(a_i) - a_i > 0.$$

(To change this to a cost frontier, change u_i to $[a_i - \min(a_i)]$ This bears resemblance to a stochastic frontier model, though in fact, it is a 'deterministic' frontier model. The signature feature is that u_i equals zero for the 'most efficient' firm in the sample. A natural interpretation of this is that what we measure with the model is not the absolute inefficiency, but inefficiency of firm *i* relative to the other firms in the sample. From the modeler's point of view, this approach has several substantive advantages and disadvantages: The main advantage is

- It is distribution free. It requires only the assumptions of the linear model.

The disadvantages are:

- It does not allow any time invariant variables in the model.
- It labels as inefficiency any and all omitted time invariant effects.
- It can only measure firms relative to each other.

As illustrated in the results below, this approach tends to produce very large estimates of u_i. The invariance assumption about u_i has been criticized elsewhere. Attempts to relax this assumption are a recurrent theme in the literature, including the Battese and Coelli and true fixed and random effects approaches described later. Other early work on the model suggested direct manipulation of the fixed effects, for example,

$$\alpha_{it} = \theta_{i0} + \theta_{i1}t + \theta_{i2}t^2.$$

Other more recent research (Han, Orea and Schmidt (2005)) has proposed factor analytic forms for α_{it}. The sections to follow will include several of these different approaches.

Application

This Cornwell, Schmidt and Sickles (CSS) approach requires only a linear fixed effects regression and a few instructions to manipulate the fixed effects. The following analyzes the airline data with this approach. The following computes the CSS estimates and compares them to the unstructured pooled estimates (using the normal-half normal model from Chapter E62) and the Pitt and Lee model introduced above. The commands for the analysis are as follows:

```
SAMPLE     ; All $
CREATE     ; Railroad = id $
CREATE     ; If(railroad > 20)railroad = railroad - 1 $ (There is a gap in the data)
HISTOGRAM  ; Rhs = railroad
           ; Title = Number of Observations for Firms in Swiss Railroad Sample $
SETPANEL   ; Group = id ; Pds = ti $
REJECT     ; ti = 1 $
FRONTIER   ; Lhs = lnc ; Cost ; Rhs = x ; Costeff = eusfpool $
CREATE     ; pooled = Group Mean(eusfpool, Pds = ti) $
FRONTIER   ; Lhs = lnc ; Cost ; Rhs = x ; Panel ; Costeff = pittlee $
REGRESS    ; Lhs = lnc ; Rhs = x ; Panel ; Fixed Effects $
CREATE     ; ai = alphafe(railroad) $
CALC       ; minai = Min(ai) $
CREATE     ; css = Exp((minai - ai)) $
CREATE     ; Period = Ndx(id,1) $
REJECT     ; period#1 $
PLOT       ; Lhs = railroad ; Rhs = pooled,css ; Grid ; Fill ; Limits = 0,1
           ; Vaxis = Estimated Cost Efficiency
           ; Title = Half Normal vs. Cornwell, Schmidt, Sickles FE Cost Efficiencies $
PLOT       ; Lhs = railroad ; Rhs = css,pittlee ; Grid ; Fill ; Limits = 0,1
           ; Vaxis = Estimated Cost Efficiency
           ; Title = Pitt and Lee RE vs. Cornwell, Schmidt, Sickles FE Cost Efficiencies $
```

The results below show the considerable differences in the parameter estimates produced by the three models. Figure E64.4 demonstrates the expected quite large differences between the time varying estimates (using the group means) and the time invariant results based on the CSS model. Figure E64.5 also shows a striking, albeit commonly observed result – the CSS and Pitt and Lee estimates are virtually identical.

```
----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=LNC      Mean                 =      11.30305
             Standard deviation   =       1.09984
             No. of observations  =           604  Degrees of freedom
Regression   Sum of Squares       =       726.000          52
Residual     Sum of Squares       =       3.41179         551
Total        Sum of Squares       =       729.412         603
             Standard error of e  =        .07869
Fit          R-squared            =        .99532  R-bar squared =   .99488
Model test   F[ 52,    551]       =    2254.77325  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =     706.21504  Akaike I.C.   = -5.00084
             Restricted (b=0)     =    -914.01557  Bayes  I.C.   = -4.61443
             Chi squared [ 52]    =    3240.46122  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)   =       .668792
---------------------------------------------------
Panel:Groups Empty       0,      Valid data       49
             Smallest    3,      Largest          13
             Average group size in panel       12.33
Variances    Effects a(i)          Residuals e(i,t)
              .423441                      .006192
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     LNC|  Coefficient      Error       z     |z|>Z*        Interval
--------+-------------------------------------------------------------------
    LNQ2|     .29374***      .02850    10.31  .0000      .23789    .34959
    LNQ3|     .01612***      .00543     2.97  .0030      .00547    .02676
    LPLE|     .66452***      .03580    18.56  .0000      .59434    .73469
    LPKE|     .31777***      .01863    17.05  .0000      .28125    .35430
--------+-------------------------------------------------------------------
```

(These are the estimated parameters in the estimated pooled stochastic frontier model.)

```
Constant|   -10.0907***     1.14284    -8.83  .0000    -12.3306   -7.8507
    LNQ2|     .64179***      .01371    46.80  .0000      .61491    .66867
    LNQ3|     .06855***      .00655    10.46  .0000      .05570    .08139
    LPLE|     .53971***      .08858     6.09  .0000      .36610    .71333
    LPKE|     .26045***      .03260     7.99  .0000      .19655    .32435
        |Variance parameters for compound error
  Lambda|    1.29697***      .13854     9.36  .0000     1.02545   1.56850
   Sigma|     .44345***      .00056   789.05  .0000      .44235    .44455
```

(These are the estimated parameters in the estimated Pitt and Lee model.)

```
        |Deterministic Component of Stochastic Frontier Model
Constant|   -7.25643***      .24767   -29.30  .0000    -7.74185  -6.77101
    LNQ2|     .36259***      .01503    24.12  .0000      .33312    .39205
    LNQ3|     .01902***      .00240     7.94  .0000      .01432    .02372
    LPLE|     .64148***      .02112    30.38  .0000      .60009    .68287
    LPKE|     .30842***      .00700    44.08  .0000      .29471    .32214
        |Variance parameters for compound error
  Lambda|    12.1932**      5.55909     2.19  .0283      1.2975   23.0888
Sigma(u)|     .96071***      .13303     7.22  .0000      .69998   1.22145
```

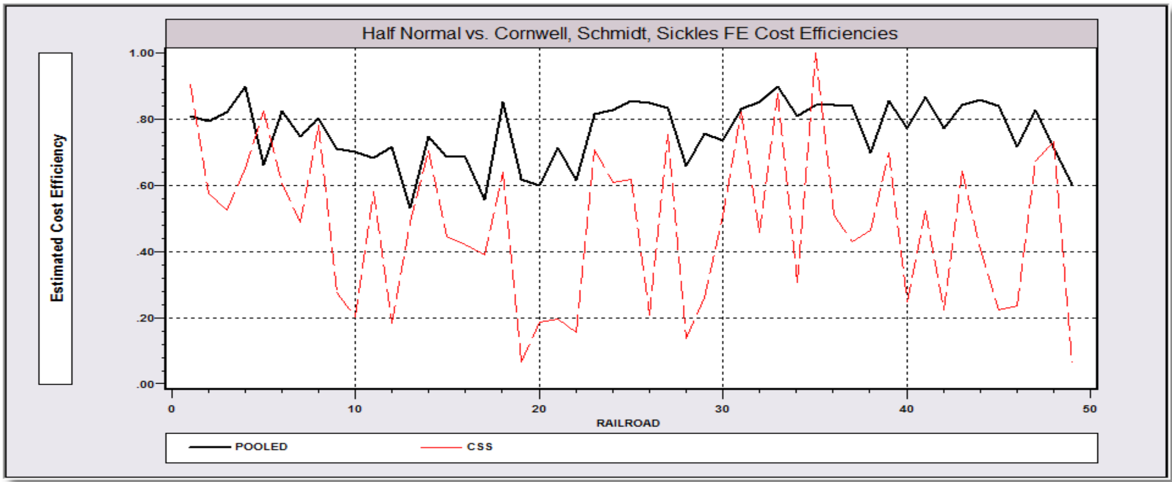


Figure E64.4 Cornwell et al. Estimates vs. Normal-Half Normal

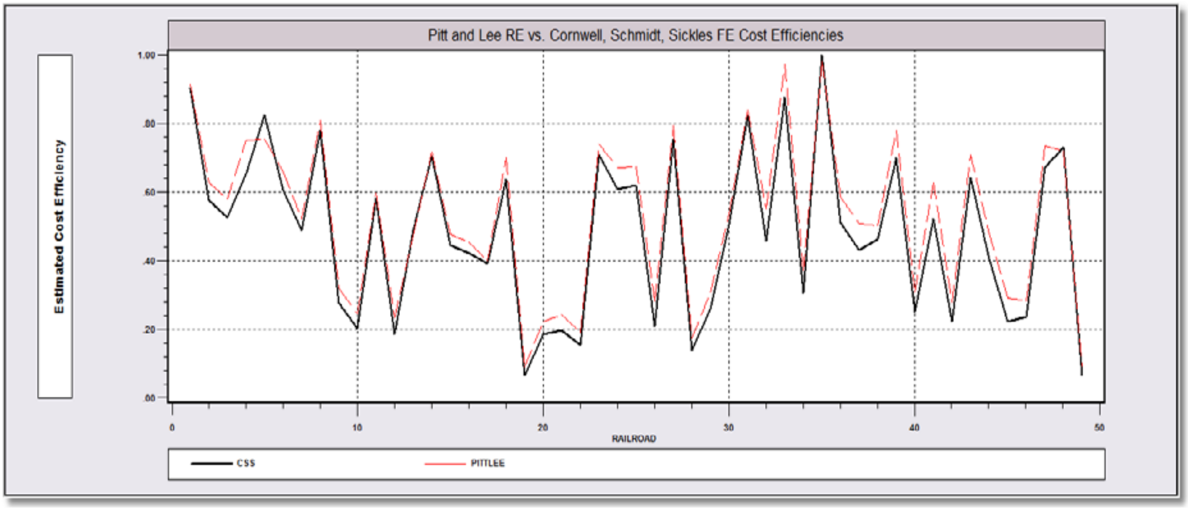


Figure E64.5 Estimated Inefficiencies from Cornwell et al. and Pitt and Lee Models

E64.5 Battese and Coelli – Time Dependent Inefficiency Models

Battese and Coelli (1992) proposed a series of models that can be collected in the general form

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}$$

$$u_{it} = g(\mathbf{z}_{it})\,|U_i| \text{ where } U_i \text{ is half normal or truncated normal.}$$

Several formulations are available. In Battese and Coelli's original formulation, the distribution was half normal and the base specification was

$$g(z_{it}) = \exp[-\eta(t - T)]$$

where T is the number of periods in their balanced panel. (Here it would be T_i.) They also suggested

$$g(z_{it}) = \exp[-\eta_1(t - T) + -\eta_2(t - T)^2].$$

The first (linear) form is taken to be the default case for this model. The second is not provided in this package. The BC92 model is requested with

```
FRONTIER    ; Lhs = ... ; Rhs = one,...
            ; Model = BC
            ; Panel $
```

A truncated normal version is requested by adding

```
            ; Rh2 = list of variables which may (generally should) include one
```

(The **; Model = T** is not needed here.)

We note a warning to practitioners. When the data are very consistent with the model, the Battese and Coelli model produces quite satisfactory results. The framework has been employed in many recent empirical applications. But, when the data are not of particularly good quality, or this is the wrong model, extreme results can emerge. The airline data examined in Chapter E63 (and the WHO data), for example, are a poor fit to this model.

We have labeled this model as 'time dependent' rather than time varying. While the inefficiency component in the model does vary through time, the variation is systematic with respect to time. A question pursued in the ongoing literature is the extent to which this model actually moves away from the time invariant specification of Pitt and Lee. Since there is actual variation, the result is clearly somewhere between Pitt and Lee and what we have labeled the unstructured 'pooled' model. If η equals zero, Pitt and Lee emerges, so it depends entirely on this parameter. We have found in some investigations that the end result is actually closer to Pitt and Lee than it is to the pooled model – that is, there is quite a lot of structure involved in the BC92 model. The example below illustrates.

E64.5.1 Application

To illustrate the Battese and Coelli models, we return to the railroad data used previously. The base case is the pooled data stochastic cost frontier. This is followed by the Pitt and Lee model and, finally, by the original Battese Coelli 'time decay' model,

$$g(\mathbf{z}_{it}) = \exp[-\eta(t - T_i)].$$

The commands are

```
SAMPLE      ; All $
REJECT      ; ti = 1 $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; Costeff = eusfpool $
FRONTIER    ; Lhs = lnc ; Cost ; Rhs = x ; Model = BC ; Panel ; Costeff = eucbc92 $
DSTAT       ; Rhs = eucbc92,eusfpool $
KERNEL      ; Rhs = eucbc92,eusfpool
            ; Title = Estimated Cost Efficiencies - Battese-Coelli 1992 vs. Pooled $
KERNEL      ; Rhs = eucbc92,pittlee
            ; Title = Estimated Cost Efficiencies - Battese-Coelli 1992 vs. Pitt and Lee $
```

The kernel density estimators are used to compare the efficiency estimates from the pooled data model to the Battese and Coelli model. The estimates of $\exp(-E[u_{it}|\varepsilon_i])$ from the Battese and Coelli model are far larger than those from the pooled model. The assumption of time invariance of the random term is a major component of this model. The second kernel estimator below compares Battese-Coelli to Pitt-Lee. The correspondence of the two results is striking, albeit to be expected given the small estimated value of η.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LNC
Log likelihood function      -209.42340
Estimation based on N =    604, K =   7
Inf.Cr.AIC  =    432.8 AIC/N =     .717
Variances: Sigma-squared(v)=     .07332
           Sigma-squared(u)=     .12333
           Sigma(v)        =     .27077
           Sigma(u)        =     .35119
Sigma = Sqr[(s^2(u)+s^2(v)]=     .44345
Gamma = sigma(u)^2/sigma^2 =     .62716
Var[u]/{Var[u]+Var[v]}     =     .37937
Stochastic Cost Frontier Model, e = v+u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0         -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.060
Kodde-Palm C*: 95%: 2.706,   99%:  5.412
```

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -10.0907***     1.14284    -8.83  .0000    -12.3306   -7.8507
    LNQ2|      .64179***     .01371    46.80  .0000      .61491    .66867
    LNQ3|      .06855***     .00655    10.46  .0000      .05570    .08139
    LPLE|      .53971***     .08858     6.09  .0000      .36610    .71333
    LPKE|      .26045***     .03260     7.99  .0000      .19655    .32435
        |Variance parameters for compound error
  Lambda|     1.29697***     .13854     9.36  .0000     1.02545   1.56850
   Sigma|      .44345***     .00056   789.05  .0000      .44235    .44455
--------+-------------------------------------------------------------

----------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                  LNC
Log likelihood function       530.16177
Estimation based on N =    604, K =   8
Inf.Cr.AIC  =  -1044.3 AIC/N =   -1.729
Stochastic frontier based on panel data
Estimation based on      49 individuals
Variances: Sigma-squared(v)=     .00613
           Sigma-squared(u)=     .97581
           Sigma(v)        =     .07828
           Sigma(u)        =     .98783
Sigma = Sqr[(s^2(u)+s^2(v)]=     .99093
Gamma = sigma(u)^2/sigma^2 =     .99376
Var[u]/{Var[u]+Var[v]}     =     .98301
Stochastic Cost Frontier Model, e = v+u
Battese-Coelli Models: Time Varying uit
Time dependent uit=exp[-eta(t-T)]*|U(i)|
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0         -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] = 1481.231
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -6.83502***      .27362   -24.98  .0000    -7.37130  -6.29873
    LNQ2|      .35459***     .01636    21.68  .0000      .32254    .38665
    LNQ3|      .02183***     .00238     9.17  .0000      .01716    .02649
    LPLE|      .61516***     .02092    29.40  .0000      .57415    .65617
    LPKE|      .30931***     .00701    44.09  .0000      .29556    .32306
        |Variance parameters for compound error
  Lambda|    12.6195***      .01188  1062.18  .0000     12.5962   12.6428
Sigma(u)|      .98783***     .15275     6.47  .0000      .68845   1.28721
        |Eta parameter for time varying inefficiency
     Eta|    -.00248***      .00086    -2.89  .0039     -.00416   -.00080
--------+-------------------------------------------------------------
```

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
EUCBC92	.514566	.231680	.085140	.982112	604	0
EUSFPOOL	.760991	.095229	.478178	.906348	604	0

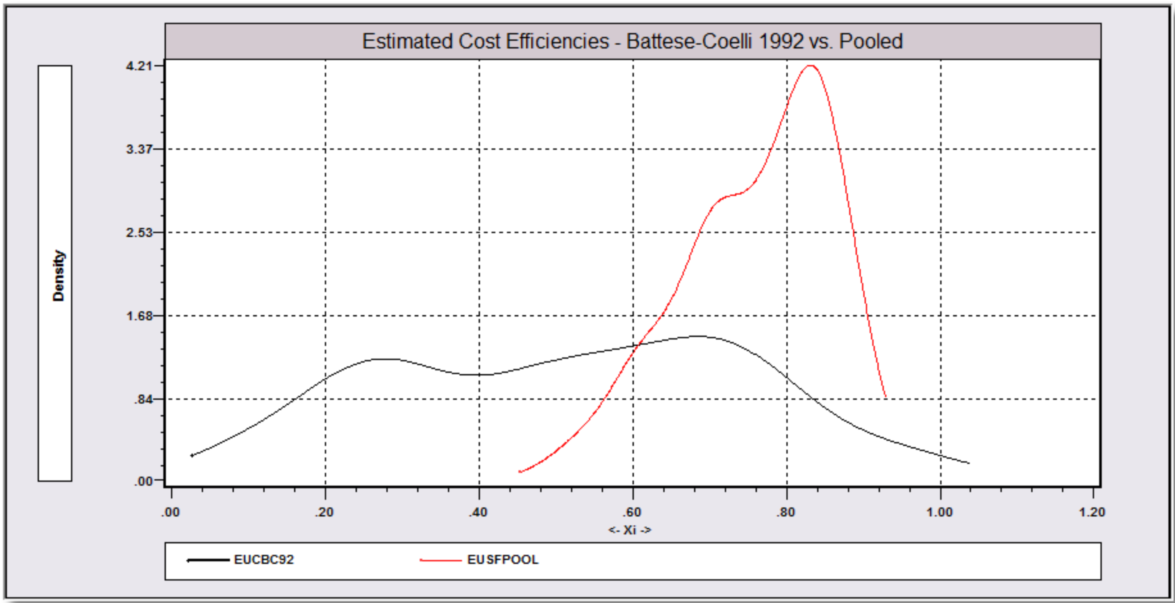


Figure E64.6 Kernel Density Estimates for Inefficiencies from Battese and Coelli Model

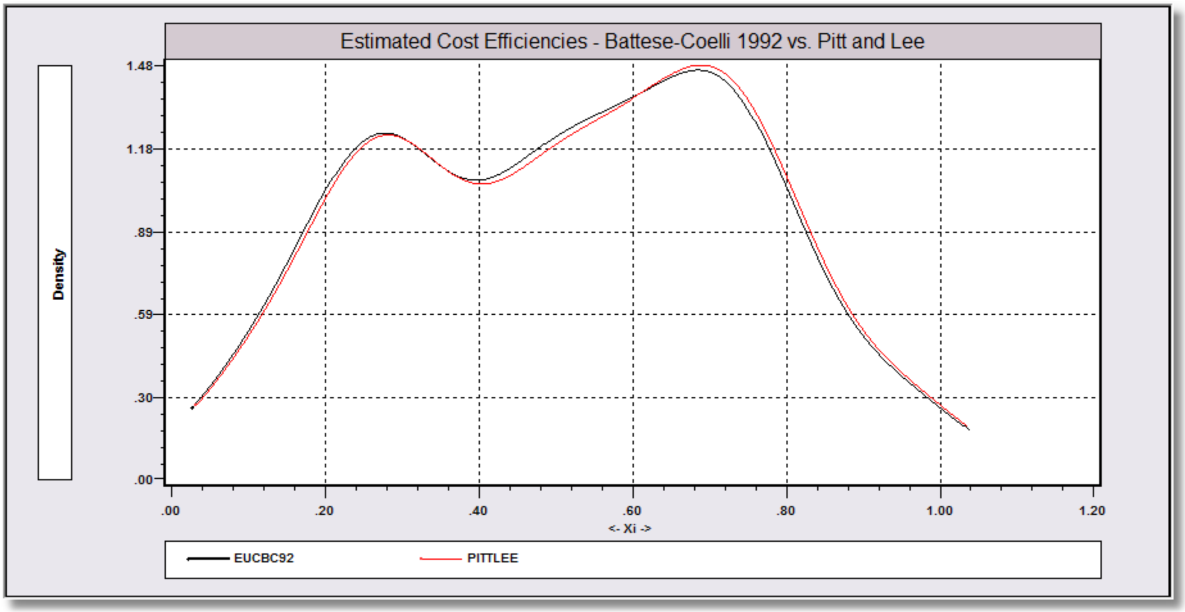


Figure E64.7 Kernel Density Estimates for Inefficiencies

E64.5.2 Technical Details

To form the log likelihood function for the model, we use Battese and Coelli's parameterization of the model. The contribution of the *i*th individual (firm, group, etc.) to the log likelihood is

$$\begin{aligned}\log L_i = &-\frac{T_i}{2}(\log 2\pi + \log\sigma^2) - \frac{(T_i-1)\log(1-\gamma)}{2} - \frac{1}{2}\sum_{t=1}^{T_i}\frac{\varepsilon_{it}^2}{(1-\gamma)\sigma^2}\\ &-\frac{1}{2}\log\left[1+\gamma\left(\left(\sum_{t=1}^{T_i} g_{it}^2\right)-1\right)\right]\\ &-\frac{1}{2}\left(\frac{\mu_i}{\sigma\sqrt{\gamma}}\right)^2 - \log\Phi\left(\frac{\mu_i}{\sigma\sqrt{\gamma}}\right) + \frac{A_i^2}{2} + \log\Phi(A_i)\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \sigma_u^2 + \sigma_v^2\\ \gamma &= \sigma_u^2/\sigma^2\\ \varepsilon_{it} &= y_{it} - \boldsymbol{\beta}'\mathbf{x}_{it}\\ \mu_i &= 0 \text{ or } \mu \text{ or } \boldsymbol{\delta}'\mathbf{w}_i\\ g_{it} &= \exp[-\eta(t-T_i)] \text{ or } \exp(\boldsymbol{\eta}'\mathbf{z}_{it})\\ S &= +1 \text{ for a production model and -1 for a cost model}\\ A_i &= \frac{(1-\gamma)\mu_i - \gamma S\Sigma_{t=1}^{T_i} g_{it}\varepsilon_{it}}{\sqrt{\gamma(1-\gamma)\left[1+\gamma\left(\left(\Sigma_{t=1}^{T_i} g_{it}^2\right)-1\right)\right]}}\end{aligned}$$

Derivatives of this function are complicated in the extreme, and are omitted here. (Some useful results for obtaining them are found in Battese and Coelli (1992, 1995).)

The Jondrow estimator of u_{it} is

$$\begin{aligned}E[u_{it} \mid \varepsilon_{i1},\varepsilon_{i2},...] &= g_{it}\, E[u_i \mid \varepsilon_{i1},\varepsilon_{i2},...]\\ &= g_{it}\left[\tilde{\mu}_i + \tilde{\sigma}_i\left(\frac{\phi(\tilde{\mu}_i/\tilde{\sigma}_i)}{\Phi(\tilde{\mu}_i/\tilde{\sigma}_i)}\right)\right]\end{aligned}$$

where
$$\tilde{\mu}_i = \frac{(1-\gamma)\mu_i - \gamma\Sigma_{t=1}^{T_i} g_{it}(S\varepsilon_{it})}{(1-\gamma)+\gamma\Sigma_{t=1}^{T_i} g_{it}^2}$$

$$\tilde{\sigma}_i^2 = \frac{\gamma(1-\gamma)\sigma^2}{(1-\gamma)+\gamma\Sigma_{t=1}^{T_i} g_{it}^2}$$

E64.6 Time Varying Inefficiency in the Battese Coelli Model

The general form of the Battese and Coelli model is,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}$$

$$u_{it} = g(\mathbf{z}_{it})\,|U_i| \text{ where } U_i \text{ is half normal or truncated normal.}$$

The default form used earlier is $g(\mathbf{z}_{it}) = \exp[-\eta(t - T_i)]$. You may also use a more general form,

$$g(\mathbf{z}_{it}) = \exp(\boldsymbol{\eta}'\mathbf{z}_{it})$$

where $\mathbf{z}_{it}$ contains any desired set of variables. For this extension, use

```
FRONTIER     ; Lhs = ... ; Rhs = one,...
             ; Model = BC ; Hfu = the variables in z
             ; Pds = the panel specification $
```

As before, the truncated normal version of the model is also supported. For an example, we have used

```
FRONTIER     ; Lhs = lnc ; Cost ; Rhs = x ; Model = BC ; Panel ; Costeff = eucbc92h
             ; Hfu = rack,virage,tunnel $
```

The estimates of cost efficiency produced by this model are identical to those from the base model in the previous section.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LNC
Log likelihood function         529.63533
Stochastic frontier based on panel data
Estimation based on       49 individuals
Variances: Sigma-squared(v)=      .00615
           Sigma-squared(u)=      .94808
           Sigma(v)        =      .07840
           Sigma(u)        =      .97369
Sigma = Sqr[(s^2(u)+s^2(v)]=      .97685
Gamma = sigma(u)^2/sigma^2 =      .99356
Var[u]/{Var[u]+Var[v]}     =      .98247
Stochastic Cost Frontier Model, e = v+u
Battese-Coelli Models: Time Varying uit
Time varying uit=exp[eta*z(i,t)]*|U(i)|
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):     1
Deg. freedom for heteroscedasticity:  3
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  4
LogL when sigma(u)=0          -210.45352
Chi-sq=2*[LogL(SF)-LogL(LS)] = 1480.178
Kodde-Palm C*: 95%: 8.761,  99%: 12.483
```

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     LNC|  Coefficient       Error       z     |z|>Z*        Interval
--------+-------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -6.89845***      .32923   -20.95  .0000    -7.54374  -6.25316
    LNQ2|     .35751***      .01591    22.47  .0000      .32632    .38870
    LNQ3|     .02149***      .00236     9.10  .0000      .01686    .02613
    LPLE|     .61741***      .02430    25.40  .0000      .56977    .66504
    LPKE|     .30892***      .00759    40.71  .0000      .29405    .32380
        |Variance parameters for compound error
  Lambda|    12.4202***      .01108  1120.76  .0000     12.3984   12.4419
Sigma(u)|     .97369***      .13513     7.21  .0000      .70884   1.23855
        |Coefficients in u(i,t)=[exp{eta*z(i,t)}]*|U(i)|
    RACK|     .00024         .01743      .01  .9889     -.03392    .03441
  VIRAGE|    -.02096         .01321    -1.59  .1126     -.04685    .00493
  TUNNEL|     .00219         .01625      .14  .8926     -.02966    .03405
--------+-------------------------------------------------------------
```

(Parameter estimates from base case Battese and Coelli)

```
--------+-------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -6.83502***      .27362   -24.98  .0000    -7.37130  -6.29873
    LNQ2|     .35459***      .01636    21.68  .0000      .32254    .38665
    LNQ3|     .02183***      .00238     9.17  .0000      .01716    .02649
    LPLE|     .61516***      .02092    29.40  .0000      .57415    .65617
    LPKE|     .30931***      .00701    44.09  .0000      .29556    .32306
        |Variance parameters for compound error
  Lambda|    12.6195***      .01188  1062.18  .0000     12.5962   12.6428
Sigma(u)|     .98783***      .15275     6.47  .0000      .68845   1.28721
        |Eta parameter for time varying inefficiency
     Eta|    -.00248***      .00086    -2.89  .0039     -.00416   -.00080
--------+-------------------------------------------------------------
```

E64.7 True Fixed Effects Models

The received applications of fixed effects to the stochastic frontier model, primarily Cornwell, Schmidt and Sickles have actually been reinterpretations of the linear regression model with fixed effects, not frontier models of the sort considered here. The estimators described below apply the fixed effects to the stochastic frontier. We label these 'true fixed effects models' to distinguish them from the linear regression models as discussed in Section E64.3. (This is not meant to apply that these are 'false fixed effects models.' Had we used 'real fixed effects models,' then the contrasting 'unreal fixed effects models' would arise which is likewise problematic. We use this purely as a concise term of art, not a characterization of the types of estimators considered.)

The stochastic frontier model with fixed effects may be fit in several forms. The base case applies the heterogeneity to the normal-half normal production function model;

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_{it},$$

where S = +1 for a production frontier and -1 for a cost frontier, and

$$u_i = |\,N[0, \sigma_u^2]\,|.$$

This model (as are the others) is fit by maximum likelihood, not least squares. The normal-half normal model is applied to the stochastic part of the model. Note that the inefficiency term in this model is time varying. The heterogeneity may appear in Stevenson's truncated normal model as follows. This is a true fixed effects, normal-truncated normal model.

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$u_i = |\,N[\mu_i, \sigma_u^2]\,|$$

$$\mu_i = \boldsymbol{\delta}'\mathbf{z}_i.$$

In this form, the heterogeneity is still retained in the production function part of the model. Another possibility is to allow the heterogeneity to enter the mean of the inefficiency distribution rather than the production function – this seems the most natural of the three forms. In this case,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$u_{it} = |\,N[\mu_{it}, \sigma_u^2]\,|$$

$$\mu_{it} = \alpha_i + \mu \text{ (nonzero) or } \boldsymbol{\delta}'\mathbf{z}_i.$$

The mean of the inefficiency distribution shifts in time, but also has a firm specific component. Finally, the heterogeneity may be shifted to the variance of the inefficiency distribution. In this form, we have

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$u_{it} = |\,N[0, \sigma_{ui}^2]\,|$$

$$\sigma_{uit}^2 = \sigma_u^2 \times \exp(\alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it}).$$

The variables in the variance term may be omitted if only a groupwise heteroscedastic model is desired. Note this is a half normal model. A model with nonzero underlying mean and variation in the variance appears to be inestimable. Note that in order to secure identification, this model must have time varying inefficiency, induced by time variation in the variance.

NOTE: We have had extremely limited success with the second and third forms of the model. The likelihood function is quite volatile in the parameters of the underlying mean of the truncated distribution with the result that the estimated variance parameters λ and σ generally become negative in the early iterations and estimation must be halted. This occurs even when very good starting values are used, which suggests that estimation of this model as stated is likely to be extremely problematic in all but the most favorable of cases. An alternative approach which is simple, but can be used only with small panels (up to 100 groups), is suggested below.

In terms of implementation, we note that these forms of the models, though they are new with *LIMDEP*, have long been feasible. The panels typically used by researchers in this setting are often fairly small – our airline data for example have only 25 units and the Swiss railroad data has 49 firms. It would always have been possible to create these models simply by adding dummy variables to the familiar model. However, *LIMDEP*'s implementation of the model obviates this by using the methodology described in Chapter R23. In principle, this allows up to 100,000 firms in the data set.

Results that are kept for this model are

Matrices: *b* = estimate of **β**
varb = asymptotic covariance matrix for estimate of **β**.
alphafe = estimated fixed effects (if **; Par** is in the command)

Scalars: *kreg* = number of variables in Rhs
nreg = number of observations
logl = log likelihood function

Last Model: *b_variables*

The upper limit on the number of groups is 100,000.

E64.7.1 Commands for the Fixed Effects Stochastic Frontier Model

The command for fitting the normal-half normal model with fixed effects is as follows:

FRONTIER ; Lhs = ... ; Rhs = one,... $
FRONTIER ; Lhs = ... ; Rhs = one,...
; FEM ; Pds = specification $

The model must be fit twice. The first model is a pooled data model which provides the starting values for the second. The second command is identical to the first save for the addition of the panel data specification. In order to set up the initial values correctly, it is essential that your initial model include the constant term first in the Rhs list and that the second model specification be identical to the first. Other options and specifications for the fixed effects models are the same as in other applications. (See Chapter R23 for details.) The fixed effects command also contains the constant term, but this will be removed by the command processor later. See the example below for the operation of the command.

NOTE: Starting values must be provided by the first estimator. The specification **; Start = list of values** is not available for this model. You must fit both models each time you fit an FEM. The starting values are not retained after the FEM is estimated.

All fixed effects forms are estimated by maximum likelihood. You may also fit a two way fixed effects model

$$y_{it} = \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_i, \text{ (change to } v + u \text{ for a stochastic cost frontier)},$$

$$u_i = |\,N[0, \sigma_u^2]\,|$$

where γ_t is an additional, time (period) specific effect. The time specific effect is requested by adding

; Time

to the command if the panel is balanced, and

; Time = variable name

if the panel is unbalanced.

For the unbalanced panel, we assume that overall, the sample observation period is t = 1,2,..., T_{max} and that the time variable gives for the specific group, the particular values of t that apply to the observations. Thus, suppose your overall sample is five periods. The first group is three observations, periods 1, 2, 4, while the second group is four observations, 2, 3, 4, 5. Then, your panel specification would be

; Pds = Ti, for example, where *Ti* = 3, 3, 3, 4, 4, 4, 4
and **; Time = Pd**, for example, where *Pd* = 1, 2, 4, 2, 3, 4, 5.

E64.7.2 Model Specifications for Fixed Effects Stochastic Frontier Models

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps ancillary parameter σ in main results vector *b*.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list gives starting values for a nonlinear model.
; Tlg[= value] sets convergence value for gradient.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.

E64.7.3 Application of the True Fixed Effects Model

We have fit the fixed effects model with the airline data used in the previous chapter. These are simple models that do not use the observed heterogeneity in load factor, stage length or number of points served. Additional variables which vary over time can also be included in the function. The commands employed for the example are

```
SETPANEL    ; Group = firm ; Pds = ti $
FRONTIER    ; Lhs = lq ; Rhs = one,lf,lm,le,ll,lp,lk$
FRONTIER    ; Lhs = lq ; Rhs = one,lf,lm,le,ll,lp,lk,
            ; FEM ; Panel ; Techeff = euitfe ; Par $
REGRESS     ; Lhs = lq ; Rhs = one,lf,lm,le,ll,lp,lk
            ; Panel ; Fixed Effects $
CREATE      ; ai = alphafe(firm) $
CALC        ; maxai = Max(ai) $
CREATE      ; euicss = exp(-(maxai - ai)) $
CREATE      ; meuitfe = Group Mean(euitfe, Pds = ti) $
SAMPLE      ; All $
CREATE      ; Period = Ndx(firm,1) $
PLOT        ; For[period=1] ; Lhs = firm ; Rhs = euitfe,euicss
            ; Fill ; Symbols ; Limits = 0,1 ; Grid
            ; Title = Technical Efficiency Estimates, CSS vs. True Fixed Effects
                      (Group Means)
            ; Vaxis = Estimated Technical Efficiency $
```

This command recovers the estimated fixed effects from the Cornwell et al. model. then replicates them for each year in the data set. This is used to create the plot of the two sets of estimates of u_i shown below.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LQ
Log likelihood function         108.43918
Estimation based on N =    256, K =    9
Inf.Cr.AIC  =   -198.9 AIC/N =    -.777
Model estimated: Aug 17, 2011, 06:36:42
Variances: Sigma-squared(v)=     .01902
           Sigma-squared(u)=     .01692
           Sigma(v)        =     .13791
           Sigma(u)        =     .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=     .18957
Gamma = sigma(u)^2/sigma^2 =     .47074
Var[u]/{Var[u]+Var[v]}     =     .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0            108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
```

```
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***       .72136    -4.14  .0000    -4.40206  -1.57439
      LF|     .37257***       .07038     5.29  .0000      .23463    .51052
      LM|     .69910***       .07580     9.22  .0000      .55054    .84766
      LE|    2.09473***       .68790     3.05  .0023      .74647   3.44299
      LL|    -.42909***       .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***       .09498     4.69  .0000      .25917    .63149
      LK|   -2.09806***       .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***       .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***       .00064   297.81  .0000      .18832    .19082
--------+-------------------------------------------------------------

Normal exit from iterations. Exit status=0.
-----------------------------------------------------------------------------
FIXED EFFECTS Frontr Model
Dependent variable                   LQ
Log likelihood function       205.05799
Estimation based on N =    256, K =  33
Inf.Cr.AIC  =   -344.1 AIC/N =   -1.344
Model estimated: Aug 17, 2011, 06:36:46
Unbalanced panel has     25 individuals
Skipped    0 groups with inestimable ai
Half normal stochastic frontier
Sigma( u) (1 sided)  =           .11713
Sigma( v) (symmetric)=           .08347
--------+-------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
        |Production / Cost parameters
      LF|     .20090**        .09879     2.03  .0420      .00727    .39453
      LM|     .78173***       .07495    10.43  .0000      .63483    .92863
      LE|     .56626          .62357      .91  .3638     -.65591   1.78843
      LL|    -.16687          .11488    -1.45  .1464     -.39204    .05830
      LP|     .17273*         .09414     1.83  .0665     -.01177    .35724
      LK|    -.29167          .69055     -.42  .6728    -1.64513   1.06179
        |Variance parameter for v +/- u
   Sigma|     .14383***       .00045   317.51  .0000      .14294    .14472
        |Asymmetry parameter, lambda
  Lambda|    1.40326***       .21468     6.54  .0000      .98248   1.82403
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
LSDV          least squares with fixed effects ....
LHS=LQ        Mean                 =     -1.11237
              Standard deviation   =      1.29728
              No. of observations  =          256  Degrees of freedom
Regression    Sum of Squares       =      426.103          30
Residual      Sum of Squares       =      3.04876         225
Total         Sum of Squares       =      429.152         255
              Standard error of e  =       .11640
Fit           R-squared            =       .99290  R-bar squared =   .99195
Model test    F[ 30,    225]       =   1048.21999  Prob F > F*   =   .00000
Diagnostic    Log likelihood       =    203.84835  Akaike I.C.   = -4.18825
              Restricted (b=0)     =   -429.37729  Bayes  I.C.   = -3.75896
              Chi squared [ 30]    =   1266.45126  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)    =      .575211
--------------------------------------------------------
Panel:Groups Empty      0,      Valid data       25
              Smallest   2,      Largest          15
              Average group size in panel      10.24
Variances     Effects a(i)         Residuals e(i,t)
               .030410                    .013550
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error         t   |t|>T*        Interval
--------+-------------------------------------------------------------------
      LF|     .14860        .09677      1.54  .1259     -.04107    .33828
      LM|     .80497***     .07843     10.26  .0000      .65125    .95868
      LE|     .68672        .67075      1.02  .3069     -.62792   2.00136
      LL|    -.15977        .11829     -1.35  .1780     -.39162    .07208
      LP|     .16227        .09973      1.63  .1050     -.03320    .35774
      LK|    -.37897        .74689      -.51  .6123    -1.84284   1.08490
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

Figure E64.8 plots the Jondrow estimates of exp(-$E[u_{it}|\varepsilon_{it}]$) from the true fixed effects model and the estimates of u_i from the Cornwell, Schmidt and Sickles model of Section E64.4 for each firm. Since the true FE estimates vary by period, we have plotted the group means. The implication of the regression based model is clear in the figure. The estimates of technical efficiency from the true FEM are generally considerably larger than those from the deterministic model.

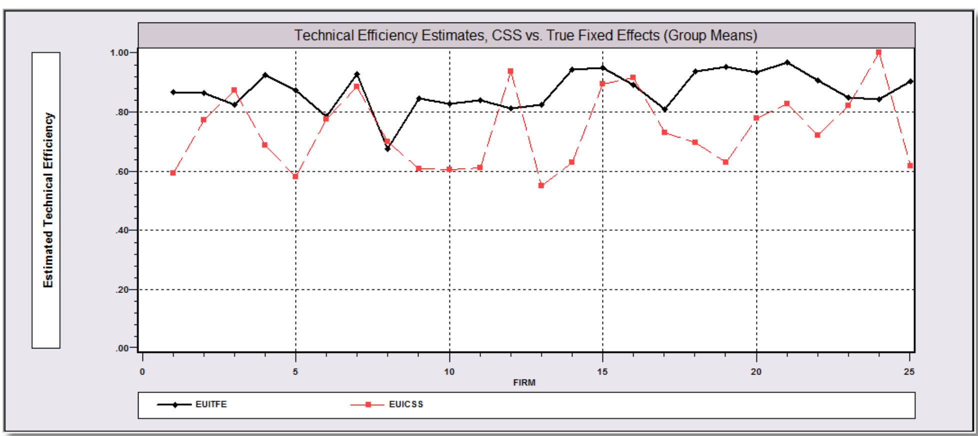


Figure E64.8 True Fixed Effects vs. Fixed Effects Estimates of u_i

E64.7.4 Fixed Effects in the Normal-Truncated Normal Model

The preceding may be extended to the truncated normal (with earlier caveats) as follows: For a model with heterogeneity appearing in the production (or cost) function,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$u_{it} = |\, N[\mu_{it}, \sigma_u^2]\,|$$

$$\mu_{it} = \mu \text{ (nonzero) or } \boldsymbol{\delta}'\mathbf{z}_{it},$$

use **FRONTIER ; Lhs = ... ; Rhs = one, ... ; Rh2 = one, ...**
; Model = T $
FRONTIER ; Lhs = ... ; Rhs = one, ... ; Rh2 = one, ...
; FEM ; Panel $

The Rh2 is optional in the first equation if you have only a constant term in the mean of the truncated distribution. But, you should include it nonetheless so as to insure the match between the first and second commands. Also, it is essential that both Rhs and Rh2 include constant terms in the first positions.

To move the heterogeneity to the mean of the underlying truncated normal distribution,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$u_i = |\, N[\mu_{it}\sigma_u^2]\,|$$

$$\mu_{it} = \alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it},$$

use **FRONTIER ; Lhs = ... ; Rhs = one, ... ; Rh2 = one, ...**
; Model = T $
FRONTIER ; Lhs = ... ; Rhs = one, ... ; Rh2 = one, ...
; Model = T
; FEM ; Panel $

Note that this version differs from the earlier one only in the presence of **; Model = T** in the second form and its absence in the first. Again, the variable specifications in the two commands must be identical, and both must include constant terms in the first position in both lists. As before, you may use **; Rh2 = one** if you do not require variables $\mathbf{z}_{it}$ in the mean. (This constant term will be removed from the fixed effects model, but this common value is used as the starting value for the firm specific estimates.)

We note, we have had scant success with this model even with a carefully constructed data set and good starting values. The problem appears to be Newton's method, which must be used for the general fixed effects program which this is part of. If you have a small panel with no more than 100 groups, an alternative approach appears to work better. You may provide a stratification variable in the cross section template to request that a set of dummy variables be inserted directly into the function.

To fit a model of the first form above, use

```
FRONTIER    ; Lhs = ... ; Rhs = one,...
            ; Model = T [ ; Rh2 = list is optional ]
            ; Str = a variable which provides a group indicator for the panel $
```

The stratification variable must take the full set of values from 1 to N up to 100 and all groups must have at least two observations. For the second form, with the heterogeneity embedded in the mean of the truncated normal distribution, add

```
            ; Mean
```

to the command.

This provides four possible forms of the model, which we illustrate with the airline data:

```
NAMELIST    ; x = one,lf,lm,le,ll,lp,lk $
```

This is a true fixed effects model with normal-truncated normal structure for u_{it}.

```
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Model = T
            ; Str = firm $
```

This model is the same as the preceding one except now $\mu_i = \delta_1 + \delta_2 loadfctr_i$.

```
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Model = T
            ; Rh2 = one,loadfctr
            ; Str = firm $
```

This is a true fixed effects model with the fixed effects appearing in μ_i rather than in the production function.

```
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Model = T
            ; Mean
            ; Str = firm $
```

This model is the same as the preceding model except that *loadfctr* now also appears in the mean of the truncated variable.

```
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Model = T
            ; Rh2 = one,loadfctr ; Mean
            ; Str = firm $
```

E64.7.5 Fixed Effects in the Heteroscedasticity Model

The firmwise heteroscedasticity model,

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},\\
u_{it} &= |\,N[0, \sigma_{uit}^2]\,|\\
\sigma_{uit}^2 &= \sigma_u^2 \times \exp(\alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it})
\end{aligned}
$$

is requested in the same fashion as the normal-truncated normal model, using a stratification variable in the cross section formulation. (This likelihood function is likewise quite ill behaved, though less so than the truncation form.) The command is

```
FRONTIER    ; Lhs = ... ; Rhs = one, ...
            ; Het
            ; Hfu = list of variables ; Hfv = one
            ; Str = stratification variable $
```

This model also allows for the doubly heteroscedastic form,

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it},\\
u_{it} &= |\,N[0, \sigma_{uit}^2]\,|\\
\sigma_{uit}^2 &= \sigma_u^2 \times \exp(\alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it})\\
v_{it} &\sim N[0,\sigma_{vit}^2]\\
\sigma_{vit}^2 &= \sigma_v^2 \times \exp(\boldsymbol{\gamma}'\mathbf{w}_{it})
\end{aligned}
$$

The command would be

```
FRONTIER    ; Lhs = ... ; Rhs = one, ...
            ; Het
            ; Hfu = list of variables ; Hfv = list of variables
            ; Str = stratification variable $
```

To continue the earlier example, the following fits a model of heteroscedasticity to the airline data. The first model has heteroscedasticity and the fixed effects in the variance of u_i. The second is doubly heteroscedastic, again with the fixed effects in the variance of u_i.

```
NAMELIST    ; x = one,lf,lm,le,ll,lp,lk $
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Het ; Hfu = one,loadfctr ; Hfv = one ; Str = firm $
FRONTIER    ; Lhs = lq ; Rhs = x
            ; Het ; Hfu = one,loadfctr ; Hfv = one,loadfctr ; Str = firm $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                      LQ
Log likelihood function       182.50025
Variances: Sigma-squared(v)=     .00876
           Sigma-squared(u)=     .04920
           Sigma(v)        =     .09357
           Sigma(u)        =     .22182
Sigma = Sqr[(s^2(u)+s^2(v)]=     .24075
Gamma = sigma(u)^2/sigma^2 =     .84892
Var[u]/{Var[u]+Var[v]}     =     .67126
Variances averaged over observations
Stochastic Production Frontier, e = v-u
Stratified by FIRM    ,  25 groups
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.70847***      .75902    -4.89  .0000    -5.19612  -2.22081
      LF|     .38142***      .08642     4.41  .0000      .21204    .55079
      LM|     .57659***      .09175     6.28  .0000      .39676    .75642
      LE|    2.78934***      .72692     3.84  .0001     1.36459   4.21408
      LL|    -.41646***      .08641    -4.82  .0000     -.58582   -.24710
      LP|     .59190***      .11704     5.06  .0000      .36251    .82129
      LK|   -2.87861***      .80566    -3.57  .0004    -4.45767  -1.29956
        |Parameters in variance of v (symmetric)
Constant|   -4.73798***      .21921   -21.61  .0000    -5.16764  -4.30833
        |Parameters in variance of u (one sided)
Constant|    8.11346        7.80244     1.04  .2984    -7.17903  23.40596
LOADFCTR|   -23.6678***     6.88328    -3.44  .0006    -37.1588  -10.1768
 FIRM001|    1.35540        7.37739      .18  .8542   -13.10403  15.81482
 FIRM002|     .25791        7.25149      .04  .9716   -13.95476  14.47057
 FIRM003|     .68176        7.22190      .09  .9248   -13.47290  14.83643
(Firms 4-20 omitted)
 FIRM021|     .73089        7.21226      .10  .9193   -13.40488  14.86666
 FIRM022|    -.38963        7.46091     -.05  .9584   -15.01274  14.23347
 FIRM023|    -.63171        7.53984     -.08  .9332   -15.40952  14.14610
 FIRM024|   -7.77451       41.07339     -.19  .8499   -88.27688  72.72786
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       190.29998
Estimation based on N =    256, K =  35
Inf.Cr.AIC  =   -310.6 AIC/N =   -1.213
Model estimated: Aug 22, 2011, 22:57:54
Variances: Sigma-squared(v)=     .00906
           Sigma-squared(u)=     .04124
           Sigma(v)        =     .09519
           Sigma(u)        =     .20307
Sigma = Sqr[(s^2(u)+s^2(v)]=     .22427
Gamma = sigma(u)^2/sigma^2 =     .81986
Var[u]/{Var[u]+Var[v]}     =     .62318
Variances averaged over observations
Stochastic Production Frontier, e = v-u
Stratified by FIRM    ,  25 groups
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.00340***      .65319    -4.60  .0000    -4.28364  -1.72316
      LF|     .24071***      .07721     3.12  .0018      .08938    .39204
      LM|     .60992***      .07600     8.03  .0000      .46096    .75887
      LE|    2.19046***      .62677     3.49  .0005      .96202   3.41890
      LL|    -.38679***      .07314    -5.29  .0000     -.53015   -.24344
      LP|     .49345***      .09820     5.03  .0000      .30098    .68591
      LK|   -2.09638***      .69385    -3.02  .0025    -3.45631   -.73646
        |Parameters in variance of v (symmetric)
Constant|   -13.5487***     2.64897    -5.11  .0000    -18.7406   -8.3569
LOADFCTR|    15.5221***     4.48367     3.46  .0005      6.7343   24.3099
        |Parameters in variance of u (one sided)
Constant|    8.01865        5.60084     1.43  .1522    -2.95879  18.99609
LOADFCTR|   -23.3031***     6.88508    -3.38  .0007    -36.7976   -9.8086
 FIRM001|     .88200        5.06220      .17  .8617    -9.03972  10.80373
 FIRM002|    -.83198        4.67591     -.18  .8588    -9.99660   8.33264
 FIRM003|    -.18608        4.65296     -.04  .9681    -9.30573   8.93356
(Firms 4-20 omitted)
 FIRM021|     .35047        4.63405      .08  .9397    -8.73210   9.43303
 FIRM022|    -.68781        4.83235     -.14  .8868   -10.15903   8.78342
 FIRM023|    -.96206        4.88186     -.20  .8438   -10.53033   8.60622
 FIRM024|   -2.86357        4.82675     -.59  .5530   -12.32383   6.59670
--------+--------------------------------------------------------------------
```

E64.8 True Random Effects Models

We call the stochastic frontier model with a random as opposed to a fixed effect term a 'true random effects' model. The structure is the normal-half normal stochastic frontier model,

$$y_{it} = w_i + \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} + u_{it}$$
$$v_{it} \sim N[0,\sigma_v^2]$$
$$u_{it} = |U_{it}|,\ U_{it} \sim N[0,\sigma_u^2]$$
$$w_i \sim N[0,\sigma_w^2].$$

At first look, this appears to be a model with a three part disturbance, which would surely be inestimable. But, that is incorrect. It is a model with a traditional random effect, but with the additional feature that the time varying disturbance is not normally distributed. Specifically, the model may be written in our familiar form for the stochastic frontier model,

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + w_i$$
$$\varepsilon_{it} \sim (2/\sigma)\phi(\varepsilon_{it}/\sigma)\Phi(-\varepsilon_{it}\lambda/\sigma)$$
$$w_i \sim N[0,\sigma_w^2].$$

The model is estimable by maximum simulated likelihood, as shown below. Contrast this to the Pitt and Lee form,

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} + u_i$$
$$v_{it} \sim N[0,\sigma_v^2]$$
$$u_i = |U_i|,\ U_i \sim N[0,\sigma_u^2].$$

In this form, u_i, the time invariant effect, is the inefficiency. In the true random effects model, u_{it} is the inefficiency, and it is time varying. The latent heterogeneity, the random effect, is w_i. Thus, in the Pitt and Lee model, the 'inefficiency' term also contains all other time invariant unmeasured sources of heterogeneity. In the true random effects model, these effects appear in w_i, and u_{it} picks up the inefficiency. By this interpretation, we will expect (and always find) that estimated inefficiencies from the Pitt and Lee are larger than those from the true random effects model, sometimes far larger. The same result is at work in the difference between the Cornwell et al. fixed effects model and the true fixed effects model. Figure E64.8 clearly shows the effect at work.

The true random effects model is estimated as a form of random parameters (RP) model, in which the only random parameter in the model is the constant term. Thus, we write the model in the canonical RP form

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} + u_{it}$$
$$v_{it} \sim N[0,\sigma_v^2]$$
$$u_{it} = |U_{it}|,\ U_{it} \sim N[0,\sigma_u^2]$$
$$\alpha_i = \alpha + w_i$$
$$w_i \sim N[0,\sigma_w^2]$$

Details on estimating random parameters models appear in Chapter R24, so they will be omitted here.

The command structure for the true random effects model is similar to that for the true fixed effects model. The frontier model must be fit twice, first with no effects to generate the starting values, then with the effect specified. The commands are

FRONTIER ; Lhs = ... ; Rhs = one,... ; Par $
FRONTIER ; Lhs = ... ; Rhs = one,...
; RPM ; Fcn = one(n) $

If desired, the Jondrow estimates are requested as usual with

; Eff = the variable name

The computation of random parameters models is fairly time consuming because of the simulations. You can control this in part with

; Pts = the number of replications

For exploratory work (or for examples in program documentation), small values such as 25 or 50 are sufficient. For final results destined for publication, larger values, in the range of several hundred are advisable. Also, we advise using Halton sequences rather than pseudorandom numbers for the simulations (see Chapter R24). The parameter is

; Halton

The random parameters formulation also allows a variety of specifications for the mean of the underlying u_{it} – the normal-truncated normal model – and for heteroscedasticity. These are discussed in Section E64.9.

Application

To illustrate the true random effects model, we continue the analysis of the airline data. The commands below estimate the pooled model, then the true RE model. In like fashion to the analysis of fixed effects, we then compare the true random effects estimates of inefficiency to the Pitt and Lee estimates. Figure E64.8 illustrates the general result that the estimated inefficiencies in the true fixed effects model will differ considerably from those produced by the Cornwell et al. approach to fixed effects. Figure E64.9 shows the same result for the two approaches to random effects. Numerous studies in the literature (see Greene (2005) for discussion) have documented the similarity of the random and fixed approaches – when the same overall structure is used. Thus, Figure E64.10 shows similar results for the true fixed and random effects models and for the Pitt and Lee and Cornwell et al. models.

The commands used for this application are as follows:

```
NAMELIST    ; x = one,lf,lm,le,ll,lp,lk $
FRONTIER    ; Lhs = lq ; Rhs = x ; Panel ; Eff = uplre $
FRONTIER    ; Lhs = lq ; Rhs = x ; Par $
FRONTIER    ; Lhs = lq ; Rhs = x ; Panel ; RPM ; Eff  = utre
            ; Fcn = one(n) ; Pts = 50 ; Halton $
FRONTIER    ; Lhs = lq ; Rhs = x ; Par $
FRONTIER    ; Lhs = lq ; Rhs = x ; Panel ; FEM ; Eff = utfe $
DSTAT       ; Rhs = uplre,utre $
CREATE      ; utrebar = Group Mean(utre, Str = firm) $
PLOT        ; Lhs = uplre ; Rhs = utrebar ; Grid
            ; Title = Group Means of u(i,t) vs. Time Invariant u(i) $
PLOT        ; Lhs = utfe ; Rhs = utre ; Grid
            ; Title = Time Varying FE u(i) vs. Time Varying RE u(i) $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                       LQ
Log likelihood function         156.04955
Estimation based on N =    256, K =   9
Stochastic frontier based on panel data
Estimation based on      25 individuals
Variances: Sigma-squared(v)=     .01342
           Sigma-squared(u)=     .06529
           Sigma(v)        =     .11582
           Sigma(u)        =     .25552
Sigma = Sqr[(s^2(u)+s^2(v)]=     .28054
Gamma = sigma(u)^2/sigma^2 =     .82955
Var[u]/{Var[u]+Var[v]}     =     .63879
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0            108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =   95.950
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -1.70327***      .41761     -4.08  .0000    -2.52176   -.88477
      LF|     .19534**       .09759      2.00  .0453      .00407    .38662
      LM|     .81312***      .06954     11.69  .0000      .67682    .94941
      LE|    1.12741***      .34589      3.26  .0011      .44947   1.80534
      LL|    -.32931***      .07230     -4.55  .0000     -.47102   -.18760
      LP|     .22206***      .06265      3.54  .0004      .09927    .34485
      LK|    -.86072**       .42646     -2.02  .0436    -1.69657   -.02488
        |Variance parameters for compound error
  Lambda|    2.20605*       1.31249      1.68  .0928     -.36639   4.77849
Sigma(u)|     .25552**       .10148      2.52  .0118      .05661    .45442
--------+--------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                       LQ
Log likelihood function       108.43918
Estimation based on N =    256, K =   9
Variances: Sigma-squared(v)=     .01902
           Sigma-squared(u)=     .01692
           Sigma(v)        =     .13791
           Sigma(u)        =     .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=     .18957
Gamma = sigma(u)^2/sigma^2 =     .47074
Var[u]/{Var[u]+Var[v]}     =     .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***      .72136    -4.14  .0000    -4.40206  -1.57439
      LF|     .37257***      .07038     5.29  .0000      .23463    .51052
      LM|     .69910***      .07580     9.22  .0000      .55054    .84766
      LE|    2.09473***      .68790     3.05  .0023      .74647   3.44299
      LL|    -.42909***      .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***      .09498     4.69  .0000      .25917    .63149
      LK|   -2.09806***      .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***      .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***      .00064   297.81  .0000      .18832    .19082
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

These are the estimates of the true random effects model. Note that the variation of the random terms in the model has been rearranged. In the pooled model, s_v = 0.138 and s_u = 0.130. In the random effects model, we have s_v = .099 and s_u= .100. But, s_w = .140. The proportional allocation of the total to u and v has stayed roughly the same, but some additional variation is now attributed to the random effect. Note that the production function parameters have changed substantially as well.

```
-----------------------------------------------------------------------------
Random Coefficients  Frontier Model
Dependent variable                    LQ
Log likelihood function       160.58066
Restricted log likelihood        .00000
Chi squared [   1 d.f.]       321.16131
Significance level               .00000
Estimation based on N =    256, K =  10
Inf.Cr.AIC  =   -301.2 AIC/N =   -1.176
Model estimated: Aug 22, 2011, 23:15:44
Unbalanced panel has     25 individuals
Stochastic frontier (half normal model)
Simulation based on     50 Halton draws
Sigma( u) (1 sided)   =          .09962
Sigma( v) (symmetric) =          .09857
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Production / Cost parameters, nonrandom first
      LF|      .20387***      .05183     3.93  .0001      .10229    .30545
      LM|      .79450***      .04660    17.05  .0000      .70318    .88583
      LE|     1.10745***      .33573     3.30  .0010      .44943   1.76547
      LL|     -.32691***      .04277    -7.64  .0000     -.41074   -.24308
      LP|      .22812***      .05403     4.22  .0000      .12223    .33401
      LK|     -.84947**       .38344    -2.22  .0267    -1.60101   -.09794
        |Means for random parameters
Constant|    -1.83727***      .35442    -5.18  .0000    -2.53191  -1.14263
        |Scale parameters for dists. of random parameters
Constant|      .11729***      .00934    12.56  .0000      .09898    .13559
        |Variance parameter for v +/- u
   Sigma|      .14015***      .01373    10.21  .0000      .11325    .16705
        |Asymmetry parameter, lambda
  Lambda|     1.01064**       .43792     2.31  .0210      .15234   1.86895
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|       Mean        Std.Dev.      Minimum      Maximum     Cases Missing
--------+---------------------------------------------------------------------
   UPLRE|      .221170      .117670      .016992      .435912       256       0
    UTRE|      .078815      .031677      .026405      .305595       256       0
--------+---------------------------------------------------------------------
```

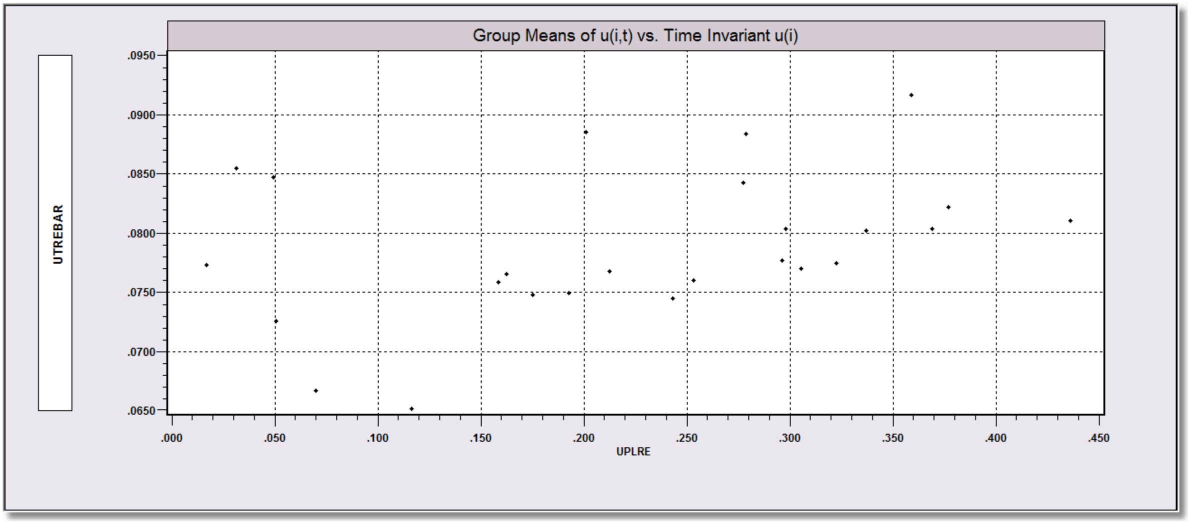


Figure E64.9 Time Varying vs. Time Invariant Estimates of u(i)

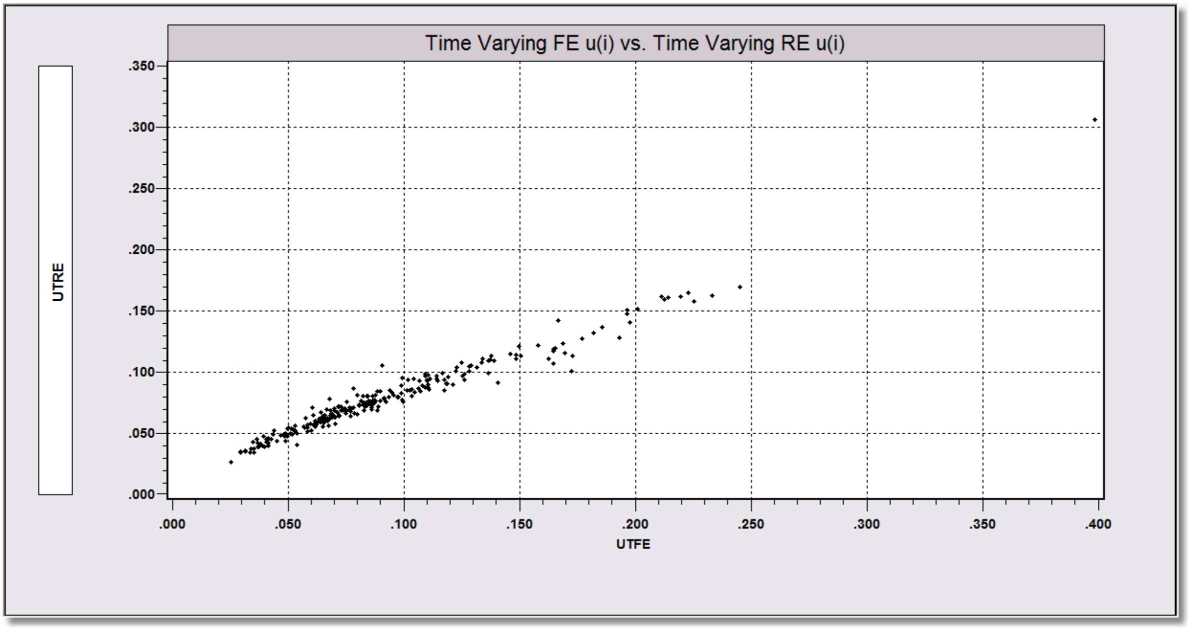


Figure E64.10 Comparison of Time Varying Fixed and Random Effects Estimates

E64.9 Random Parameters Stochastic Frontier Models

The random parameters stochastic frontier model in *LIMDEP* is very general, and embodies all three of the formulations discussed in the preceding sections on fixed and random effects.

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}_i'\mathbf{x}_{it} + v_{it} - u_{it},\\
u_i &= |\, \text{N}[\mu_{it}, \sigma_{uit}^2] \,|\\
\mu_{it} &= \boldsymbol{\delta}_i'\mathbf{m}_{it}.\\
\sigma_{uit}^2 &= \sigma_u^2 \times \exp(\boldsymbol{\gamma}_i'\mathbf{w}_{it}).
\end{aligned}
$$

The model allows, all at once, half normal or truncated normal distribution for u_i and firmwise and/or timewise heteroscedasticity in u_{it}. The model form allows parameters to be random in all three parts of the specification with the single restriction noted below. (Only the variance of the 'disturbance,' v_{it} is assumed to be constant. In addition, this model form does not accommodate heteroscedasticity in v_{it}.) As will be clear in what follows, the true random effects model developed in the previous section is a special case of this model with nonrandom parameters in μ_{it} and σ_{uit}^2 and only a random constant term in $\boldsymbol{\beta}_i$.

NOTE: The random parameters normal-truncated normal model with heteroscedasticity (in u_{it}) at the same time is not identified. Only one of these two should be specified. The command parser will not prevent you from specifying such a model, but it will ultimately be impossible to obtain the parameter estimates.

The general structure of the random parameters stochastic frontier model is based on the conditional density

$$f(y_{it}|\, \mathbf{x}_{it}, \boldsymbol{\beta}_i) = f(\boldsymbol{\beta}_i'\mathbf{x}_{it}),\ i = 1,...,N,\ t = 1,...,T_i$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$$

and $f(.)$ is the density for the stochastic frontier regression model. The model assumes that parameters are randomly distributed with possibly heterogeneous (across individuals) means

$$\text{E}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

(the second term is optional – the mean may be constant), and

$$\text{Var}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] = \boldsymbol{\Sigma}.$$

As noted earlier, the heterogeneity term is optional. In addition, it may be assumed that some of the parameters are nonrandom by placing rows of zeros in the appropriate places in $\boldsymbol{\Delta}$and $\boldsymbol{\Gamma}$. The general form of random parameter vector $\boldsymbol{\beta}_i$ is also extended to $\boldsymbol{\delta}_i$ and $\boldsymbol{\gamma}_i$. The general aspects of random parameters model estimation in *LIMDEP* are described in Chapter R24.

Command for the Random Parameters Model

The model command for the random parameters form of the stochastic frontier model is as follows. The first **FRONTIER** command is mandatory, and is needed to obtain the starting values. This is a pooled data version of the model. Note that it does not include the heteroscedasticity or truncation specification, even if the second command does.

FRONTIER ; Lhs = dependent variable ; Rhs = independent variables
; Parameters $
FRONTIER ; Lhs = dependent variable
; Rhs = independent variables
[; Rh2 = list is optional for the truncated normal model]
[; Hfn = list is optional for the heteroscedasticity model]
; Pds = fixed periods or count variable
; RPM (may include = variables in z)
; Fcn = random parameters specification $

(Note, again, only one of the two optional specifications noted should be specified.)

NOTE: For this model, your Rhs list must include a constant term. Though not strictly necessary, you should also include constants in Rh2 or Hfn if they are specified.

Specifying Random Parameters

The **; Fcn = specification** is used to define the random parameters. It is constructed from the list of Rhs names as follows: Suppose your model is specified by

; Rhs = one, x1, x2, x3, x4

This involves five coefficients. Any or all of them may be random; any not specified as random are assumed to be constant. For those that you wish to specify as random, use the following for production (cost, profit) function parameters,

; Fcn = variable name (distribution),
variable name (distribution), ...

There are two other sets of parameters in the model, in the mean of and variance of the one sided disturbance. To specify random parameters in the underlying mean of the truncated normal variable, use the following:

; Fcn = variable name [distribution],
variable name [distribution], ...

(Note square brackets designate the terms in μ_{it}.) For parameters in the computation of the variance of u_{it}, use

; Fcn = variable name <distribution>,
variable name <distribution>, ...

The difference in the three formulations is in the enclosures, ( ) for production function, [ ] for mean of the truncated distribution, and <> for the variance of the one sided disturbance. This distinction is necessary because the lists might have variables in common, and this is the only way to distinguish them. In particular, it is likely that all three lists would include *one*, so this device is used to distinguish the three functions.

Three distributions may be specified All random variables have mean 0.

n = standard normal distribution, variance = 1,
t = triangular (tent shaped) distribution in [-1,+1], variance = 1/6,
u = standard uniform distribution [-1,1], variance = 1/3.

Note that each of these is scaled as it enters the distribution, so the variance is only that of the random draw before multiplication. (See Chapter R23 for discussion of this computation and for other distributions that can be specified.) The latter two distributions are provided as one may wish to reduce the amount of variation in the tails of the distribution of the parameters across individuals and to limit the range of variation. (See Train (2010) for discussion.) For example, to specify that the constant term and the coefficient on *x*1 are normally distributed with fixed mean and variance, and a normally distributed constant in the mean of the truncated distribution, you might use

; Fcn = one(n), x1(n), one[n]

This specifies that the first and second coefficients are random while the remainder are not. The parameters estimated will be the mean and standard deviations of the distributions of these two parameters and the fixed values of the other three.

NOTE: If you use the wrong enclosures for the variables, a diagnostic will appear that the program does not recognize a variable. For example:

```
FRONTIER    ; Lhs = lq ; Rhs = one,lf,lm,le,ll,lp
            ; Hfn = one,lf ; RPM ; Pds = ni
            ; Fcn = one(n),lf(n),lf[n] $
```

```
Variable in FCN=name[type] is not in RHS/RH2/HFN list.
```

The reason for the diagnostic is that the **lf[n]** would indicate a specification for the truncation model, using **; Rh2 = list**. But, this command specifies only heteroscedasticity, which is denoted with <> enclosures. Hence, when the **lf[n]** is encountered, *LIMDEP* searches for *lf* in an Rh2 list, and finding no such list, issues the diagnostic.

Correlated Random Parameters

The stochastic frontier model does not support correlated random parameters. The model is not identified with this extension.

Heterogeneity in the Means

The preceding examples have specified that the mean of the random variable is fixed over individuals. If there is measured heterogeneity in the means, in the form of

$$E[\beta_{ki}] = \beta_k + \Sigma_m \delta_{km} z_{mi}$$

where z_{mi} is a variable that is measured for each individual, then the command may be modified to

; RPM = list of variables in z

In the data set, these variables must be repeated for each observation in the group. Since the coefficients are assumed to be time invariant, the variables in $\mathbf{z}_i$ must be also.

The Parameter Vector and Retained Results

The variances of the underlying random variables are given earlier, 1 for the normal distribution, 1/3 for the uniform, and 1/6 for the tent distribution. The σ_k parameters are only the standard deviations for the normal distribution. For the other two distributions, σ_k is a scale parameter. The standard deviation is obtained as $\sigma_k/\sqrt{3}$ for the uniform distribution and $\sigma_k/\sqrt{6}$ for the triangular distribution. When the parameters are correlated, the implied covariance matrix is adjusted accordingly. The correlation matrix is unchanged by this.

Results saved by this estimator are:

Matrices:	*b*	= estimate of $\boldsymbol{\theta}$
	varb	= asymptotic covariance matrix for estimate of $\boldsymbol{\theta}$.
	beta_i	= individual specific parameters, if **; Par** is requested.
Scalars:	*kreg*	= number of variables in Rhs
	nreg	= number of observations
	logl	= log likelihood function

Last Model: *b_variables*

Last Function: None

Standard Model Specifications for the Stochastic Frontier Random Parameters Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par keeps individual specific parameter estimates.
; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Robust requests a 'sandwich' estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Tlg[= value] sets convergence value for gradient.
; Tlf [= value] sets convergence value for function.
; Tlb[= value] sets convergence value for parameters.
; Alg = name requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n sets the maximum iterations.
; Output = n requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List displays a list of fitted values with the model estimates.
; Keep = name keeps fitted values as a new (or replacement) variable in data set.
; Res = name keeps residuals as a new (or replacement) variable.

Hypothesis Tests and Restrictions

; Test: spec defines a Wald test of linear restrictions.
; Wald: spec defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec defines a constrained maximum likelihood estimator.
; Rst = list specifies equality and fixed value restrictions.

Application

We continue the earlier application by fitting the stochastic frontier model with random parameters. The random parameters truncation model appears to be unidentified in these data, so the second model fit is with heteroscedasticity. In the first model, the constant and one of the production coefficients is specified to be random. In the second, these two coefficients and the parameter on the variable that enters the variance function are all taken to be random. The kernel density estimators compare the efficiency estimates from the random parameters model to those from the simplest pooled estimator.

The commands are:

```
NAMELIST    ; x = one,lf,lm,le,ll,lp,lk $
FRONTIER    ; Lhs = lq ; Rhs = x ; Eff = u $
FRONTIER    ; Lhs = lq ; Rhs = x
            ; RPM ; Panel ; Pts = 50 ; Halton; Fcn = one(n),lf(n) ; Eff = urp1 $
KERNEL      ; Rhs = urp1,u $
FRONTIER    ; Lhs = lq ; Rhs = x $
FRONTIER    ; Lhs = lq ; Rhs = x ; Hfn = one,loadfctr
            ; RPM ; Panel ; Pts = 50 ; Halton
            ; Fcn = one(n),lf(n),loadfctr<n> $
```

```
-----------------------------------------------------------------------------
Random Coefficients  Frontier Model
Dependent variable                       LQ
Log likelihood function           161.33196
Restricted log likelihood            .00000
Chi squared [   2 d.f.]           322.66392
Significance level                   .00000
Estimation based on N =    256, K =  11
Inf.Cr.AIC  =   -300.7 AIC/N =   -1.174
Model estimated: Aug 22, 2011, 23:28:18
Unbalanced panel has     25 individuals
Stochastic frontier (half normal model)
Simulation based on     50 Halton draws
Sigma( u) (1 sided)   =          .10598
Sigma( v) (symmetric) =          .09399
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Production / Cost parameters, nonrandom first
      LM|     .81447***      .04526    18.00  .0000      .72577    .90317
      LE|    1.16342***      .31391     3.71  .0002      .54817   1.77867
      LL|    -.33712***      .04111    -8.20  .0000     -.41769   -.25654
      LP|     .24213***      .04782     5.06  .0000      .14841    .33585
      LK|    -.94502***      .35520    -2.66  .0078    -1.64119   -.24886
        |Means for random parameters
Constant|   -1.89056***      .33140    -5.70  .0000    -2.54009  -1.24103
      LF|     .21430***      .05277     4.06  .0000      .11088    .31773
        |Scale parameters for dists. of random parameters
Constant|     .12526***      .00926    13.53  .0000      .10711    .14341
      LF|     .04979***      .00823     6.05  .0000      .03366    .06592
        |Variance parameter for v +/- u
   Sigma|     .14165***      .01265    11.20  .0000      .11686    .16645
        |Asymmetry parameter, lambda
  Lambda|    1.12768***      .42335     2.66  .0077      .29792   1.95743
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Figure E64.11 shows the distributions of the estimates of inefficiencies from the random parameters model and the simple, pooled fixed parameters model. The figure suggests that the RP formulation is moving some of the variation of the outcome variable out of the inefficiency term and into the production model, in the form of parameter variation.

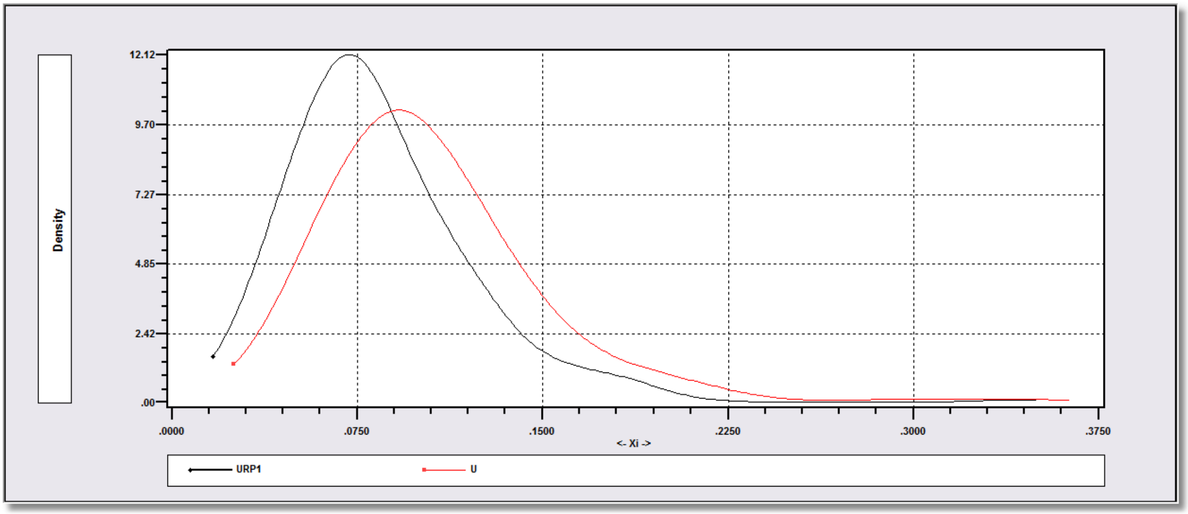


Figure E64.11 Kernel Density Estimator for Random Parameters Model Inefficiencies

```
----------------------------------------------------------------------------
Random Coefficients  FrntrTrn Model
Dependent variable                     LQ
Log likelihood function       199.14429
Estimation based on N =    256, K =  13
Unbalanced panel has     25 individuals
Stochastic frontier, truncation/hetero.
Simulation based on     50 Halton draws
Estimated parameters of efficiency dstn
s(u)  =     .189842  s(v)=        .07165
avgE[u|e]=   .10986  avgE[TE|e]= .90303
Lambda  = su/sv =               2.64974
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Nonrandom parameters
      LM|     .62243***      .04223    14.74  .0000      .53966    .70521
      LE|     .38353         .28063     1.37  .1717    -.16649    .93355
      LL|    -.36579***      .03589   -10.19  .0000    -.43614   -.29544
      LP|     .15282***      .04217     3.62  .0003     .07017    .23547
      LK|    -.16125         .31392     -.51  .6075    -.77652    .45401
   suONE|    9.05239***     1.65934     5.46  .0000    5.80014  12.30464
        |Means for random parameters
Constant|   -1.17144***      .29799    -3.93  .0001   -1.75549   -.58739
      LF|     .49011***      .04904     9.99  .0000     .39398    .58623
suLOADFC|   -16.4160***     3.47560    -4.72  .0000   -23.2281   -9.6039
        |Scale parameters for dists. of random parameters
Constant|     .12591***      .00859    14.65  .0000     .10906    .14275
      LF|     .01186**       .00593     2.00  .0456     .00023    .02350
suLOADFC|    1.47653***      .36192     4.08  .0000     .76718   2.18589
        |Sigma(v) from symmetric disturbance.
Sigma(v)|     .07165***      .00670    10.69  .0000     .05851    .08478
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

E64.10 Alvarez et al. – Fixed Management Model

Alvarez, Arias and Greene (2006) suggested a production model in which an unobserved factor enters as a latent variable. The core production model is

$$y_{it} = f(x_{it,1},x_{it,2},...,x_{it,K}, m_i)$$

where the unobservable, time invariant factor, 'm_i' is labeled 'management' in their paper. By treating the unobserved factor as a random component in the model, the authors develop a stochastic frontier model in which the resultant functional form is such that all random parameters are functions of the same single random effect, v_i, and the v_i appears in squared form in the equation as well. In generic terms, this model is a random parameters stochastic frontier model with random constant term and first order terms, and nonrandom second order terms in a translog model. The functional form is

$$\begin{aligned}
\log y_{it} &= \alpha_i + \sum_{k=1}^{K}\beta_{k,i}\ln x_{it,k} + \sum_{k=1}^{K}\sum_{m=1}^{K}\gamma_{km}\ln x_{it,k}\ln x_{it,m} + v_{it} - u_{it} \\
\alpha_i &= \alpha + \theta_\alpha w_i + \theta_{\alpha\alpha}(\tfrac{1}{2}w_i^2) \\
\beta_{k,i} &= \beta_k + \lambda_k w_i \\
w_i &\sim N[0,1] \\
v_{it} &\sim N[0,\sigma_v^2] \\
u_{it} &= |N[0,\sigma_u^2]|
\end{aligned}$$

This model is specified simply by creating the necessary variables, then building a random parameters model with the two additional specifications,

; Common ; Mgt

The **; Common** specification alone is generic, and applies to all random parameters models. Use it to specify that the same random component appears in all random parameters. The **; Mgt** specification has no function outside the frontier model. It is used only with the frontier model to specify this particular form. For example, consider the following three factor translog model:

```
FRONTIER    ; Lhs = yit ; Rhs = one,x1,x2,x3,x11,x12,x13,x22,x23,x33 $
FRONTIER    ; Lhs = yit ; Rhs = one,x1,x2,x3,x11,x12,x13,x22,x23,x33
            ; RPM ; Pds = the panel specification ; Halton
            ; Fcn = one(n),x1(n),x2(n),x3(n)
            ; Common ; Mgt $
```

(It is always necessary to fit the frontier model with fixed parameters first to generate the starting values.)

An extension of this model that the authors considered was intended to ameliorate the probable correlation between the random effect w_i and the independent variables (factors). The Mundlak approach to this problem is to incorporate the group means of the variables in the model. For this model, they proposed

$$w_i = \tau + \sum_{k=1}^{K} \gamma_k \overline{\log x_{i,k}} + f_i$$

where f_i is now the structural random variable that drives the random parameters. This extension is requested with

; Means

(The program deduces internally which variables are nonconstant and should be used.)

Application

The following is the Alvarez, Arias and Greene application. The data consists of six years of observations on 247 Spanish dairy farms. The output, *yit* is milk production. The four inputs, *x*1, *x*2, *x*3 and *x*4 are feed, land, labor and cows. Commands for fitting the model are as follows: (We have restricted the number of iterations and the number of replications for purpose of this numerical illustration.) Both models (with and without the Mundlak adjustment) are shown.

```
FRONTIER    ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4,x11,x12,x13,x14,x23,x24,x34,x44 ; Par $
FRONTIER    ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4,x11,x12,x13,x14,x23,x24,x34,x44
            ; RPM ; Halton ; Pts = 25 ; Pds = 6 ; Maxit = 25 ; Common ; Mgt
            ; Fcn = one(n),x1(n),x2(n),x3(n),x4(n) $
FRONTIER    ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4,x11,x12,x13,x14,x23,x24,x34,x44 ; Par $
FRONTIER    ; Lhs = yit
            ; Rhs = one,x1,x2,x3,x4,x11,x12,x13,x14,x23,x24,x34,x44
            ; RPM ; Halton ; Pts = 25 ; Pds = 6 ; Maxit = 25
            ; Common ; Mgt ; Means
            ; Fcn = one(n),x1(n),x2(n),x3(n),x4(n) $
```

The first set of results is the pooled stochastic frontier model with no extensions or modifications.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                  YIT
Log likelihood function       851.16734
Estimation based on N =   1482, K =  15
Variances: Sigma-squared(v)=     .00876
           Sigma-squared(u)=     .02831
           Sigma(v)        =     .09359
           Sigma(u)        =     .16825
Sigma = Sqr[(s^2(u)+s^2(v)]=     .19253
Gamma = sigma(u)^2/sigma^2 =     .76371
Var[u]/{Var[u]+Var[v]}     =     .54012
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          829.23705
Chi-sq=2*[LogL(SF)-LogL(LS)] =   43.861
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.6942***      .00529  2209.86  .0000     11.6838   11.7046
      X1|     .60483***      .02133    28.35  .0000      .56302    .64664
      X2|     .02246**       .01140     1.97  .0489      .00011    .04480
      X3|     .02336*        .01245     1.88  .0606    -.00104    .04776
      X4|     .44945***      .01172    38.34  .0000      .42647    .47242
     X11|     .59297***      .13525     4.38  .0000      .32789    .85806
     X12|    -.17183***      .04842    -3.55  .0004    -.26673   -.07693
     X13|     .20033***      .06903     2.90  .0037      .06502    .33563
     X14|    -.32993***      .07299    -4.52  .0000    -.47297   -.18688
     X23|     .00386         .04203      .09  .9268    -.07852    .08624
     X24|     .06473**       .03009     2.15  .0314      .00576    .12369
     X34|    -.07096*        .03853    -1.84  .0655    -.14648    .00455
     X44|     .20854***      .04328     4.82  .0000      .12373    .29336
        |Variance parameters for compound error
  Lambda|    1.79780***      .10292    17.47  .0000     1.59608   1.99951
   Sigma|     .19253***      .00011  1715.95  .0000      .19231    .19275
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the fixed management model without the Mundlak correction.

```
+----------------------------------------------+
| Random Coefficients  Frontier Model          |
| Dependent variable                 YIT       |
| Log likelihood function      1327.58807      |
| Estimation based on N =    1482, K =  21     |
| Sample is  6 pds and    247 individuals      |
+----------------------------------------------+
-----------------------------------------------------------------------------
All parameters have the same random effect
Alvarez/Arias/Greene Fixed Mgt. SF Model
Stochastic frontier (half normal model)
Simulation based on     25 Halton draws
Sigma( u) (1 sided)   =          .09355
Sigma( v) (symmetric) =          .05799
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Production / Cost parameters, nonrandom first
     X11|     .19550**       .08392     2.33  .0198      .03101    .35999
     X12|    -.00410         .02903     -.14  .8876     -.06100    .05279
     X13|    -.03972         .04116     -.96  .3346     -.12039    .04095
     X14|    -.08681**       .04220    -2.06  .0397     -.16952   -.00410
     X23|     .02377         .02534      .94  .3483     -.02590    .07344
     X24|    -.01893         .01743    -1.09  .2775     -.05310    .01524
     X34|     .02550         .02305     1.11  .2684     -.01967    .07067
     X44|     .09988***      .02339     4.27  .0000      .05403    .14572
        |Means for random parameters
Constant|    11.6506***      .00445  2620.80  .0000     11.6418   11.6593
      X1|     .65048***      .01227    53.03  .0000      .62643    .67452
      X2|     .03525***      .00681     5.17  .0000      .02190    .04861
      X3|     .04531***      .00759     5.97  .0000      .03043    .06019
      X4|     .40147***      .00646    62.16  .0000      .38881    .41413
        |Coefficients on unobservable fixed management
Constant|     .12579***      .00238    52.96  .0000      .12114    .13045
      X1|    -.02248*        .01218    -1.85  .0649     -.04635    .00139
      X2|     .00767         .00851      .90  .3676     -.00902    .02436
      X3|     .00794         .00939      .85  .3979     -.01047    .02635
      X4|    -.00967         .00657    -1.47  .1410     -.02255    .00320
Alpha_mm|    -.02835***      .00414    -6.85  .0000     -.03646   -.02024
        |Variance parameter for v +/- u
   Sigma|     .11007***      .00289    38.04  .0000      .10439    .11574
        |Asymmetry parameter, lambda
  Lambda|    1.61332***      .11959    13.49  .0000     1.37893   1.84771
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------+
| Random Coefficients  Frontier Model          |
| Dependent variable                  YIT      |
| Log likelihood function      1273.63070      |
| Sample is  6 pds and    247 individuals      |
+----------------------------------------------+
--------------------------------------------------------------------------------
All parameters have the same random effect
Alvarez/Arias/Greene Fixed Mgt. SF Model
Stochastic frontier (half normal model)
Simulation based on     25 Halton draws
Sigma( u) (1 sided)   =          .12577
Sigma( v) (symmetric) =          .05376
--------+-----------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Production / Cost parameters, nonrandom first
     X11|    -.06957         .08521     -.82  .4142     -.23658    .09743
     X12|     .00164         .02989      .05  .9562     -.05693    .06022
     X13|     .31592***      .04339     7.28  .0000      .23087    .40097
     X14|    -.08946*        .04767    -1.88  .0606     -.18289    .00398
     X23|    -.02088         .02784     -.75  .4533     -.07545    .03369
     X24|    -.04357**       .01912    -2.28  .0227     -.08103   -.00610
     X34|    -.15581***      .02350    -6.63  .0000     -.20187   -.10975
     X44|     .16310***      .02763     5.90  .0000      .10895    .21725
        |Means for random parameters
Constant|    11.6829***      .00449  2601.72  .0000     11.6741   11.6917
      X1|     .60260***      .02198    27.41  .0000      .55951    .64569
      X2|     .05221***      .01636     3.19  .0014      .02015    .08427
      X3|     .10728***      .02775     3.87  .0001      .05290    .16166
      X4|     .39780***      .01047    38.00  .0000      .37728    .41832
        |Coefficients on unobservable fixed management
Constant|     .11398***      .00235    48.52  .0000      .10937    .11858
      X1|    -.05393***      .01134    -4.76  .0000     -.07616   -.03171
      X2|     .03061***      .00916     3.34  .0008      .01265    .04857
      X3|     .01309         .01202     1.09  .2760     -.01046    .03665
      X4|     .01621**       .00707     2.29  .0218      .00236    .03007
Alpha_mm|    -.03575***      .00368    -9.72  .0000     -.04296   -.02855
        |Variance parameter for v +/- u
   Sigma|     .13678***      .00368    37.19  .0000      .12957    .14399
        |Asymmetry parameter, lambda
  Lambda|    2.33925***      .14491    16.14  .0000     2.05524   2.62326
        |Variable Means in Unobserved Management
  X1_bar|    -.12466         .22073     -.56  .5722     -.55728    .30796
  X2_bar|     .00045         .15758      .00  .9977     -.30839    .30930
  X3_bar|     .01632         .25437      .06  .9489     -.48224    .51487
  X4_bar|     .15107         .11332     1.33  .1825     -.07102    .37316
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

E64.11 Latent Class Stochastic Frontier Models

The latent class framework discussed in Chapter E20 is available for the stochastic frontier model. The structural equations of the basic model are

$$y_{it} \mid j = \boldsymbol{\beta}_j'\mathbf{x}_{it} + v_{it} - u_{it},$$

$$v_i \mid j = \mathrm{N}[0, \sigma_{vj}^2]$$

$$u_i \mid j = \mid \mathrm{N}[\sigma_{uj}^2] \mid$$

where 'j' indicates class j. The truncation and heteroscedasticity models are not supported by this estimator. However, the Battese and Coelli model, in which

$$u_{it} \mid j = g(\mathbf{z}_{it})| j \times |U_i|$$

is available for both forms of $g(\mathbf{z}_{it})$.

The estimation command for the latent class stochastic frontier model is

FRONTIER ; Lhs = dependent variable
; Rhs = one, remaining variables ; Parameters $
FRONTIER ; Lhs = dependent variable
; Rhs = one, remaining variables
; Pds = fixed periods or count variable
; LCM ; Pts = number of classes (2, 3, ..., 9) $

(As in other panel data settings, it is necessary to fit the pooled model first to compute the starting values.)

The Battese and Coelli models may be specified here with

; Model = BC

for the decay model and

; Model = BC
; Hfu = one, heteroscedasticity variables

For this model, you must fit the identical Battese and Coelli model without the latent class specification first. The application below demonstrates.

The basic form of the latent class model assumes that the class probabilities are fixed values. You may make them dependent on time invariant variables, w_i with

; LCM = list of variables in w

Do not include *one* in the list.

Some particular variables computed for the latent class model are

; Group = the index of the most likely latent class
; Cprob = estimated probability for the most likely latent class

You can obtain a listing of these two results by using

; List

An example appears below. You can also use the **; Rst = list** option to structure the latent class model so that different variables appear in different classes or that certain coefficients are equal across classes. Examples are given in Chapter E20.

Estimates retained by this model include:

Matrices:	*b*	=	full parameter vector, $[\boldsymbol{\beta}_1'\lambda_1\sigma_1, \boldsymbol{\beta}_2',\lambda_2\sigma_2, \ldots F_1,\ldots,F_J]$
	varb	=	full covariance matrix
	beta_i	=	individual specific parameters, if **; Par** is requested

Note that *b* and *varb* involve $J\times(K+2)$ estimates. Two additional matrices are created,

b_class	=	a $J\times K$ matrix with each row equal to the corresponding $\boldsymbol{\beta}_j$
class_pr	=	a $J\times 1$ vector containing the estimated class probabilities

Scalars:	*kreg*	=	number of variables in Rhs list
	nreg	=	total number of observations used for estimation
	logl	=	maximized value of the log likelihood function
	exitcode	=	exit status of the estimation procedure

Standard Model Specifications for the Latent Class Stochastic Frontier Model

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Par	keeps individual specific parameter estimates.
; Partial Effects	displays marginal effects, same as **; Marginal Effects**.
; OLS	displays least squares starting values when (and if) they are computed.
; Table = name	saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.
; Robust	requests a ‘sandwich’ estimator or robust covariance matrix for TSCS and several discrete choice models.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Res = name	keeps residuals as a new (or replacement) variable.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

Application

The airline data used in the preceding examples are clearly not compatible with this model; no configuration of the equation produces meaningful results. To illustrate the estimator, we have borrowed the Spanish dairy data used in the previous section. The following commands fit a two class, Battese and Coelli decay model.

```
NAMELIST    ; x  = one,x1,x2,x3,x4 $
FRONTIER    ; Lhs = yit ; Rhs = x
            ; Model = BC
            ; Pds  = 6 $
FRONTIER    ; Lhs = yit ; Rhs = x
            ; Model = BC
            ; LCM ; Pts = 2 ; Pds = 6 ; List $
```

These are the initial results from the first command.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    YIT
Log likelihood function      1390.20024
Stochastic frontier based on panel data
Estimation based on     247 individuals
Variances: Sigma-squared(v)=     .00549
           Sigma-squared(u)=     .03940
           Sigma(v)        =     .07413
           Sigma(u)        =     .19848
Sigma = Sqr[(s^2(u)+s^2(v)]=     .21187
Gamma = sigma(u)^2/sigma^2 =     .87759
Var[u]/{Var[u]+Var[v]}     =     .72263
Stochastic Production Frontier, e = v-u
Battese-Coelli Models: Time Varying uit
Time dependent uit=exp[-eta(t-T)]*|U(i)|
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          809.67610
Chi-sq=2*[LogL(SF)-LogL(LS)] = 1161.048
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.7882***      .00716  1646.05  .0000     11.7742   11.8022
      X1|     .62230***      .01365    45.59  .0000      .59555    .64905
      X2|     .06001***      .01069     5.61  .0000      .03905    .08096
      X3|     .05708***      .01454     3.93  .0001      .02858    .08557
      X4|     .35510***      .00700    50.69  .0000      .34137    .36883
        |Variance parameters for compound error
  Lambda|    2.67761***      .02351   113.88  .0000     2.63152   2.72369
Sigma(u)|     .19848***      .00060   332.72  .0000      .19731    .19965
        |Eta parameter for time varying inefficiency
     Eta|     .08030***      .00432    18.60  .0000      .07184    .08877
--------+--------------------------------------------------------------------
```

```
Warning   141: Iterations:current or start estimate of sigma is nonpositive
Normal exit from iterations. Exit status=0.
-----------------------------------------------------------------------------
Latent Class / Panel Frontier Model
Dependent variable                  YIT
Log likelihood function      1462.93500
Estimation based on N =   1482, K =  17
Sample is  6 pds and    247 individuals
Stoch. frontier (B&C,time varying U)
Ineff=u(i,t)=exp(-eta*(t-T))|U(i)|
Model fit with  2 latent classes.
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    11.8355***      .02201   537.84  .0000     11.7923   11.8786
      X1|     .60324***      .03499    17.24  .0000      .53467    .67181
      X2|     .13327***      .04014     3.32  .0009      .05459    .21195
      X3|     .10581***      .03248     3.26  .0011      .04216    .16947
      X4|     .33560***      .01392    24.11  .0000      .30832    .36288
        |Square root of variance sum, sqr(s2u + s2v)
   Sigma|     .71161**       .35935     1.98  .0477      .00730   1.41591
        |Asymmetry parameter in compound distn, su/sv
  Lambda|     .02071         .02565      .81  .4194    -.02956    .07098
        |Scale factor in time varying inefficiency
     Eta|     .19551***      .01986     9.84  .0000      .15658    .23444
        |Model parameters for latent class 2
Constant|    11.7611***      .01279   919.62  .0000     11.7360   11.7862
      X1|     .61866***      .01873    33.04  .0000      .58196    .65536
      X2|     .05041***      .01289     3.91  .0001      .02514    .07567
      X3|     .06232***      .01830     3.40  .0007      .02645    .09820
      X4|     .30614***      .01029    29.76  .0000      .28598    .32631
        |Square root of variance sum, sqr(s2u + s2v)
   Sigma|     .92839***      .02938    31.60  .0000      .87081    .98597
        |Asymmetry parameter in compound distn, su/sv
  Lambda|     .05084         .22185      .23  .8187    -.38398    .48566
        |Scale factor in time varying inefficiency
     Eta|     .07059***      .00475    14.87  .0000      .06129    .07990
        |Estimated prior probabilities for class membership
Class1Pr|     .30612***      .05178     5.91  .0000      .20463    .40760
Class2Pr|     .69388***      .05178    13.40  .0000      .59240    .79537
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+--------------------------------------------------------+
| Stochastic Frontier Model Variance Parameters           |
| Class     Lambda      Sigma   Sigma(u)   Sigma(v) |
|     1    .020709    .711607    .014734    .711454 |
|     2    .050840    .928393    .047139    .927195 |
+--------------------------------------------------------+
===============================================================================
Predictions computed for the group with the largest posterior probability
Obs.  Periods Estimated inefficiencies, E[u|v -/+ u]
===============================================================================
Ind.=    1  J* = 1  P(j)=  .889  .111
       01-06  .3105  .2554  .2100  .1727  .1421  .1168
Ind.=    2  J* = 2  P(j)=  .295  .705
       01-06  .0813  .0757  .0706  .0657  .0613  .0571
Ind.=    3  J* = 2  P(j)=  .012  .988
       01-06  .2254  .2100  .1957  .1824  .1699  .1584
Ind.=    4  J* = 1  P(j)=  .955  .045
       01-06  .1778  .1463  .1203  .0989  .0814  .0669
Ind.=    5  J* = 1  P(j)=  .650  .350
       01-06  .2453  .2018  .1659  .1365  .1122  .0923
Ind.=    6  J* = 2  P(j)=  .138  .862
       01-06  .0517  .0482  .0449  .0418  .0390  .0363
Ind.=    7  J* = 1  P(j)=  .985  .015
       01-06  .3010  .2476  .2036  .1674  .1377  .1132
Ind.=    8  J* = 2  P(j)=  .165  .835
       01-06  .0561  .0523  .0487  .0454  .0423  .0394
Ind.=    9  J* = 2  P(j)=  .450  .550
       01-06  .0134  .0125  .0116  .0108  .0101  .0094
Ind.=   10  J* = 1  P(j)=  .999  .001
       01-06  .1039  .0855  .0703  .0578  .0475  .0391
 (Farms 11-247 omitted)
```

E65: Data Envelopment Analysis

E65.1 Introduction

There are two broad paradigms used by researchers to analyze efficiency in production, stochastic frontier analysis (SFA) and data envelopment analysis (DEA). No formulation has yet been devised that unifies SFA and DEA in a single analytical framework. Arguably, the former is a fully parameterized model whereas the latter is 'nonparametric,' albeit also atheoretical in nature. DEA is currently the conventional approach to deterministic frontier estimation. This is usually handled with linear programming techniques. The analysis assumes that there is a frontier technology (in the same spirit as the stochastic frontier production model) that can be described by a piecewise linear hull that envelopes the observed outcomes. Some (efficient) observations will be on the frontier while other (inefficient) individuals will be inside. The technique produces a deterministic frontier that is generated by the observed data, so by construction, some individuals are 'efficient.' This is one of the fundamental differences between DEA and SFA. This chapter presents *LIMDEP*'s programs for data envelopment analysis (DEA).

E65.2 Data Envelopment Analysis

Stochastic frontier modeling is based on maximum likelihood or other classical or Bayesian, parametric econometric techniques. In contrast, DEA is based on nonparametric, linear programming methods. Both paradigms are based on an underlying construct of the efficient production frontier that relates maximal output to inputs for the 'firm' (decision making unit, or DMU). Using SFA methods, the analyst defines, then estimates a continuous, regular relationship that defines the frontier. DEA uses linear programming methods to fit a piecewise linear 'hull' around the data, under the assumption that the hull adequately approximates the underlying frontier, the more so as the number of observations increases. (Since the technique is nonstatistical, this is difficult to establish analytically.) There is a vast literature on the two techniques and comparisons, none of which will be reviewed here. Our purpose here is only to document the estimator. We recommend, as a departure point in the literature, a working paper by Coelli (1996a), which describes the techniques documented here and introduces some of the theoretical notions. He also provides several useful citations.

E65.2.1 Input and Output Oriented Efficiency

The discussion of DEA efficiency measurement begins with the notion of a measure of the ratio of outputs to inputs for firm '*i*,'

$$Ratio_i = \boldsymbol{\alpha}'\mathbf{y}_i / \boldsymbol{\beta}'\mathbf{x}_i,\ i = 1,..,N,$$

where $\mathbf{y}_i$ is the vector of M outputs and $\mathbf{x}_i$ is the vector of K inputs. The optimal weights are defined by the programming problem,

$$\begin{aligned}
&\text{Maximize wrt } \boldsymbol{\alpha},\boldsymbol{\beta}: && \boldsymbol{\alpha}'\mathbf{y}_i / \boldsymbol{\beta}'\mathbf{x}_i \\
&\text{Subject to} && \boldsymbol{\alpha}'\mathbf{y}_s / \boldsymbol{\beta}'\mathbf{x}_s \leq 1,\ s = 1,...,N \\
& && \alpha_m \geq 0,\ m = 1,...,M \\
& && \beta_k \geq 0,\ k = 1,...,K
\end{aligned}$$

The optimization program seeks the optimal weights to maximize the 'efficiency' of firm *s* subject to the restriction that the efficiencies of all firms are less than or equal to one, and that all weights are nonnegative. Because the objective function is homogeneous of degree zero – any multiple of the weights produces the same solution – it is normalized with a restriction such as $\boldsymbol{\alpha}'\mathbf{x}_i = 1$. Transforming and simplifying the problem a bit produces the equivalent program,

$$
\begin{aligned}
&\text{Maximize wrt } \boldsymbol{\alpha},\boldsymbol{\beta}: && \boldsymbol{\alpha}'\mathbf{y}_i \\
&\text{Subject to} && \boldsymbol{\beta}'\mathbf{x}_i = 1 \\
& && \boldsymbol{\alpha}'\mathbf{y}_s - \boldsymbol{\beta}'\mathbf{x}_s \leq 0,\ s = 1,\ldots,N \\
& && \boldsymbol{\alpha} \geq \mathbf{0} \\
& && \boldsymbol{\beta} \geq \mathbf{0}
\end{aligned}
$$

An equivalent form of the problem is the envelopment form (hence the name),

$$
\begin{aligned}
&\text{Minimize wrt } \theta_i, \boldsymbol{\lambda}: && \theta_i \\
&\text{Subject to} && \Sigma_s\, \lambda_s\mathbf{y}_s - \mathbf{y}_i \geq \mathbf{0} \\
& && \theta_i\, \mathbf{x}_i - \Sigma\, \lambda_s\mathbf{x}_s \geq \mathbf{0} \\
& && \lambda_s \geq 0.
\end{aligned}
$$

The value of θ_i is the *input oriented technical efficiency score* for the *i*th firm

$$TE_{INPUT,i} = \theta_i.$$

It measures the extent to which the firm could reduce inputs to obtain the same output – relative to other firms in the sample. Note that the program is solved for each firm in the sample – an efficiency score θ_i is generated for each firm. For some firms in the sample, the efficiency score will be 1.0. This indicates firms deemed to be technically efficient. Otherwise, $\theta_i \leq 1$.

The preceding formulation includes an implicit assumption of constant returns to scale (CRS). The assumption is relaxed to variable returns to scale (VRS), by adding a restriction

$$\Sigma_s\, \lambda_s = 1.$$

Variable returns to scale is the standard assumption in contemporary applications. This provides a means by which the 'scale efficiency' of the firm can be measured. Let θ_{iC} denote the technical efficiency measure obtained assuming constant returns and θ_{iV} be the variable returns to scale counterpart. Then, the '*scale efficiency*' may be measured by

$$SE_i = \theta_{iC} / \theta_{iV}.$$

This can be computed using the results of the two different programs after computation. A 'nonincreasing returns to scale' (NRS) version of the program can be obtained by changing the adding up restriction to

$$\Sigma_s\, \lambda_s \leq 1.$$

An alternative view of the optimization process is to consider the extent to which outputs could conceivably be increased using the same inputs – again relative to the standard of other firms in the sample. The linear program which produces this solution is

$$\begin{aligned} &\text{Maximize wrt } \phi_i, \boldsymbol{\lambda}: && \phi_i \\ &\text{Subject to} && \Sigma_s \lambda_s \mathbf{y}_s - \phi_i \, \mathbf{y}_i \geq \mathbf{0} \\ & && \mathbf{x}_i - \Sigma \lambda_s \mathbf{x}_s \geq \mathbf{0} \\ & && \lambda_s \geq 0. \end{aligned}$$

Once again, this assumes constant returns to scale. The variable returns to scale form is obtained by adding the constraint $\Sigma_s\lambda_s = 1$. In this solution, $1 < \phi_i < \infty$. The technical efficiency measure is

$$0 < TE_{OUTPUT,i} = 1/\phi_i \leq 1$$

As before, some firms in the sample (the same firms) will be found to be technically efficient by this *output oriented efficiency measure*.

E65.2.2 Economic and Allocative Efficiency

With input price information, $\mathbf{w}_i$, (and assuming cost minimization) a cost minimization program to find the optimal inputs given the input prices is

$$\begin{aligned} &\text{Minimize wrt } \boldsymbol{\chi}_i, \boldsymbol{\lambda}: && \mathbf{w}_i' \, \boldsymbol{\chi}_i \\ &\text{Subject to} && \Sigma_s \lambda_s \mathbf{y}_s - \mathbf{y}_i \geq \mathbf{0} \\ & && \boldsymbol{\chi}_i - \Sigma \lambda_s \mathbf{x}_s \geq \mathbf{0} \\ & && \lambda_s \geq 0. \end{aligned}$$

As before, to allow for variable returns to scale (VRS), we add $\Sigma_s \lambda_s = 1$. In this program, $\boldsymbol{\chi}_i$ gives the cost minimizing vector of inputs for output $\mathbf{y}_i$ and input prices $\mathbf{w}_i$. The cost efficiency for the *i*th firm is then the ratio

$$0 < CE_i = \mathbf{w}_i \boldsymbol{\chi}_i / \mathbf{w}_i' \mathbf{x}_i < 1.$$

Allocative efficiency may be measured using

$$0 < AE_i = CE_i / TE_{INPUT,i} < 1.$$

E65.2.3 Solutions to the Optimization Problems

We note briefly the mathematical form of *LIMDEP*'s solutions to the linear programs above. The programming problem is defined in terms of

- Activity vector, $\boldsymbol{\gamma}$ = the solution vector
- Coefficient vector, $\mathbf{c}$ so that the objective function is $\mathbf{c}'\boldsymbol{\gamma}$
- Constraint matrix, $\mathbf{A}$
- Lower and upper limits for constraints, $\mathbf{b}_L$ and $\mathbf{b}_U$
- Lower and upper limits for activities, $\mathbf{d}_L$ and $\mathbf{d}_U$

The linear program solution, in general is, then,

$$\begin{aligned}
&\text{Optimize wrt } \gamma: && \mathbf{c}'\boldsymbol{\gamma} \\
&\text{Subject to} && \mathbf{b}_L < \mathbf{A}\boldsymbol{\gamma} < \mathbf{b}_U \\
& && \mathbf{d}_L < \boldsymbol{\gamma} < \mathbf{d}_U.
\end{aligned}$$

We will define the components for the three programs defined earlier. Note, first, for convenience, we define the data matrices, $\mathbf{Y}$ and $\mathbf{X}$. $\mathbf{Y}$ is an $N\times M$ matrix of outputs whose ith row is the vector of outputs for firm i; $\mathbf{X}$ is the $N\times K$ matrix of inputs, defined likewise. For an individual firm, we define $\mathbf{y}_i$ to the $M\times 1$ column vector of outputs for firm i; thus, $\mathbf{y}_i$ is the transpose of the ith row of $\mathbf{Y}$. Likewise, $\mathbf{x}_i$ is the column vector of K inputs for firm i, the transpose of the ith row of $\mathbf{X}$. Finally, the column vector of weights is $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_N)'$. Thus,

$$\Sigma_s\, \lambda_s\, \mathbf{y}_s \;=\; \mathbf{Y}'\boldsymbol{\lambda} \;\text{ and }\; \Sigma_s\, \lambda_s\, \mathbf{x}_s \;=\; \mathbf{X}'\boldsymbol{\lambda}.$$

Finally, we note once again, the programs about to be defined are solved for each firm to obtain the efficiency scores. (In fact, $\boldsymbol{\lambda}$ should be indexed by firm, since it is recomputed each time. For convenience, we have omitted this subscript.) We use the symbol ∞_K and ∞_M to indicate a vector whose each element equals infinity (or sometimes minus infinity) and boldface $\mathbf{1}$ or $\mathbf{0}$ to indicate a vector of ones or zeros with a subscript to indicate the number of elements. Finally, our tableaus include the VRS restriction, which may be suppressed by the user for the CRS form.

With all this in place, we can define the solutions to the optimization problems just by identifying the components of the linear programming problems. These are as follows:

Input Oriented Technical Efficiency

$$\mathbf{d}_L = \begin{bmatrix} \mathbf{0}_N \\ 0 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} \mathbf{0}_N \\ 1 \end{bmatrix}, \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\lambda} \\ \phi_i \end{bmatrix}, \mathbf{d}_U = \begin{bmatrix} \mathbf{1}_N \\ 1 \end{bmatrix}$$

$$\mathbf{b}_L = \begin{bmatrix} -\infty_K \\ \mathbf{y}_i \\ 1 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{X}' & -\mathbf{x}_i \\ \mathbf{Y}' & \mathbf{0}_M \\ \mathbf{1}'_N & 0 \end{bmatrix}, \mathrm{b}_U = \begin{bmatrix} \mathbf{0}_K \\ \infty_M \\ 1 \end{bmatrix}$$

Output Oriented Technical Efficiency

$$\mathbf{d}_L = \begin{bmatrix} \mathbf{0}_N \\ 1 \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} \mathbf{0}_N \\ 1 \end{bmatrix},\ \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\lambda} \\ \phi_i \end{bmatrix},\ \mathbf{d}_U = \begin{bmatrix} \mathbf{1}_N \\ \infty \end{bmatrix}$$

$$\mathbf{b}_L = \begin{bmatrix} -\infty_K \\ \mathbf{0}_M \\ 1 \end{bmatrix},\ \mathbf{A} = \begin{bmatrix} \mathbf{X}' & \mathbf{0}_K \\ \mathbf{Y}' & -\mathbf{y}_i \\ \mathbf{1}'_N & 0 \end{bmatrix},\ \mathbf{b}_U = \begin{bmatrix} \mathbf{x}_i \\ \infty_M \\ 1 \end{bmatrix}$$

Allocative Efficiency

$$\mathbf{d}_L = \begin{bmatrix} \mathbf{0}_N \\ \mathbf{0}_K \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} \mathbf{0}_N \\ \mathbf{w}_i \end{bmatrix},\ \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\chi}_i \end{bmatrix},\ \mathbf{d}_U = \begin{bmatrix} \mathbf{1}_N \\ \infty_{\mathrm{K}} \end{bmatrix}$$

$$\mathbf{b}_L = \begin{bmatrix} -\infty_K \\ \text{-}\mathbf{y}_i \\ 1 \end{bmatrix},\ \mathbf{A} = \begin{bmatrix} \mathbf{X}' & \text{-}\mathbf{I}_K \\ \mathbf{Y}' & \mathbf{0}_{M\times K} \\ \mathbf{1}_N' & \mathbf{0}_K' \end{bmatrix},\ \mathbf{b}_U = \begin{bmatrix} \mathbf{0}_K \\ \infty_M \\ 1 \end{bmatrix}$$

One final note, DEA requires a fair amount of computation. The linear program involves M+K+1 constraints and N+1 activities, and it is computed once for each of the N firms in the sample. The amount of computation increases with the square of N. The particular computations are quite fast, however

E65.3 Confidence Limits for Efficiency Scores

A major shortcoming of the DEA approach to modeling production is the absence of a statistical underpinning. One approach that has been used to try to produce some statistical characterization of the estimator is to use bootstrapping to obtain confidence limits for the estimated efficiency scores. A popular method used is that of Simar and Wilson (1998). In brief, their method amounts to the following: We have in hand for each firm a θ_i estimated using the linear program defined above. To carry out the bootstrap, we use the following experiment. The data on $\mathbf{x}_m$ for all firms, including this one, are proportionally scaled using a randomly generated (see their paper for the algorithm) scale factor, θ_i/τ_{mb} for replication b. Then, $\theta_{i,b}$ is recomputed using the revised data, with the same method. The experiment is repeated B times. The 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles of the B observations provide the confidence limits. This is repeated B times for each firm. To obtain bootstrapped confidence use the command syntax described below, with the simple addition of the request for the number of bootstrap replications.

It should be noted, bootstrapping adds considerably to the amount of computation. In general, the analysis requires the computation of 2N linear programs, two for each firm, to compute the input and output oriented efficiency scores, plus one more if input prices are supplied for the allocative efficiency computation. Bootstrapping adds $B\times N$ more programs. Each program involves N+1 activities and K+M+1 constraints, so overall, the amount of computation is considerable. Nonetheless, each component of each linear program is very fast. In the example below, we have 123 observations. We requested 50 bootstrap replications, so we computed altogether 53×123 = 6,519 programs, each with 123 activities. The LP computations plus all the ancillary computations and the display took altogether only 3.84 seconds on our desktop computer.

E65.4 Command Structure

The command for the data envelopment analysis routine is simply

```
FRONTIER    ; Lhs = output variables
            ; Rhs = input variables  (will never include one)
            ; Alg  = DEA $
```

The following is the full list of specifications for this command.

The default specification uses the variable returns to scale form. If you wish to use the constant returns to scale form, add

; CRS

to the command. The nonincreasing returns to scale form ($\Sigma_i\ \lambda_i \leq 1$) is requested with

; NRS

Nondecreasing returns to scale is requested with **; NDS**.

If you wish to analyze input price data, add

; Rh2 = input price variables

The program computes the DEA efficiency scores (input and output oriented, and economic efficiency), and stores them as variables and as matrices. (See the description in the next section.) If you would like to see a listing of the scores on your screen, in the output window, add

; List

to the command. The list of 'peer' firms for each observation (see Section E65.5.1 below) may be requested by adding

; Peers

to the command. Finally, to obtain bootstrapped confidence limits for the estimator, add

; Nbt = the desired number of replications

E65.5 DEA Results

This estimator by default computes both the input and output oriented technical efficiency scores. Descriptive statistics for the results are the visible output from the estimator. The following shows an example, using the sample of 1,482 observations on Spanish dairy farms that was examined in Chapter E64. This is a one output, four input process.

FRONTIER ; Lhs = milk
; Rhs = cows,land,labor,feed
; Alg = DEA $

```
+--------------------------------------------------------------------+
| Data Envelopment Analysis                                          |
| Output Variables:  MILK                                            |
| Input Variables:   COWS     LAND     LABOR    FEED                 |
| Underlying Technology assumes VARIABLE Returns to Scale.           |
+--------------------------------------------------------------------+
| Estimated Efficiencies:     Mean     Std.Deviation     Minimum   Maximum  |
| Technical Efficiency     =======     =============     =======   =======  |
|      Input Oriented        .8301          .1416          .4823    1.0000  |
|     Output Oriented        .7388          .1268          .3875    1.0000  |
| Sample Size:           1482 Observations.    1482 Complete observations  |
| Efficiencies saved as variables DEAEFF_O, DEAEFF_I and DEAEFF_E         |
| Efficiencies saved as matrices  DEA_EFFO, DEA_EFFI and DEA_EFFE         |
| Incomplete observations are filled with zeros for efficiency values.    |
+--------------------------------------------------------------------+
```

As noted, the computed efficiency scores are saved in two places, in the data area, as variables *deaeff_i* and *deaeff_o* and *deaeff_e* if you provide input prices for the economic efficiency analysis. The same results are saved as matrices, *dea_effo*, *dea_effi*, *dea_effe*. Note that in both occurrences, the estimator is bypassing missing and bad (nonpositive) data. If any of the variables used in the analysis are missing, the observation is assigned an efficiency score of 0.0. The matrices will have row dimension equal to the original sample size, before the bypass of missing values.

The example below includes a listing of the efficiency scores. The observation identifier shows *I* = the sequence number of the observation used in the analysis. The *R* = value shows, instead, the actual location of the observation in the raw data set. *I* will not equal *R* if you have used a subset of the data (e.g., with **SAMPLE** or **REJECT)**, or if the program has bypassed missing data – the listing will only show the complete observations. If you have included observation labels, e.g., firm names, in your data set, these observation and row identifiers will be replaced with the observation names for your data set.

For a second example, the following analyzes the Christensen and Greene (1976) electricity generation data. For these data, we have the input prices, so we do the full analysis.

FRONTIER ; Alg = DEA ; List ; Nbt = 50
; Lhs = output
; Rhs = labor,capital,fuel
; Rh2 = lprice,cprice,fprice $

```
+-----------------------------------------------------------------------------+
| Data Envelopment Analysis                                                   |
| Output Variables:  OUTPUT                                                   |
| Input Variables:   LABOR    CAPITAL  FUEL                                   |
| Price Variables:   LPRICE   CPRICE   FPRICE                                 |
| Underlying Technology assumes VARIABLE Returns to Scale.                    |
+-----------------------------------------------------------------------------+
| Estimated Efficiencies:     Mean     Std.Deviation     Minimum   Maximum    |
| Technical Efficiency     =======     =============     =======   =======    |
|      Input Oriented        .7692          .1390          .3464    1.0000    |
|     Output Oriented        .7657          .1467          .2960    1.0000    |
| Economic Efficiency        .4331          .1965          .1411    1.0000    |
| Allocative Effic.          .5473          .1754          .1796    1.0000    |
| Sample Size:             123 Observations.     123 Complete observations    |
| Efficiencies saved as variables DEAEFF_O, DEAEFF_I and DEAEFF_E             |
| Efficiencies saved as matrices  DEA_EFFO, DEA_EFFI and DEA_EFFE             |
| Incomplete observations are filled with zeros for efficiency values.        |
| Compute allocative efficiency as technical divided by economic efficiency |
+-----------------------------------------------------------------------------+

Estimated Efficiency Values for Individual Decision Making Units
(Results are listed only for complete observations)
===============================================================================
Observation     | Input Oriented| Output Oriented|    Economic    | Allocative
Sample   Data   | Rank     Value| Rank      Value| Rank      Value| Rank   Value
================+===============+================+================+=============
I=     1 R=    1|     1  1.00000|     1   1.00000|    1   1.00000|    1 1.00000
I=     2 R=    2|    13   .98446|    16    .92501|   53    .43644|   87  .44333
I=     3 R=    3|    16   .96243|    28    .88393|  119    .17287|  123  .17962
I=     4 R=    4|    46   .79469|    83    .73593|   96    .29127|  103  .36652
I=     5 R=    5|   115   .57426|   118    .44224|   47    .44703|   15  .77845
I=     6 R=    6|   120   .44307|   122    .35608|  103    .26194|   43  .59120
I=     7 R=    7|    80   .73356|   100    .64826|  101    .26996|  102  .36801
I=     8 R=    8|   123   .34637|   123    .29601|  121    .15388|   85  .44425
I=     9 R=    9|   106   .62517|   110    .57829|  109    .21689|  111  .34692
I=    10 R=   10|   103   .63852|   107    .59578|   66    .38812|   39  .60783
 (Remaining observations are omitted.)
----------------------------------------------------------------------------
Results of Bootstrap analysis of technical efficiency.       50 replications
----------------------------------------------------------------------------
                  Technical  Estimated  Corrected  Standard   Confid. Limits
Observation_____  Efficiency    Bias    Tech.Eff.  Deviation  Lower    Upper
I=     1 R=    1    1.0000      .0000     1.0000      .0000   1.0000  1.0000
I=     2 R=    2     .9845    -.0634     1.0479      .1008    .6583  1.0000
I=     3 R=    3     .9624    -.0898     1.0522      .1391    .5023  1.0000
I=     4 R=    4     .7947     .1091      .6856      .0953    .7222  1.0000
I=     5 R=    5     .5743     .3006      .2737      .1215    .6007  1.0000
I=     6 R=    6     .4431     .4318      .0113      .1246    .5785  1.0000
I=     7 R=    7     .7336     .1086      .6250      .1131    .6609  1.0000
I=     8 R=    8     .3464     .5317     -.1853      .0979    .6977  1.0000
I=     9 R=    9     .6252     .2154      .4097      .1265    .5131  1.0000
I=    10 R=   10     .6385     .2267      .4118      .1062    .6645  1.0000
```

It is always interesting to compare the DEA results with those obtained using the stochastic frontier model. The following fits a translog stochastic frontier production function for the Christensen and Greene data, computes the technical efficiencies, and plots them against the DEA efficiency scores. As has been widely documented, the results are not so close to each other as one might hope.

```
FRONTIER   ; Lhs = logq
           ; Rhs = one,logcap,loglabor,logfuel,
                   loglsq,logksq,logfsq,logklogl,logklogf,logllogf
           ; Techeff = tesf $
PLOT       ; Lhs = tesf ; Rhs = deaeff_i
           ; Grid ; Title = DEA Efficiencies vs. Stochastic Frontier JLMS $
```

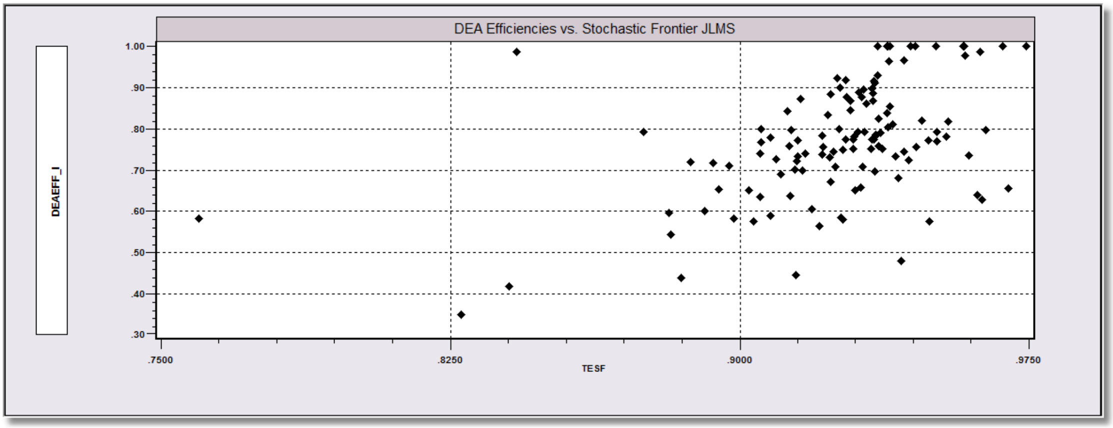


Figure E65.1 Comparison of SFA and DEA Efficiency Estimates

E65.5.1 Analysis of Peers

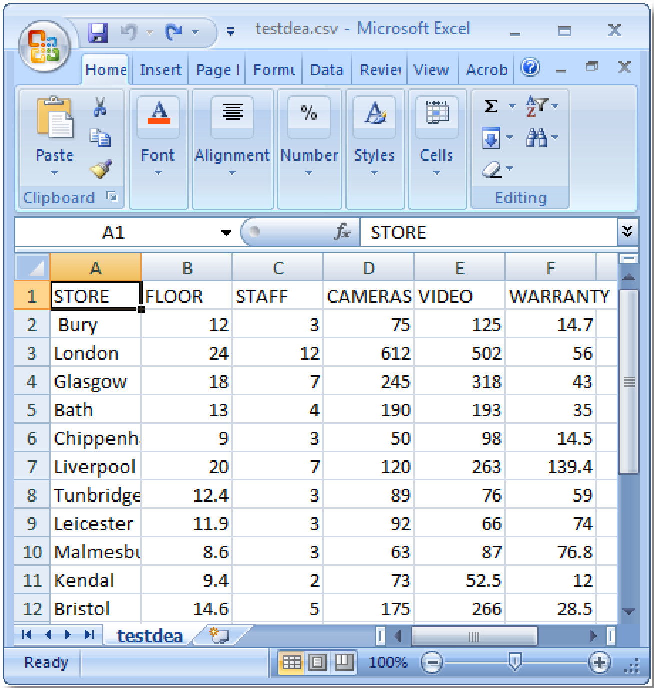
Part of the solution for the technical efficiency is the set of activity multipliers, $\lambda_{i,m}$ for the *i*th firm. The vector of *N* values, $\lambda_{i,m}$ will give the weights that produce the point on the efficient frontier for this firm. The firms with nonzero values of $\lambda_{i,m}$ – there will typically only be a few or one of them – will define the 'peers' for firm *i*. The listing of the peer firms can be requested by adding **; Peers** to the command. The first few observations for the sample above are shown below.

```
===============================================================================
Peers - By Firm
===============================================================================
Firm                     Orient.  TechEff  Peers
---------------------    -------  -------  ---------------------------------------
   1                     Inputs   1.00000   3  14 101
                         Outputs  1.00000   1  14 101
   2                     Inputs    .98446   4  71
                         Outputs   .92501   1  71
   3                     Inputs    .96243   3  71
                         Outputs   .88393   1  71
   4                     Inputs    .79469   4  14
                         Outputs   .73593   1  14
   5                     Inputs    .57426   4  71 118
                         Outputs   .44224   1  71
```

E65.5.2 Application

The following uses all the features of the routine save for the Malmquist TFP computation and the allocative efficiency routine. The sample data are in an *Excel* spreadsheet:

IMPORT ; File = … testdea.csv $
FRONTIER ; Lhs = cameras,video,warranty
; Rhs = floor,staff
; Alg = DEA ; CRS
; Peers
; Nbt = 50 $



	A	B	C	D	E	F
1	STORE	FLOOR	STAFF	CAMERAS	VIDEO	WARRANTY
2	Bury	12	3	75	125	14.7
3	London	24	12	612	502	56
4	Glasgow	18	7	245	318	43
5	Bath	13	4	190	193	35
6	Chippenh	9	3	50	98	14.5
7	Liverpool	20	7	120	263	139.4
8	Tunbridge	12.4	3	89	76	59
9	Leicester	11.9	3	92	66	74
10	Malmesbu	8.6	3	63	87	76.8
11	Kendal	9.4	2	73	52.5	12
12	Bristol	14.6	5	175	266	28.5

Figure E65.2 Sample Data for Data Envelopment Analysis

```
+----------------------------------------------------------------------------+
| Data Envelopment Analysis                                                  |
| Output Variables:  CAMERAS  VIDEO    WARRANTY                              |
| Input Variables:   FLOOR    STAFF                                          |
| Underlying Technology assumes CONSTANT Returns to Scale.                   |
+----------------------------------------------------------------------------+
| Estimated Efficiencies:     Mean     Std.Deviation     Minimum   Maximum  |
| Technical Efficiency      =======    =============     =======   =======  |
|      Input Oriented         .9132          .1270         .6387    1.0000  |
|     Output Oriented         .9132          .1270         .6387    1.0000  |
| Sample Size:               11 Observations.      11 Complete observations  |
| Efficiencies saved as variables DEAEFF_O, DEAEFF_I and DEAEFF_E            |
| Efficiencies saved as matrices  DEA_EFFO, DEA_EFFI and DEA_EFFE            |
| Incomplete observations are filled with zeros for efficiency values.       |
+----------------------------------------------------------------------------+
```

```
Estimated Efficiency Values for Individual Decision Making Units
================================================================================
Observation      | Input Oriented| Output Oriented|    Economic    | Allocative
Sample   Data    | Rank     Value| Rank      Value| Rank      Value| Rank   Value
================+===============+================+===============+=============
 Bury            |     9   .79126|      9    .79126|    0    .00000|    0  .00000
London           |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
Glasgow          |     7   .95227|      7    .95227|    0    .00000|    0  .00000
Bath             |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
Chippenham       |    11   .63869|     11    .63869|    0    .00000|    0  .00000
Liverpool        |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
Tunbridge        |     8   .90635|      8    .90635|    0    .00000|    0  .00000
Leicester        |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
Malmesbury       |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
Kendal           |    10   .75714|     10    .75714|    0    .00000|    0  .00000
Bristol          |     1  1.00000|      1   1.00000|    0    .00000|    0  .00000
================================================================================
Peers - By Firm
Firm                    Orient.  TechEff  Peers
---------------------   -------  -------  ------------------------------------
   1  Bury              Inputs    .79126   6  11
                        Outputs   .79126   6  11
   2 London             Inputs   1.00000   2
                        Outputs  1.00000   2
   3 Glasgow            Inputs    .95227   2   6  11
                        Outputs   .95227   2   6  11
   4 Bath               Inputs   1.00000   2   4   8   9
                        Outputs  1.00000   2   4
   5 Chippenham         Inputs    .63869   6  11
                        Outputs   .63869   6  11
   6 Liverpool          Inputs   1.00000   6  11
                        Outputs  1.00000   6
   7 Tunbridge          Inputs    .90635   4   8   9
                        Outputs   .90635   4   8   9
   8 Leicester          Inputs   1.00000   2   8   9
                        Outputs  1.00000   2   8
   9 Malmesbury         Inputs   1.00000   4   6   9
                        Outputs  1.00000   2   6   9
  10 Kendal             Inputs    .75714   2   4
                        Outputs   .75714   2   4
  11 Bristol            Inputs   1.00000   2  11
                        Outputs  1.00000   2  11
================================================================================
-----------------------------------------------------------------------------
Results of Bootstrap analysis of technical efficiency.       50 replications
-----------------------------------------------------------------------------
                  Technical  Estimated  Corrected  Standard   Confid. Limits
Observation_____  Efficiency    Bias    Tech.Eff.  Deviation  Lower    Upper
Bury                 .7913      .0404      .7509      .0374    .7931   .9074
London              1.0000      .0000     1.0000      .0000   1.0000  1.0000
Glasgow              .9523      .0353      .9170      .0143    .9570  1.0000
Bath                1.0000      .0000     1.0000      .0000   1.0000  1.0000
Chippenham           .6387      .0392      .5995      .0309    .6411   .7293
Liverpool           1.0000      .0000     1.0000      .0000   1.0000  1.0000
Tunbridge            .9064      .0630      .8433      .0333    .9138  1.0000
Leicester           1.0000      .0000     1.0000      .0000   1.0000  1.0000
Malmesbury          1.0000      .0000     1.0000      .0000   1.0000  1.0000
Kendal               .7571      .0389      .7183      .0551    .7614   .9307
Bristol             1.0000      .0000     1.0000      .0000   1.0000  1.0000
-----------------------------------------------------------------------------
```

E65.6 Comparing Efficiency Values and Rankings – SFA vs. DEA

In many settings, the efficiency ratings themselves are less interesting than the ranks of the observations. The WHO study used in numerous examples throughout this chapter is an example, in which the objective of the efficiency analysis was to rank the countries in terms of their measured efficiency. A perennial question in the efficiency analysis literature focuses on whether one obtains the same qualitative results with the two methodologies. We return to the WHO data to provide an illustration.

The data used are the country means of the output, *dale*, and two inputs, health expenditure, *hexp*, and education, *educ*. After the raw data are input, we use the following

```
SAMPLE      ; All $
REJECT      ; Small > 0 $
CREATE      ; dalebar  = Group Mean(dale, Str = country) $
CREATE      ; hexpbar = Group Mean(hexp, Str = country) $
CREATE      ; educbar = Group Mean(educ, Str = country) $
REJECT      ; year # 1997 $
CREATE      ; logdbar = Log(dalebar) $
CREATE      ; loghbar = Log(hexpbar) $
CREATE      ; logebar = Log(educbar) $
FRONTIER    ; Lhs = logdbar ; Rhs = one,loghbar,logebar ; Techeff  = effsfa $
FRONTIER    ; Lhs = dalebar ; Rhs = hexpbar,educbar ; Alg = DEA$
DSTAT       ; Rhs = effsfa,deaeff_i,deaeff_o ; Output = 2 $
PLOT        ; Lhs = effsfa ; Rhs = deaeff_i ; Grid
            ; Title = SFA Efficiencies vs. DEA Input Efficiencies $
PLOT        ; Lhs = effsfa ; Rhs = deaeff_o ; Limits=.4,1.1 ; Grid
            ; Title = SFA Efficiencies vs. DEA Output Efficiencies $
CREATE      ; sfarank = Rnk(effsfa) $
CREATE      ; dearanki = Rnk(deaeff_i) $
CREATE      ; dearanko= Rnk(deaeff_o) $
CALC        ; List ; Rkc(sfarank,dearanki)
                   ; Rkc(sfarank,dearanko)
                   ; Rkc(dearanki,dearanko) $
PLOT        ; Lhs = sfarank ; Rhs = dearanki
            ; Endpoints = 0,200 ; Limits = 0,200 ; Grid
            ; Title = Ranks of SFA Efficiencies vs. DEA Input Efficiencies $
PLOT        ; Lhs = sfarank ; Rhs = dearanko
            ; Endpoints = 0,200 ; Limits = 0,200 ; Grid
            ; Title = Ranks of SFA Efficiencies vs. DEA Output Efficiencies $
```

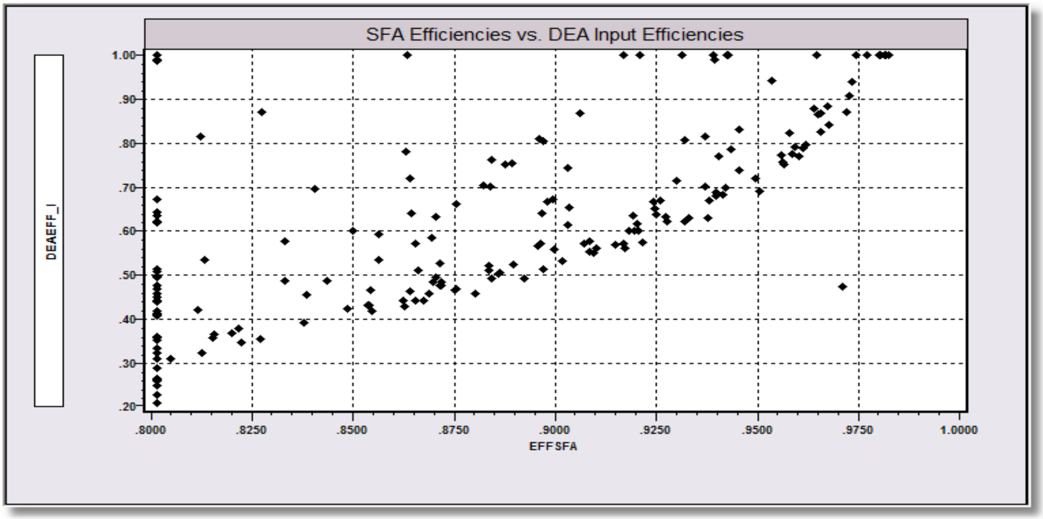
```
Normal exit:  11 iterations. Status=0, F=   -133.3834
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable              LOGDBAR
Log likelihood function       133.38343
Estimation based on N =    191, K =   5
Inf.Cr.AIC  =   -256.8 AIC/N =   -1.344
Variances: Sigma-squared(v)=     .00140
           Sigma-squared(u)=     .04405
           Sigma(v)        =     .03744
           Sigma(u)        =     .20989
Sigma = Sqr[(s^2(u)+s^2(v)]=     .21320
Gamma = sigma(u)^2/sigma^2 =     .96915
Var[u]/{Var[u]+Var[v]}     =     .91947
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          114.81039
Chi-sq=2*[LogL(SF)-LogL(LS)] =   37.146
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 LOGDBAR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    3.57889***      .04980    71.87  .0000     3.48129   3.67649
 LOGHBAR|     .06480***      .00824     7.86  .0000      .04864    .08096
 LOGEBAR|     .15292***      .01852     8.26  .0000      .11662    .18923
        |Variance parameters for compound error
  Lambda|    5.60534***     1.46657     3.82  .0001     2.73091   8.47977
   Sigma|     .21320***      .00101   211.97  .0000      .21123    .21517
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

+----------------------------------------------------------------------------+
| Data Envelopment Analysis                                                  |
| Output Variables:  DALEBAR                                                 |
| Input Variables:   HEXPBAR  EDUCBAR                                        |
| Underlying Technology assumes VARIABLE Returns to Scale.                   |
+----------------------------------------------------------------------------+
| Estimated Efficiencies:     Mean     Std.Deviation     Minimum   Maximum   |
| Technical Efficiency      =======    =============     =======   =======   |
|      Input Oriented         .6138         .2089          .2059    1.0000   |
|     Output Oriented         .8794         .1124          .5061    1.0000   |
| Sample Size:              191 Observations.     191 Complete observations  |
| Efficiencies saved as variables DEAEFF_O, DEAEFF_I and DEAEFF_E            |
| Efficiencies saved as matrices  DEA_EFFO, DEA_EFFI and DEA_EFFE            |
| Incomplete observations are filled with zeros for efficiency values.       |
+----------------------------------------------------------------------------+
```

DSTAT ; Rhs = effsfa,deaeff_i,deaeff_o ; Output = 2 $

```
Descriptive Statistics
--------+-----------------------------------------------------------------------
Variable|      Mean      Std.Dev.      Minimum      Maximum     Cases Missing
--------+-----------------------------------------------------------------------
  EFFSFA|    .882053     .059219      .801579      .982272       191       0
DEAEFF_I|    .613836     .208905      .205870          1.0       191       0
DEAEFF_O|    .879363     .112447      .506133          1.0       191       0
--------+-----------------------------------------------------------------------

--------+---------------------------
Cor.Mat.|  EFFSFA DEAEFF_I DEAEFF_O
--------+---------------------------
  EFFSFA| 1.00000   .70610   .75911
DEAEFF_I|  .70610  1.00000   .72559
DEAEFF_O|  .75911   .72559  1.00000
```



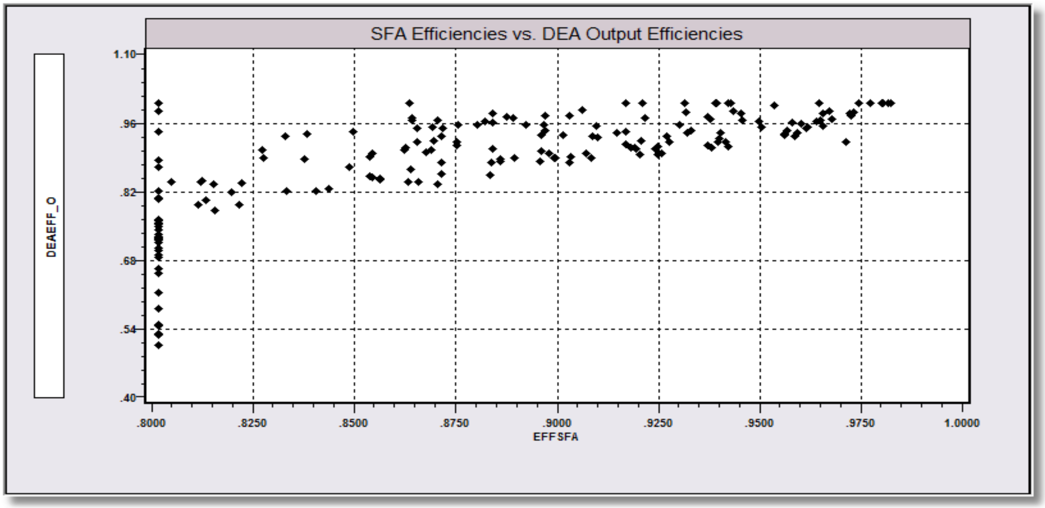
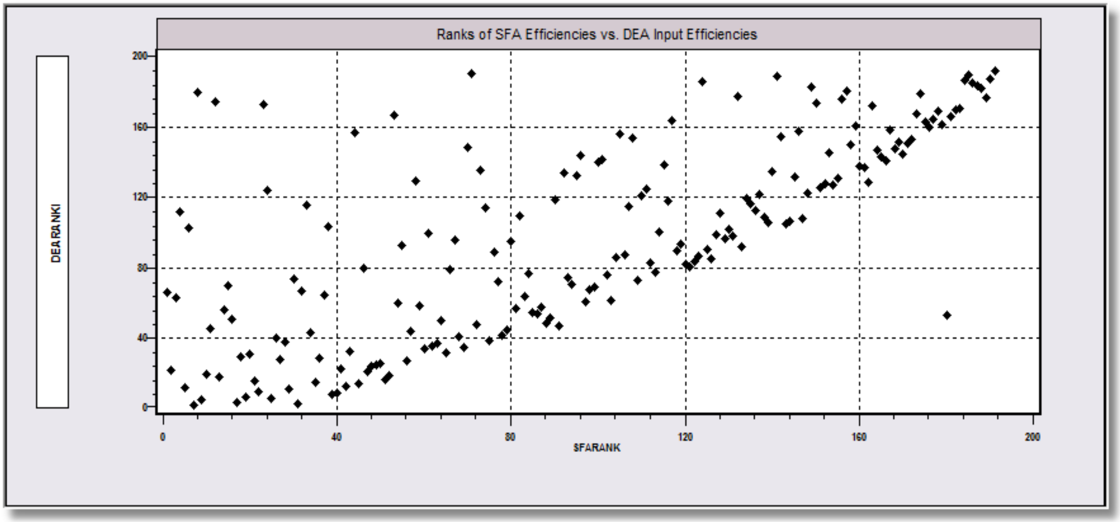


Figure E65.3 Plot of SFA Efficiency Values vs. DEA Values



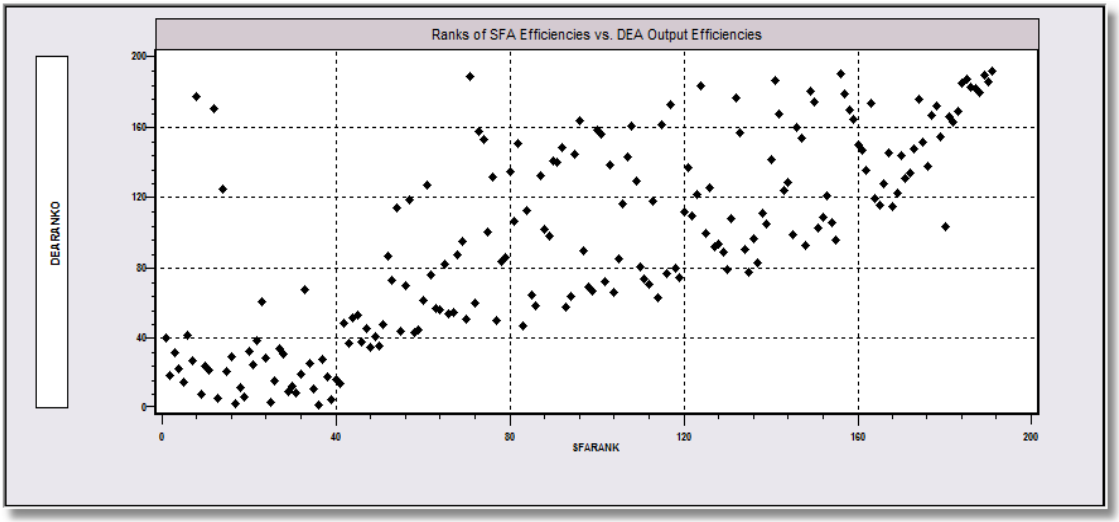


Figure E65.4 Plot of Ranks of SFA Efficiency Scores vs. Ranks of DEA Scores

E65.7 Malmquist Index of Total Factor Productivity

(Once again, the user is referred to the relevant literature, such as the numerous papers by Fare and Grosskopf) for background details. Fare's 1994 output based Malmquist productivity change may be written

$$M_{i,O}(t,t+1) = \sqrt{\frac{TE_i(t+1\,|\,t)\times TE_i(t+1\,|\,t+1)}{TE_i(t\,|\,t)\times TE_i(t\,|\,t+1)}}$$

where $TE(r|s)$ indicates the earlier defined output oriented technical efficiency index for firm i, using inputs $\mathbf{x}_{i,r}$ and producing outputs $\mathbf{y}_{i,r}$ relative to production (and input usage) for firms based in period s. This index is computed using the following program:

$$\mathbf{d}_L = \begin{bmatrix} \mathbf{0}_N \\ 0 \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} \mathbf{0}_N \\ 1 \end{bmatrix},\ \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\lambda} \\ \phi_{ir} \end{bmatrix},\ \mathbf{d}_U = \begin{bmatrix} \mathbf{1}_N \\ \infty \end{bmatrix}$$

$$\mathbf{b}_L = \begin{bmatrix} -\infty_K \\ \mathbf{0}_M \end{bmatrix},\ \mathbf{A} = \begin{bmatrix} \mathbf{X}_s' & \mathbf{0}_K \\ \mathbf{Y}_s' & \text{-}\mathbf{y}_{ir} \end{bmatrix},\ \mathbf{b}_U = \begin{bmatrix} \mathbf{x}_i \\ \infty_M \end{bmatrix}$$

This uses the constant returns to scale form. Also, since the period r output and input vectors for firm i will not appear in $\mathbf{Y}_s$ and $\mathbf{X}_s$ when r does not equal s, ϕ_{ir} need not be larger than one. Note that this requires solution of four linear programs for each firm in each period, so the total number of programs to solve will be $4\times N\times T$. Each is quite fast, so overall, the computations do not take long. In the sample of 247 firms and six periods, the nearly 6,000 programs, each involving 248 activities and six constraints, took about 10 seconds.

These computations are carried out for each firm in each period save the last one, and produce an $N\times T$ matrix of TFP values, one row for each firm, one column for each period. The TFP value for the last period is recorded as 1.0, though this is just a space filler.

To compute the Malmquist TFP indices, you will require a panel of data, at least two periods, for each of N firms. Unlike other panel data routines in *LIMDEP*, *this computation always requires a balanced panel*. Every firm must be observed in the same T periods. Also, this routine has no procedures for avoiding missing or invalid data such as zero values for inputs or outputs. The balanced panel must be 'clean' before computation begins. To request the computations, just add

; Pds = t, the fixed number of periods.

Nothing else need be changed. There is no bootstrap feature (**; Nbt = 0**); the computations assume constant returns to scale (**; CRS** is the default and cannot be changed) and no allocative efficiency (**; Rh2** is ignored).

Malmquist TFP Index Application

To illustrate the Malmquist computations, we reexamine the sample of 247 Spanish dairy farms observed for six years. The output is *milk* production. Inputs are *cows, land, labor* and *feed*.

```
FRONTIER    ; Lhs = milk
            ; Rhs = cows,land,labor,feed
            ; Alg = DEA ; Pds = 6
            ; List $
```

The following results are displayed. In addition, a matrix containing the full table, named *malmquist*, is created.

```
=============================================================================
Malmquist TFP Index for Productivity Change
Panel contained  247 firms each observed in   6 periods
Full Results saved as matrix MALMQIST
=============================================================================
Average results across firms, by period:
=============================================================================
Period:           1       2       3       4       5
TFP               1.0476  1.0233  1.0247  1.0298  1.0349
=============================================================================
Individual calculations by firm
(Only 8 periods can be displayed.  TFP for the final period is not computed.)
=============================================================================
Observation       1       2       3       4       5       6       7       8
Firm =     1      1.1301  1.1002   .9736  1.0291  1.0901  1.
Firm =     2      1.0528  1.0343  1.0212  1.0109  1.0416  1.
Firm =     3      1.0525  1.0383   .9477  1.0465  1.0395  1.
Firm =     4      1.1418  1.0129  1.0079   .9829  1.0476  1.
Firm =     5      1.1192  1.0240  1.0082  1.0245  1.0641  1.
Firm =     6       .9871  1.0073   .9785  1.0322  1.0464  1.
Firm =     7       .9851  1.1484  1.1599   .8054  1.1110  1.
Firm =     8      1.0746   .9796   .9636  1.0671   .9753  1.
Firm =     9       .8977  1.1496   .9818  1.0500   .9867  1.
Firm =    10      1.0105  1.1507   .9751  1.0055  1.0469  1.
Firm =    11      1.1276   .9867   .9636  1.0826   .9873  1.
Firm =    12      1.0310  1.1020   .9822  1.0438   .9914  1.
Firm =    13      1.0549  1.1263   .9221  1.0723  1.1945  1.
Firm =    14       .9408  1.0740   .9938   .9739  1.0336  1.
Firm =    15       .8952   .7156  1.5056   .8614   .9204  1.
(Rows 66 – 247 omitted).
```

E66: MAXIMIZE – Nonlinear Optimization

E66.1 Introduction

Chapters E14 and E21 presented programs for computing nonlinear least squares, nonlinear two and three stage least squares and GMM estimators. The discussion continued in Chapter E25 with a discussion of **NLSUR** for estimation of nonlinear systems of equations. The **NLSQ** and **NLSUR** procedures are parts of a package that allows you to define your own minimization or maximization problem. This will allow you to set up your own maximum likelihood problems for models which are not included in the program already.

The estimation criteria defined in Chapters E14 and E21 are:

Nonlinear Least Squares	Minimize$_\beta$ $\sum_{i=1}^{N} w_i \times [y_i - f(\boldsymbol{\beta},\mathbf{x}_i)]^2 = \sum_{i=1}^{N} w_i\, \varepsilon_i^2$
Nonlinear IV	Minimize$_\beta$ $\boldsymbol{\varepsilon}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}$ where $\varepsilon_i = y_i - \mathrm{f}(\boldsymbol{\beta},\mathbf{x}_i)$
GMM	Minimize$_\beta$ $\mathrm{M}(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$
GMME	Minimize$_\beta$ $q = \overline{\mathbf{m}}'\mathbf{W}^{-1}\overline{\mathbf{m}}$, $\overline{\mathbf{m}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{m}(\mathbf{x}_i,\boldsymbol{\beta})$, $\mathbf{W} = \text{a weighting matrix}$

MAXIMIZE/MINIMIZE adds to these a general program that allows you to optimize:

Single Function of a Set of Parameters	Minimize or maximize$_\beta$ $\mathrm{F}(\boldsymbol{\beta})$
Sum of Terms	Minimize or maximize$_\beta$ $\sum_{i=1}^{N} w_i \times F(\boldsymbol{\beta},\mathbf{x}_i)$

LIMDEP's **MINIMIZE/MAXIMIZE** procedure will allow you to set up your own log likelihood or method of moments criterion functions. Most of the necessary information about **MINIMIZE/MAXIMIZE** was given in Chapters E14 and E21. Users will find it useful to review these chapters with the discussion below.

NOTE: This program may be used to estimate up to 150 parameters.

NOTE: Use **MAXIMIZE** to create new models that are not in the menu of available models in the program. An example appears in Section E66.8.3.

E66.2 The MINIMIZE/MAXIMIZE Commands

For convenience, we will assume at this point that you wish to **MAXIMIZE** a function. (If appropriate, change the command to **MINIMIZE**.) The **MAXIMIZE** command is the same as **NLSQ** discussed in Chapter E14,

NLSQ ; Lhs = y
; Fcn = ...
; Labels = ...
; Start = starting values $

with two exceptions. To maximize a general function, the command is **MAXIMIZE** and there is no Lhs variable. The basic command is, then,

MAXIMIZE ; Labels = list of labels for parameters being computed
; Fcn = function definition
; Start = list of starting values $

The basic format of the command, as shown above, is used to maximize a sum of terms. The function definition defines a function that is summed over the sample observations. Here is an example that computes maximum likelihood estimates of the parameters of a probit model using 500 artificially generated observations that conform exactly to the assumptions of the model.

CALC ; Ran(12345) $
SAMPLE ; 1-500 $
CREATE ; x = Rnu(-.5,.5) ; y = (.2 - .2*x + Rnn(0,1)) > 0 $
CREATE ; q = 2*y - 1 $
MAXIMIZE ; Fcn = Log(Phi(q * (b0 + b1*x))) ; Labels = b0,b1 ; Start = 0,0
; Output = 3 $
PROBIT ; Lhs = y ; Rhs = one,x ; Output = 3 $

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator= 100
Start values:    .00000D+00   .00000D+00
1st derivs.    -.71810D+02   .11154D+02
Parameters:      .00000D+00   .00000D+00
Itr  1 F=  .3466D+03 gtHg=  .7267D+02 chg.F=  .3466D+03 max|db|=  .7181D+08
1st derivs.      .13095D+01   .87780D+01
Parameters:      .23105D+00  -.35889D-01
Itr  2 F=  .3381D+03 gtHg=  .8875D+01 chg.F=  .8474D+01 max|db|=  .2446D+03
1st derivs.    -.10851D+02   .16194D+01
Parameters:      .18685D+00  -.33218D+00
Itr  3 F=  .3368D+03 gtHg=  .1097D+02 chg.F=  .1330D+01 max|db|=  .5807D+02
1st derivs.    -.10851D+02   .16194D+01
Parameters:      .18685D+00  -.33218D+00
Itr  1 F=  .3368D+03 gtHg=  .1097D+02 chg.F=  .3368D+03 max|db|=  .5807D+02
1st derivs.      .18603D+00   .13083D+01
```

```
Parameters:      .22217D+00  -.33745D+00
Itr  2 F=  .3366D+03 gtHg=  .1323D+01 chg.F=  .1959D+00 max|db|=  .3946D+01
1st derivs.     -.17467D-01   .49557D-03
Parameters:      .22067D+00  -.39051D+00
Itr  3 F=  .3365D+03 gtHg=  .9902D-03 chg.F=  .3483D-01 max|db|=  .2532D-03
1st derivs.     -.32705D-08   .44429D-05
Parameters:      .22073D+00  -.39052D+00
Itr  4 F=  .3365D+03 gtHg=  .8938D-06 chg.F=  .4916D-06 max|db|=  .4604D-06
                                                           * Converged
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   4 iterations. Status=0, F=    336.5390
Function=  .34657359028D+03, at entry,  .33653903385D+03 at exit

-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function      -336.53903
Estimation based on N =    500, K =   2
Inf.Cr.AIC  =    677.1 AIC/N =    1.354
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|     .22073***      .05678     3.89  .0001      .10945     .33201
      B1|    -.39052*        .20368    -1.92  .0552     -.78972     .00869
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Nonlinear Estimation of Model Parameters
Method=NEWTON; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator= 100
Start values:    .58680D+00  -.15056D+00
1st derivs.      .11038D+03   .41954D+01
Parameters:      .58680D+00  -.15056D+00
Itr  1 F=  .3574D+03 gtHg=  .6515D+01 chg.F=  .3574D+03 max|db|=  .1614D+01
1st derivs.     -.28665D+01  -.30959D-01
Parameters:      .21147D+00  -.39358D+00
Itr  2 F=  .3366D+03 gtHg=  .1632D+00 chg.F=  .2085D+02 max|db|=  .4378D-01
1st derivs.     -.12484D-02   .13273D-03
Parameters:      .22073D+00  -.39051D+00
Itr  3 F=  .3365D+03 gtHg=  .7438D-04 chg.F=  .1332D-01 max|db|=  .1787D-04
1st derivs.     -.37516D-09   .10573D-09
Parameters:      .22073D+00  -.39052D+00
Itr  4 F=  .3365D+03 gtHg=  .2927D-10 chg.F=  .2766D-08 max|db|=  .1036D-10
                                                           * Converged
Normal exit:   4 iterations. Status=0, F=    336.5390
Function=  .35739976440D+03, at entry,  .33653903385D+03 at exit
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                       Y
Log likelihood function      -336.53903
Restricted log likelihood    -338.42927
Chi squared [   1 d.f.]         3.78048
Significance level               .05185
McFadden Pseudo R-squared      .0055853
Estimation based on N =    500, K =   2
Inf.Cr.AIC  =    677.1 AIC/N =    1.354
Model estimated: Aug 24, 2011, 12:17:51
Hosmer-Lemeshow chi-squared =  13.43761
P-value=  .09765 with deg.fr. =       8
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|     .22073***      .05680     3.89  .0001      .10941    .33205
       X|    -.39052*        .20125    -1.94  .0523     -.78496    .00393
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E66.2.1 Function Definitions

The following describes the various components of the function definition. The next section describes a very important variation of this specification, the use of subfunctions. Users should be sure to read all of both these sections before using this program.

Labels for Parameters

The labels definition is optional. If you do not provide labels, the defaults are **b1**, **b2**, ..., **bk**. The number of parameters in the model, *k*, is the number of starting values you provide. Thus, for example, a linear regression could be requested with

MINIMIZE ; Fcn = (y - b1 - b2*x) ^ 2 ; Start = 0,0 $

Because there are a variety of named entities which can appear in the function, you should use the

; Labels = list of labels

part of the command to identify which of them are the parameters being estimated. You must then use these labels in the function you specify. Labels may be anything you like, up to eight characters.

WARNING: Use new names! Do not use program names that are in use otherwise, such as *s*, *rho*, *sigma*, *b*, etc., and the names of existing scalars or matrices. Such labels would be accepted when your command is translated, because you are free to use these entities in your function definition to supply specific values. But, later, when *LIMDEP* scans your expression to see what you have specified, it checks all other tables first, and your label list last. For example, if you use **s** as a label, and this command is the first model command that you have given, **s** will simply be taken as the as yet undefined result of a regression. The actual value would, in fact, always be fixed at 0. An attempt is made to prevent you from doing this at the time your function definition is translated. For example, here is what happens if we try to use **s** instead of **b0** in the probit model estimator above

MAXIMIZE ; Fcn = Log(Phi(q * (s + b1*x))) ; Labels = s,b1 ; Start = 0,0 $

```
Conflict: param. and scalar have the same name: S.
```

For large problems, you may use a shortcut for the labels definition,

; Labels = number_label

produces ‘number’ sequentially numbered repetitions of the label. For example, **5_b** gives **b1,b2,b3,b4,b5**. The number may be a literal value or a scalar. With this device, you can make your model command independent of the size of the model, and you can accommodate a model of any size. For example:

```
NAMELIST    ; xa = ... (up to 100 names)
            ; xb = ... (up to 100 names) $
CALC        ; ka = Col(xa) ; kb = Col(xb) $
MATRIX      ; ca = Init(ka,1,0.) ; cb = Init(kb,1,0.0) $
MINIMIZE    ; Start = ca,cb, ... any other parameters
            ; Labels = ka_ba, kb_bb, any other labels
            ; Fcn = Index = ba1’xa  +  bb1’xb | ... the rest of the function $
```

This template could be used for a model of any size. Only the namelists would have to be changed from one specification to another.

LIMDEP will ensure that there is a correspondence between your labels and your starting values. However, it is not possible for the program to ensure that you have used all of the parameters in your function specification. If you define a parameter, but you do not use it in your function definition, then one of two things will occur. Either the iterations will never converge and they will exit on maximum iterations, with one of the parameters not changing from its initial value, or what appears to be convergence will be reached, but the estimated covariance matrix of the estimated parameters will be singular, as it will contain a row and column of zeros corresponding to the unused parameter.

Here is an example. Note that the defined model parameter *c*3 does not appear in the function.

MINIMIZE ; Fcn = (y - c0 - c1*x1 - c2*x2)^2
; Start = 0,0,0,0
; Labels = c0,c1,c2,c3 $

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .1000D-05 chg.F   .0000D+00 max|dB|   .0000D+00
Nodes for quadrature: Laguerre=40;Hermite=20.
Replications for GHK simulator= 100
Start values:   .00000D+00   .00000D+00   .00000D+00   .00000D+00
1st derivs.     .13962D+01   .26612D+01  -.31547D+01   .00000D+00
Parameters:     .00000D+00   .00000D+00   .00000D+00   .00000D+00
Itr  1 F=  .5325D+02 gtHg=  .4357D+01 chg.F=  .5325D+02 max|db|=  .3155D+07
1st derivs.    -.47552D+00   .17547D+00  -.62430D-01   .00000D+00
Parameters:    -.20089D-01  -.38291D-01   .45391D-01   .00000D+00
Itr  2 F=  .5311D+02 gtHg=  .5107D+00 chg.F=  .1366D+00 max|db|=  .2367D+02
1st derivs.     .13412D-01   .25228D-01  -.31252D-01   .00000D+00
Parameters:    -.15250D-01  -.40076D-01   .46027D-01   .00000D+00
Itr  3 F=  .5311D+02 gtHg=  .4234D-01 chg.F=  .1327D-02 max|db|=  .8795D+00
1st derivs.     .13412D-01   .25228D-01  -.31252D-01   .00000D+00
Parameters:    -.15250D-01  -.40076D-01   .46027D-01   .00000D+00
Itr  1 F=  .5311D+02 gtHg=  .4234D-01 chg.F=  .5311D+02 max|db|=  .8795D+00
1st derivs.    -.45648D-02   .19336D-02  -.39819D-03   .00000D+00
Parameters:    -.15443D-01  -.40439D-01   .46476D-01   .00000D+00
Itr  2 F=  .5311D+02 gtHg=  .4973D-02 chg.F=  .1290D-04 max|db|=  .2836D+00
1st derivs.     .31087D-05   .91316D-05   .87055D-05   .00000D+00
Parameters:    -.15398D-01  -.40463D-01   .46485D-01   .00000D+00
Itr  3 F=  .5311D+02 gtHg=  .1299D-04 chg.F=  .1271D-06 max|db|=  .2261D-03
1st derivs.     .67446D-12   .86003D-11  -.18233D-10   .00000D+00
Parameters:    -.15398D-01  -.40463D-01   .46485D-01   .00000D+00
Itr  4 F=  .5311D+02 gtHg=  .2425D-11 chg.F=  .8811D-12 max|db|=  .5445D-11
                        * Converged
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit from iterations. Exit status=0.
Function=  .53247681205D+02, at entry,  .53109768475D+02 at exit
Models - estimated variance matrix of estimates is singular
Current estimated covariance matrix for slopes is singular.
```

Algebraic Form of the Function

The **; Fcn** specification is written using the rules and operators of algebra (+, -, *, /), ^ (for raise to the power), and @ (for the Box-Cox transformation). The usual rules are observed; ^ and @ are computed first, then * and /, and finally + and -. Two additional operators which have the same precedence as multiplication are ! for maximum (5 ! 6 = 6) and ~ for minimum (5 ~ 6 = 5). Parentheses may be used freely to force the order of evaluation of expressions. Use as many levels of parentheses as required. Entities which may appear in the specification include:

- numbers,
- variable names,
- namelists,
- any existing scalars,
- matrix elements,
- your parameters, using your labels.

To use a subscripted matrix element, enclose the subscript in curled brackets, **{ }**, not parentheses. I.e., **gamma(1,1)** will confuse the compiler, use **gamma{1,1}**.

NOTE: This construction, with curled brackets, is specific to the function definition part of the **NLSQ**, **NLSUR**, **MAXIMIZE**, **MINIMIZE** and **GMME** commands. Elsewhere, such as in **CALC** and **CREATE**, matrix subscripts are indicated with ordinary parentheses.

The function is evaluated by 'looping' through your current sample, computing the function at each observation, and summing the terms. Let **Z**(*i*) denote all the variables in your data set where '*i*' denotes a specific observation. It is assumed that some variables appear in your function definition, so the function is computed by summing all observations. If no variables actually appear in the function, then the same function will simply be summed *N* times. Thus, in the preceding probit example, the function evaluated is

$$F = \sum_{i=1}^{N} \text{Log}\ (\ \text{Phi}\ (\ (2y_i - 1) * (\beta_0 + \beta_1 x_i))) = \sum_{i=1}^{N} g[\mathbf{Z}(i)].$$

An example that appears below is a four dimensional Rosenbrock function,

$$F(\mathbf{c}) = (c_1+10c_2)^2 + 5(c_3 - c_4)^2 + (c_2 - 2c_3)^4 + 10(c_1 - c_4)^4.$$

The function definition for this minimization problem would be

; Fcn = (c1+10*c2)^2 + 5*(c3-c4)^2 + (c2-2*c3)^4 + 10*(c1-c4)^4

Since this function does not involve any variables, the function value each time this is calculated would be just *N* times the value shown in the actual function. Since this would be a waste of time and effort, one would normally precede this kind of optimization problem with

SAMPLE ; 1 $

so that it would be evaluated only once.

Functions that May Appear in the Definition

The following functions may be used in your function definition

Abs(*z*)	= absolute value, $\lvert z \rvert$
Atn(*z*)	= arctangent, atan(*z*)
Cos(*z*)	= cosine, cos(*z*)
Exp(*z*)	= exponent, exp(*z*)
Gma(*z*)	= gamma, $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$
Inp(*z*)	= inverse normal probability = $\Phi^{-1}(z)$, $0 < z < 1$
Lgd(*z*)	= logit density = Lgp*(1-Lgp)
Lgm(*z*)	= log of gamma
Lgt(*z*)	= logit = log(z/(1-z)), $0 < z < 1$
Lgp(*z*)	= logit probability = Exp(z)/(1+Exp(z)) = Prob($Z \le z$)
Lmm(*z*)	= -N01(z)/Phi(z) = E[$z \mid z \le 0$] for $z \sim$ N[0,1]
Lmp(*z*)	= N01(z)/Phi(-z) = E[$z \mid z \ge 0$] for $z \sim$ N[0,1], Lmp(-z)=$\phi(z)/\Phi(z)$
Log(*z*)	= natural logarithm, log(z)
N01(*z*)	= standard normal density, $\phi(z)$
Phi(*z*)	= standard normal CDF, $\Phi(z)$
Psi(*z*)	= log derivative of Gma, $\Psi(z) = \Gamma'(z)/\Gamma(z)$, (digamma function)
Psp(*z*)	= $\Psi'(z) = \Gamma''(z)/\Gamma(z) - \Psi^2(z)$, (trigamma function)
Sgn(*z*)	= signum(*z*) = -1,0,+1 if z <, =, > 0
Sin(*z*)	= sine, sin(*z*)
Tvm(*z*)	= 1 - Lmm $\times$ (*z* + Lmm) = Var[$z \mid z < 0$] for $z \sim$ N[0,1]
Tvp(*z*)	= 1 - Lmp $\times$ (*z* + Lmp) = Var[$z \mid z < 0$] for $z \sim$ N[0,1]
Bds(*z*,*a*,*c*)	= incomplete beta function; (Bds(0,*a*,*c*) = 0, Bds(1,*a*,*c*) = 1)
Gmp(*z*,*p*,*a*)	= incomplete gamma integral, normalized to the probability
Bvn(*z*1,*z*2,ρ)	= bivariate normal CDF
Bvd(*z*1,*z*2,ρ)	= bivariate normal density
Min(*z*1,*z*2)	= minimum of *z*1 and *z*2
Max(*z*1,*z*2)	= maximum of *z*1 and *z*2
Ash(*z*)	= hyperbolic arc sin(*z*) = $\log(z + (1 + z^2)^{1/2})$
As1(*z*)	= derivative of Ash(*z*) = $(1 + z^2)^{-1/2}$
Ach(*z*)	= hyperbolic arc cos(*z*) = $\log(z + (z^2 - 1))$
Ac1(*z*)	= derivative of Ach(*z*) = $(z^2 - 1)^{-1/2}$
Ath(*z*)	= hyperbolic arc tan(*z*) = $.5\log((1 + z)/(1 - z))$
At1(*z*)	= derivative of Ath(*z*) = $(1 - z^2)^{-1}$
Hsn(*z*)	= hyperbolic sin(*z*) = .5(exp(2*z*)-1)/exp(*z*)
Hs1(*z*)	= derivative of Hsn(*z*) = Hcs(*z*)
Hcs(*z*)	= hyperbolic cos(*z*) = .5(exp(2*z*)+1)/exp(*z*)
Hc1(*z*)	= derivative of Hcs(*z*) = Hsn(*z*)
Htn(*z*)	= hyperbolic tan(*z*) = Hsn(*z*)/Hcs(*z*)
Ht1(*z*)	= derivative of Htn(*z*) = $1/\mathrm{Hcs}^2(z)$

The incomplete beta function is

$$\text{Bds}(z,a,c) \;=\; [\Gamma(a)\Gamma(c)/\Gamma(a+c)] \int_0^z t^{a-1}(1-t)^{b-1}dt \text{ for } 0 < z < 1.$$

The normalized incomplete gamma function is

$$\text{Gmp}(z,p,a) \;=\; [a^p / \Gamma(p)] \int_0^z t^{P-1}\, e^{-at}\, dt.$$

Note that this returns a probability; $\text{Lim}_{z\to 0}$ Gmp(z,p,a) = 0, $\text{Lim}_{z\to\infty}$ Gmp(z,p,a) = 1, 0 < Gmp < 1, ∂Gmp/∂z > 0. To get the unnormalized gamma integral, you may use the construction

$$\text{Gma}(p) \;/\; a\text{\^{}}p \;\mathbf{*}\; \text{Gmp}(z,p,a) \;=\; \int_0^z t^{p-1}\, e^{-at}\, dt.$$

Do note, however, that this integral can become very large. This function is a generalization of p! for noninteger p. Some particular values to note, Gmp(z,p,a) = 0 if $z \le 0$; Gmp(z,p,a) = 1 if $p \le 0$, and Gmp(z,p,a) = 0 if $a \le 0$ and, finally, Gma(.5) = $\sqrt{\pi}$.

In the beta, gamma and bivariate normal functions, if any of the parameters separated by commas are expressions, it is necessary to enclose them in parentheses. E.g., use **Bvn((1+x'b),z,r)**, not **Bvn(1+x'b,z,r)**. The list may contain variables, labels, scalars, and expressions contained in parentheses. Functions may be nested to any depth and expressions may appear as arguments in the functions, as in

Log (Phi (a1 + a2 * (x/y)^2)).

This would be a valid expression and would evaluate exactly as given.

Linear Functions and Dot Products

Many expressions in econometric models will involve dot products of parameters and variables. For example, a model built as an extension of a probit model will likely involve an expression of the form **Phi(b'x)**. Dot products may appear in exactly this form in your function definitions. Typically, the '**x**' would be a namelist. To use the parameter vector, use the first name in your labels list. For example, in

```
NAMELIST    ; x = one,x1,z,p  $
MAXIMIZE    ; ...  ; Labels = b0,b1,b2,b3
                   ; Fcn = ... Phi(b0'x) $
```

the term **b0'x** is evaluated as $b0\times one + b1\times x1 + b2\times z + b3\times p$. Once again, in a dot product, the sum is evaluated from left to right using your list of labels in the order in which they appear in **; Labels = list**.

NOTE: If the namelist and the labels list do not have the same number of elements, then the dot product is simply evaluated out to the shorter of the two lists. In the example, if there were additional names in **x**, they would not change **b0'x** because starting at *b*0, there are only four parameters.

NOTE: This replaces the function Dot[.] used in earlier versions of *LIMDEP*. The Dot[.] function is retained for backwards compatibility, but you will find it easier to use the more natural syntax. Also, the operation described above does allow a bit more flexibility. For completeness, we note the counterparts to the constructions described above are **Dot[x]** = **b0'x** and **Dot[b3,second]** = **b3'second**. You may use either form.

Suppose you want to pick up just a few of the parameters in a dot product. For example, suppose your parameters are **; Labels = b1,b2,b3,b4,b5,b6,b7** and as part of your function, you want **b3*x14 + b4*xyz + b5*wvs**. You could first define the namelist for the dot function, with, say, **NAMELIST ; second = x14,xyz,wvs $**. Then, to obtain that function, just begin the dot product with *b*3 instead of *b*1. Thus, **b3'second** evaluates exactly to the sum given above.

It is also possible to skip over parameters in dot products, by putting columns of zeros in your namelists. This may be convenient in specifying your function, especially if it involves many parameters. For example, using the list above, you could obtain $b2 \times x14 + b5 \times xyz$

```
CREATE     ; zero = 0 $
NAMELIST   ; second = x14,zero,zero,xyz $
MAXIMIZE   ; ... b2'second ...
```

Dot products need not be only a mix of variables and parameters. They may also include vectors (matrices) that do not appear elsewhere in the function, and they may be products of variables or parameters. When you are specifying your functions, there are several ways you can shorten your commands by making use of the dot product notation, and using lists. The following constructions can all be used in specifying your functions: Let

a,d = the names of any vectors in your matrix work area
x,y = the names of any namelists
cj = any of the labels in your **; Labels = ...** specification

Then, any of the following can appear in your function

a'a = inner product of the vector
a'd = dot product of two vectors
a'x = linear combination of variables, at the *i*th observation
x'y = sum of cross products of the variables, at *i*th observation
x'x = sum of squares
cj'a = product of vector elements and parameters
cj'x = the familiar product of coefficients and variables.

This product can be computed beginning with any of the parameters in the list. For example, consider fitting a probit model:

```
MAXIMIZE    ; Labels = a1,a2,a3
            ; Start = 0,0,0
            ; Fcn = Log(Phi((2*y-1)*(a1+a2*x1+a3*x2))) $
```

Alternatively, if you define a namelist with

```
NAMELIST    ; xa = one,x1,x2  ; xb = x1,x2 $
```

Then, **; Fcn = Log(Phi((2*y-1) * a1'xa))**

is the same as **; Fcn = Log(Phi((2*y-1) * (a1 + xb'a2)))**.

Bilinear and Quadratic Forms

Bilinear and quadratic forms may also appear in function definitions. Suppose that *c* and *d* indicate elements of the parameter vector, which point to specific parts of the vector, and **z** is a namelist and **A** is a matrix. The following forms may appear in your function definition

(bilinear) **c'[z]d** $= \Sigma_j c_j d_j z_j$,

c'[z]c $= \Sigma_j c_j^2 z_j$

(quadratic) **c'[A]c** $= \Sigma_j \Sigma_l c_j c_l A_{jl}$

In the quadratic form, **A** may denote a namelist. In this case, you must indicate how many rows and columns are to appear in the matrix. Thus, suppose the namelist contains 12 variables. These could be arranged in a 2×6, 3×4, 4×3, or other matrix. To indicate how many rows the matrix has, you append the number of rows in the name between backslashes. For example, if

```
NAMELIST    ; x = x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12 $
```

then, **c**'[**X** \ 4\]**d** is the bilinear form $\begin{pmatrix} c1 & c2 & c3 & c4 \end{pmatrix} \begin{bmatrix} x1 & x2 & x3 \\ x4 & x5 & x6 \\ x7 & x8 & x9 \\ x10 & x11 & x12 \end{bmatrix} \begin{pmatrix} d1 \\ d2 \\ d3 \end{pmatrix}$.

Each observation is inserted in turn into the matrix in order to set up the computation. The function evaluation then involves summing (possibly functions of) the quadratic forms.

E66.2.2 Gauss-Hermite and Gauss-Laguerre Quadrature in Functions

The function optimization programs such as **MAXIMIZE** or **MINIMIZE** may use functions that contain integrals of the form

$$F(\boldsymbol{\beta}) \;=\; \int_{-\infty}^{\infty} \exp(-v^2)G(\boldsymbol{\beta},v)dv$$

by using Gauss-Hermite quadrature. This is a very accurate approximation which is computed using

$$F(\boldsymbol{\beta}) \;\approx\; \sum_{h=1}^{H} w_h G(\boldsymbol{\beta},z_h)$$

where H is the number of points for the quadrature, w_h is the weight and z_h is the node at point h of the quadrature. You set the number of points, H for the quadrature. The $G(.)$ function is unrestricted – it can be any function that is allowable in **MINIMIZE/MAXIMIZE**. The variable of the integration, v, may or may not actually appear in the function. You can also include functions of the form

$$F(\boldsymbol{\beta}) \;=\; \int_{0}^{\infty} \exp(-v\;)G(\boldsymbol{\beta},v)dv\,.$$

(Notice that the exponent is exp(-v) rather than exp(-v^2), and the range of integration is from 0 to +∞ rather than from -∞ to ∞. Integrals of this form are accurately approximated using Gauss-Laguerre integration, rather than Gauss-Hermite integration.) Commands that use quadrature are of the form

MAXIMIZE **; Fcn = name = Ntg(the function to be integrated) |**
the rest of the function, which will probably involve 'name'
; Hrq = the name of the variable over which integration is done
for Hermite integration
or **; Glq = the name** of the variable over which integration is done
for Gauss-Laguerre integration

The accuracy of the quadrature is directly a function of the number of quadrature points, the more the better. You may set the number of points with

; Hpt = one of 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 20, 32, 48, 96 for the Hermite quadrature
and **; Lpt** = one of 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 40, 68 for the Laguerre quadrature

As noted, more points are better than less. However, the amount of computation varies linearly with the number of quadrature points, so if time is a consideration, you may wish to choose a lower number. The default values for the numbers of quadrature points are 20 for Hermite and 40 for Laguerre.

To use one of these integrals in your function to be maximized, you must set up the operation as follows: The command will include

; Hrq = **the name** of the variable over which the integration is done for integration by Hermite quadrature

or **; Glq =** **the name** of the variable over which the integration is done for integration by Laguerre quadrature

and **; Start =** parameter values, as usual
; Labels = labels for parameters in the model, as usual
; other options $

Note the following requirements:

- You can have more than one integral in the final function, but each must be a named subfunction. If you specify 'Ntg(...)' within a function definition, an error will occur during compilation claiming that you have an unidentified symbol.
- Integrals should not be functions of other integrals. The results will be unpredictable, but almost certainly incorrect.
- You may have only one kind of integral in your function definition. Each Hrq, Glq, (or Sim, see below) which appears in a command overrides previous ones.

Two examples follow. Note that neither of these are 'good models,' and unless the data actually do satisfy the assumptions of the model, estimation of these will not produce very appealing results. For the first one, in particular, for a cross section formulation, without multiplying *v* by *x*, the variance term diverges; it is not identified.) The examples are intended only to illustrate use of the tools.

Heterogeneity in a Probit Model

Consider a probit model in which there is normally distributed, unobserved individual heterogeneity which multiplies one of the variables in the model,

$$y = 0 \text{ or } 1,$$

$$\text{Prob}[\, y = 1 \mid v\,] = \Phi(\boldsymbol{\beta}'\mathbf{z} + \theta\, x\, v) \text{ where } v \text{ is standard normally distributed.}$$

(A nonunitary standard deviation of v would be absorbed into the free parameter θ.) The probability that enters the log likelihood is Prob[$y = j$] = E_v [Prob[$y = j \mid v$]], j = 0,1. The expectation is exactly equal to

$$\text{Prob}\,[\, y = j\,] = \int_{-\infty}^{\infty} (1/\sqrt{2\pi})\exp(-v^2/2)\Phi[(2j-1)(\boldsymbol{\beta}'\mathbf{z}+\theta xv)]dv = P(y).$$

In the integral, let $u = v/\sqrt{2}$, so $v = u\sqrt{2}$ and the Jacobian is $dv/du = \sqrt{2}$. Make the change of variable in the integral, to produce

$$\text{Prob}\,[\, y = j\,] = \int_{-\infty}^{\infty} (1/\sqrt{\pi})\exp(-u^2)\Phi[(2j-1)(\boldsymbol{\beta}'\mathbf{z}+\sqrt{2}\theta xu)]du = P(y).$$

This is now exactly in the form noted earlier for Hermite quadrature. (It can be simplified a bit more by defining $\gamma = \sqrt{2}\,\theta$.) The following commands simulate and estimate this model: (The command uses a subfunction, which is described in the next section.)

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-200 $
CREATE      ; z1 = Rnn(0,1) ; z2 = Rnn(0,1) ; v = Rnn(0,1) ; x = Rnu(-.5,.5) $
CREATE      ; y = ( .2 + z1 + z2 + v*x + Rnn(0,1) ) > 0 ; q = 2*y - 1 $
NAMELIST    ; z = one,z1,z2$
CALC        ; kz = Col(z) $
PROBIT      ; Lhs = y ; Rhs = z $
MAXIMIZE    ; Fcn = Prob = Ntg(1/Sqr(pi) * Phi(q*(b1'z + t*u*x))) | Log(Prob)
            ; Start = b,.1
            ; Labels = kz_b,t
            ; Hrq = u ; Hpt = 20 ; Output = 3 $
```

The following output results. (The probit results for the starting values are omitted.)

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=20.
Replications for GHK simulator=  10
Start values:   .36125D+00   .99240D+00   .10483D+01   .10000D+00
1st derivs.    -.21423D-02  -.37952D-02  -.84224D-02  -.30909D-01
Parameters:     .36125D+00   .99240D+00   .10483D+01   .10000D+00
Itr  1 F=  .8173D+02 gtHg=  .3233D-01 chg.F=  .8173D+02 max|db|=  .3091D+00
1st derivs.     .27562D-01  -.21212D-01   .13418D+00  -.35863D-01
Parameters:     .36217D+00   .99404D+00   .10519D+01   .11333D+00
Itr  2 F=  .8173D+02 gtHg=  .1432D+00 chg.F=  .2248D-03 max|db|=  .3164D+00
1st derivs.     .73291D-02   .71339D-01   .25074D-03  -.35622D-01
Parameters:     .36170D+00   .99440D+00   .10496D+01   .11394D+00
Itr  3 F=  .8173D+02 gtHg=  .8007D-01 chg.F=  .1755D-03 max|db|=  .3126D+00
1st derivs.     .73291D-02   .71339D-01   .25074D-03  -.35622D-01
Parameters:     .36170D+00   .99440D+00   .10496D+01   .11394D+00
(Iterations omitted)
Itr  9 F=  .8115D+02 gtHg=  .2698D-02 chg.F=  .6202D-03 max|db|=  .1173D-02
1st derivs.    -.26627D-03  -.78866D-03  -.41065D-03   .15565D-03
Parameters:     .42901D+00   .11749D+01   .12798D+01   .33647D+01
Itr 10 F=  .8115D+02 gtHg=  .2211D-03 chg.F=  .4468D-05 max|db|=  .4763D-04
1st derivs.    -.33513D-05  -.59476D-04   .65059D-04  -.13674D-05
Parameters:     .42901D+00   .11750D+01   .12798D+01   .33645D+01
Itr 11 F=  .8115D+02 gtHg=  .9649D-05 chg.F=  .2507D-07 max|db|=  .8219D-06
                                                             * Converged
Normal exit:  11 iterations. Status=0, F=    81.14638
Function=  .81730029251D+02, at entry,  .81146377711D+02 at exit

**************************************************
* Hermite  quadrature with  10 nodes (points)    *
**************************************************
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable               Function
Log likelihood function         -81.14638
Estimation based on N =    200, K =   4
Inf.Cr.AIC  =    170.3 AIC/N =     .851
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|     .42901**       .18477     2.32  .0202       .06686    .79116
      B2|    1.17495***      .33368     3.52  .0004       .52095   1.82895
      B3|    1.27978***      .40403     3.17  .0015       .48790   2.07165
       T|    3.36454        3.06312     1.10  .2720     -2.63907   9.36815
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

A Gamma Integral

(This is, admittedly, a bit contrived.) The gamma integral is

$$\Gamma(P) = \int_0^\infty \exp(-v)v^{P-1}dv = (P\text{-}1)! \text{ if } P \text{ is a positive integer.}$$

Consider the following trivial modification of a probit log likelihood function:

$$F = \Sigma \{ \log \Phi(q * \boldsymbol{\beta}'\mathbf{x}) + \frac{1}{\Gamma(P)} \int_0^\infty \exp(-v)v^{P-1}dv \}.$$

Since the second term is exactly equal to one, the end result of maximizing the function shown above should be identical to the simple probit estimates, though the function value will equal the probit log likelihood plus the sample size. This could be done with **MAXIMIZE** as follows:

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-200 $
CREATE      ; z1= Rnn(0,1) ; z2 = Rnn(0,1) $
NAMELIST    ; z = one,z1,z2 $
CREATE      ; y = ( .2 + z1 + z2 + Rnn(0,1) ) > 0 ; q = 2*y - 1 $
CALC        ; p = 2 ; k = Col(z)  $
MAXIMIZE    ; Fcn = gamma = Ntg(u^(p-1)) | Log(Phi(q * b1'z)) - gamma/Gma(P)
            ; Start = k_0 ; Labels = k_b ; Glq = u ; Pts = 24 $
PROBIT      ; Lhs = y ; Rhs = z $
```

To illustrate this program, we used the data from the previous example, and set $P = 2$ in the gamma function. As expected, the coefficients are identical to the probit model and the function differs by 200. (Our maximizer translates the problem into a minimization, so the sign changes.)

```
Normal exit:   6 iterations. Status=0, F=    261.8135
**************************************************
* Laguerre quadrature with  20 nodes (points)    *
**************************************************
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable                 Function
Log likelihood function          -261.81346
Estimation based on N =     200, K =   3
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|      .04754       .12831      .37  .7110      -.20394     .29901
      B2|     1.44294***    .19271     7.49  .0000      1.06525    1.82064
      B3|     1.33166***    .18350     7.26  .0000       .97200    1.69132
--------+--------------------------------------------------------------------
Binomial Probit Model
Log likelihood function           -61.81346
--------+--------------------------------------------------------------------
Constant|      .04754       .12751      .37  .7093      -.20237     .29744
      Z1|     1.44294***    .19664     7.34  .0000      1.05753    1.82836
      Z2|     1.33166***    .18499     7.20  .0000       .96909    1.69424
--------+--------------------------------------------------------------------
```

E66.2.3 Integration by Simulation

You can include functions that include expectations of the form

$$F(\beta) = E_v\,[F(\beta,v)]$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} F(\beta,v)\,dv$$

where v is distributed as standard normal. These can be approximated reasonably accurately by simulation, by using

$$F(\beta) \approx (1/R) \sum_{r=1}^{R} F(\beta,v_r)$$

where v_r is one of R random draws from the standard normal distribution. To replace the Hermite integration with this integration by simulation, change **; Hrq = name** to

; Sim = name

To set R, the number of points for the approximation, you will use (as with other applications)

; Pts = number of points for simulations.

NOTE: The seed for the random number generator is set to the same value each time a computation is done for a specific individual. Thus, you can replicate a computation done earlier by setting the main seed for the program before estimation.

Consider the first example above. A second way to approximate the expected value would be by simulation and averaging. That is, the probability can be approximated by averaging the probabilities obtained with a sample of random draws from the distribution of v. The change in the preceding would be only to the method of integration.

The commands are:

CALC	**; Ran(12345) $**
SAMPLE	**; 1-200 $**
CREATE	**; z1 = Rnn(0,1) ; z2 = Rnn(0,1) $**
NAMELIST	**; z = one,z1,z2 $**
CREATE	**; y = (.2 + z1 + z2 + Rnn(0,1)) > 0 ; q = 2*y - 1 $**
PROBIT	**; Lhs = y ; Rhs = z $**
MAXIMIZE	**; Fcn = Prob = Ntg(Phi(q*(b1'z + t*u*x))) | Log(Prob)**
	; Start = b,.1 ; Labels = kz_b,t ; Sim = u ; Pts = 50 ; Output = 3 $

We also drop the scaling in the integral by $1/\pi^{1/2}$, since that is specific to the change of variable for the Hermite integration. The results are shown below.

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator=  50
Start values:   .47536D-01   .14429D+01   .13317D+01   .10000D+00
1st derivs.     .66861D-02  -.16079D-01  -.24850D-01  -.12111D-01
Parameters:     .47536D-01   .14429D+01   .13317D+01   .10000D+00
Itr  1 F=  .6181D+02 gtHg=  .3267D-01 chg.F=  .6181D+02 max|db|=  .1407D+00
1st derivs.    -.15593D-01  -.16631D-01   .12701D-01  -.12525D-01
Parameters:     .47194D-01   .14438D+01   .13329D+01   .10062D+00
Itr  2 F=  .6181D+02 gtHg=  .2895D-01 chg.F=  .2728D-04 max|db|=  .3304D+00
1st derivs.     .20610D-02   .22397D-02  -.68093D-02  -.12445D-01
Parameters:     .47478D-01   .14441D+01   .13327D+01   .10085D+00
Itr  3 F=  .6181D+02 gtHg=  .1451D-01 chg.F=  .7628D-05 max|db|=  .1234D+00
1st derivs.     .20610D-02   .22397D-02  -.68093D-02  -.12445D-01
Parameters:     .47478D-01   .14441D+01   .13327D+01   .10085D+00
Itr  1 F=  .6181D+02 gtHg=  .1451D-01 chg.F=  .6181D+02 max|db|=  .1234D+00
1st derivs.    -.58554D-02  -.14640D-01   .15903D-01  -.12306D-01
Parameters:     .47355D-01   .14439D+01   .13331D+01   .10159D+00
Itr  2 F=  .6181D+02 gtHg=  .2555D-01 chg.F=  .6290D-05 max|db|=  .5011D+00
1st derivs.    -.17754D-01   .38005D-01   .22105D-01  -.83349D-02
Parameters:     .47183D-01   .14464D+01   .13348D+01   .11786D+00
Itr  3 F=  .6181D+02 gtHg=  .4867D-01 chg.F=  .1051D-03 max|db|=  .1639D+01
1st derivs.     .54207D-01   .87991D-02   .25379D-01  -.35629D-02
Parameters:     .48345D-01   .14456D+01   .13345D+01   .13189D+00
Itr  4 F=  .6181D+02 gtHg=  .6060D-01 chg.F=  .8614D-04 max|db|=  .2299D+01
1st derivs.     .71268D-03  -.69440D-03  -.32695D-03  -.23829D-04
Parameters:     .47457D-01   .14448D+01   .13336D+01   .14114D+00
Itr  5 F=  .6181D+02 gtHg=  .2135D-03 chg.F=  .5609D-04 max|db|=  .9703D-03
1st derivs.    -.11707D-04   .14852D-05  -.11693D-05  -.10977D-05
Parameters:     .47446D-01   .14449D+01   .13336D+01   .14128D+00
Itr  6 F=  .6181D+02 gtHg=  .2485D-05 chg.F=  .2324D-07 max|db|=  .2519D-04
1st derivs.    -.55376D-07  -.26099D-07  -.67956D-08   .36571D-08
Parameters:     .47446D-01   .14449D+01   .13336D+01   .14129D+00
Itr  7 F=  .6181D+02 gtHg=  .1080D-07 chg.F=  .3112D-11 max|db|=  .6844D-07
                                                           * Converged
Normal exit:   7 iterations. Status=0, F=    61.81036
Function=  .61810647342D+02, at entry,  .61810358837D+02 at exit
**************************************************
* Integration by simulation using   50 draws.    *
**************************************************
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function        -61.81036
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
(Results obtained by Monte Carlo simulation)
      B1|     .04745          .12868      .37  .7123     -.20476    .29965
      B2|    1.44488***       .19721     7.33  .0000     1.05836   1.83140
      B3|    1.33361***       .19775     6.74  .0000      .94603   1.72120
       T|     .14129         3.46202      .04  .9674    -6.64415   6.92672
--------+--------------------------------------------------------------------
 (Results obtained by maximum likelihood estimation)
Log likelihood function        -61.81346
--------+--------------------------------------------------------------------
Constant|     .04754          .12751      .37  .7093     -.20237    .29744
      Z1|    1.44294***       .19664     7.34  .0000     1.05753   1.82836
      Z2|    1.33166***       .18499     7.20  .0000      .96909   1.69424
--------+--------------------------------------------------------------------
```

Note the results are nearly the same as those computed using Hermite quadrature. The difference in the fourth coefficient results from the scaling by $\sqrt{2}$. The differences across the other coefficients can be partly explained by the relatively small number of simulation points (50). Of course, the Hermite integral is also only an approximation.

E66.2.4 Maximum Simulated Likelihood Estimation

The simulation procedure described above can be extended to a vector of up to five standard normally distributed variables. The function is defined as

$$F(\beta) = \sum_{i=1}^{N} \frac{1}{R} \sum\nolimits_{r=1}^{R} g[\beta, \mathbf{Z}(i), v_{1ir}, v_{2ir}, ..., v_{Mir}].$$

(M may be up to five.) The variables in the simulated function may be freely correlated. The specification is as follows:

MAXIMIZE **; Labels** = the list of labels for the parameters
; Start = the starting values for the parameters
; Sim = a symbol for the draws
; Sdv = M specifications for the standard deviations of the *v*s
; Pts = the number of draws, R
; Fcn = the function definition
; ... any other options $

The number of variables simulated will equal the number of specifications you provide in the **; Sdv** list. Use a ‘1’ for a fixed (at 1.0) standard deviation or a ‘*’ for a free standard deviation, to be estimated. The names for the variables will then be the symbol you place in the **; Sim** specification plus the integer index.

For example,

; Sdv = *,*,*
; Sim = uab

produces a vector of three random draws with unrestricted standard deviation, named *uab*1, *uab*2, *uab*3. Optional features include

; Cor	to allow the *M* variables to be freely correlated
; Halton	to use Halton draws (see Chapter R24) instead of uniform random numbers to power the simulation
; Antithetic	to use pairs of draws, *v* and **-***v*.

E66.3 Subfunctions in Functions

Functions for **MINIMIZE** may be built up recursively by using subfunctions. The **; Fcn** part of the command will consist of

; Fcn = name1 = expression |
name2 = expression ... |
expression $

The last expression is the one being minimized, so it does not have a name. Any expression can use the name of any previous expression, as many times as desired. For examples:

; Fcn = bx = c0 + c1*x | (y-bx)^2
; Fcn = bx = c0 + c1*x | e = y - bx | e^2
; Fcn = bx = c0 + c1*x | e = y - bx | e^2 + e^4
; Fcn = d = (2*y-1)*b'x| Log(Phi(d)) (log likelihood for a probit model).

This may bring enormous gains in simplifying expressions. Functions often involve repeated use of the same function. For an example, consider the probit model, which might be inefficiently set up as

; Fcn = y*Log(Phi (b'x)) + (1-y)*Log(1 - Phi(b'x))

This can be written

; Fcn = bx = b'x |
fbx = Phi(bx) |
lfbx = Log(fbx) |
y * lfbx + (1-y)*(1-lfbx)

(Obviously, there are yet more efficient ways to do this, but this illustrates the point.) This feature will never increase the amount of computation, and will usually decrease it. It reduces the chance for error in lengthy functions. And, it will reduce overall the length of your commands. You should take advantage of this feature of the command whenever possible.

Note that functions are compiled from left to right (or, top to bottom). That means if you try to use a name which is defined after the function you are defining, an error will occur in which the name you are using does not appear to be defined.

E66.4 Supplying Derivatives for Functions

All optimizations for user defined problems (**NLSQ**, **NLSUR**, **MINIMIZE**, **MAXIMIZE, GMME**) use first difference approximations to obtain derivatives of the functions. This provides sufficient accuracy to obtain appropriate solutions to most problems. However, it is relatively slow (sometimes extremely so) compared to using analytic derivatives and, for some problems, may not be sufficiently accurate.

The subfunctions defined as above may be the derivatives of the function you are optimizing. This can speed up the computations. To indicate that a subfunction is a derivative, you just precede the name with an underscore, then the name of the parameter. For example:

```
MINIMIZE    ; Start = 0,0
            ; Labels = c0,c1
            ; Fcn = e = y - c0 - c1*x |
                    _c0 = -2*e         |
                    _c1 = -2*e*x       |
                    e^2 $
```

NOTE: If the derivatives you provide do not match the function, the optimization procedure will eventually break down, claiming to be unable to minimize the function. The optimizer cannot check your differentiation for you by any other way. However, if you have not differentiated the function correctly, the optimization will break down eventually.

Any derivatives that you do not provide are evaluated numerically as usual.

For an example, the following shows three ways to estimate the parameters of a simple probit model.

```
TIMER $
CALC        ; Ran(12345) $
SAMPLE      ; 1-2000 $
CREATE      ; x = Rnn(0,1) ; y = x + Rnn(0,1) ; y = y>0 ; q = 2*y-1 $
PROBIT      ; Lhs = y ; Rhs = one,x $
MAXIMIZE    ; Labels = b0,b1 ; Start = 0,0
            ; Fcn = Log(Phi(q*(b0+b1*x))) $
MAXIMIZE    ; Labels = b0,b1 ; Start = 0,0
            ; Fcn = bx = q*(b0+b1*x)         |
                    d = q*N01(bx)/Phi(bx)    |
                    _b0 = d                  |
                    _b1 = d*x                |
                    Log(Phi(bx)) $
```

Absolute timings are not particularly meaningful as they are specific to computers. But, for the three methods shown, on the same machine, we find identical results to six decimal places, while the formal probit estimator requires 0.04 seconds, the solver with numerical derivatives took 0.33 seconds, and the analytical derivatives reduced this to 0.23 seconds. On a comparable basis, the time savings could be substantial.

```
Normal exit:   5 iterations. Status=0, F=    1003.258
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y
Log likelihood function     -1003.25797
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.03852         .03222    -1.20  .2319     -.10166    .02463
       X|    1.02181***      .04376    23.35  .0000      .93604   1.10757
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Elapsed time:     0 hours,  0 minutes,   .04 seconds.
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   3 iterations. Status=0, F=    1003.258
-----------------------------------------------------------------------------
User Defined Optimization
Log likelihood function     -1003.25797
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.03852         .03228    -1.19  .2328     -.10179    .02476
      B1|    1.02181***      .04223    24.19  .0000      .93903   1.10458
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Elapsed time:     0 hours,  0 minutes,   .33 seconds.
Normal exit:   3 iterations. Status=0, F=    1003.258
-----------------------------------------------------------------------------
User Defined Optimization
Log likelihood function     -1003.25797
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.03852         .03228    -1.19  .2328     -.10179    .02476
      B1|    1.02181***      .04223    24.19  .0000      .93903   1.10458
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Elapsed time:     0 hours,  0 minutes,   .23 seconds.
```

Derivatives for an Index Function

If your function contains an 'index' function, such as the **β′x** that appears in a probit model, then there is a shortcut whereby you can give a full set of derivatives with a very small amount of programming. To illustrate, consider the following example which uses this 'trick' for a probit model:

```
NAMELIST    ; xvars = a namelist of  up to 100 variable names $
CREATE      ; y = the dependent variable $
CREATE      ; q = 2*y-1 $
CALC        ; kvars = Col(xvars) $
MAXIMIZE    ; Labels = kvars_c ? labels are c1,c2,...,c100
            ; Start = the set of starting values (might be a matrix)
            ; Fcn = index = q*c1'xvars                         |
                    _d(c1'xvars) = q*N01(index)/Phi(index) |
                    Log(Phi(index)) $
```

This **MAXIMIZE** command specifies the log likelihood function. The middle specification of the function definition provides analytic derivatives for all 100 variables in the model. The specification **_d(c1'xvars)** contains the label of the desired first parameter for which the derivatives are provided, then an apostrophe, followed by a namelist. An expression appears after the equals sign. This shorthand states that the analytic derivatives are obtained by multiplying the variables in the namelist by the expression. This produces derivatives for as many parameters as there are variables in the namelist. A small, specific example would be

```
NAMELIST    ; x = one,x1,x2,x3 $
MAXIMIZE    ; Labels = b1,b2,b3,b4
            ; Start = 0,0,0,0
            ; Fcn = bx = q*b1'x                  |
                    _d(b1'x) = q*N01(bx)/Phi(bx)  |
                    Log(Phi(bx)) $
```

The second line is equivalent to the four lines

```
_dc1  = q*N01(bx)/Phi(bx) * 1
_dc2  = q*N01(bx)/Phi(bx) * x1
_dc1  = q*N01(bx)/Phi(bx) * x2
_dc1  = q*N01(bx)/Phi(bx) * x3
```

E66.5 Model Specifications for the MAXIMIZE Command

This is the full list of general specifications that are applicable to this model estimator.

Controlling Output from Model Commands

; Table = name saves model results to be combined later in output tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix displays estimated asymptotic covariance matrix (normally not shown), same as **; Printvc**.

Optimization Controls for Nonlinear Optimization

; Start = list	gives starting values for a nonlinear model.
; Tlg[= value]	sets convergence value for gradient.
; Tlf [= value]	sets convergence value for function.
; Tlb[= value]	sets convergence value for parameters.
; Alg = name	requests a particular algorithm, Newton, DFP, BFGS, etc.
; Maxit = n	sets the maximum iterations.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.
; Lpt = n	sets the number of points to use Laguerre quadrature.
; Hpt = n	sets the number of points to use for Hermite quadrature.
; Set	keeps current setting of optimization parameters as permanent.

Predictions and Residuals

; List	displays a list of fitted values with the model estimates.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set.
; Fill	fills missing values (outside estimating sample) for fitted values.

Hypothesis Tests and Restrictions

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	specifies equality and fixed value restrictions.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

Predictions requested with **; List, ; Keep** and **; Fill** are the individual values of the function. There are no residuals. You may also compute a weighted log likelihood (or any other function) with

; Wts = weighting variable

Parameters may be fixed at the starting values with

; Fix = list of labels

E66.6 Output from MINIMIZE/MAXIMIZE

The results from this procedure consist of the minimized function value, and, if a sum of terms was minimized, estimates of the standard errors and asymptotic 't ratios'. If you have computed a simple function, not a sum of terms, *LIMDEP* reports 1.0 for the estimated standard errors. The asymptotic covariance matrix is estimated by the BHHH method if BFGS, DFP or steepest descent is used, or with the estimated Hessian if you request Newton's method. The latter can occasionally be problematic because the second differencing method used to estimate the Hessian does not insure positive definiteness.

Results saved by the **MINIMIZE** procedure are:

Matrices: *b* and *varb* for all parameters, including those fixed,
gradient = the first derivative vector.

Scalars: *logl* = function value,
nreg and *kreg* = the dimensions of the problem,
exitcode = the termination status for the procedure.

Last Model: The labels in your **; Labels = list** specification.

Last Function: Your function as defined.

You can obtain simulations of the function you have maximized, or functions of the parameters you have computed as well as partial effects of any function based on those parameters using **SIMULATE** and **PARTIALS**. An example appears in Section E66.8.3.

E66.7 Types of Optimization Problems

There are broadly two types of functions, one that does not require summing over a sample of observations and one that does, in the manner of a log likelihood function. You can also use **MAXIMIZE** to find the zeros of an equation and to solve a linear programming problem.

E66.7.1 Simple Function of Parameters

If the function you are minimizing or maximizing is not a sum of terms, just specify it as described above. Then, be sure to precede your command with

SAMPLE ; 1 $

For instance, one of the examples, from Goldfeld and Quandt (1972) is this four dimensional Rosenbrock function:

$$F(\boldsymbol{\beta}) = (\beta_1+10*\beta_2)^2 + 5*(\beta_3 - \beta_4)^2 + (\beta_2 - 2*\beta_3)^4 + 10*(\beta_1 - \beta_4)^4.$$

The correct function minimizing values of all four parameters are 0.0. The **; Fcn** part of the command is exactly as it is shown above. The commands for minimizing this function are:

```
SAMPLE      ; 1 $
MINIMIZE    ; Labels = b1,b2,b3,b4 ; Start = .1,-.1,.3,.05
            ; Fcn = (b1+10*b2)^2 + 5*(b3-b4)^2 + (b2-2*b3)^4 + 10*(b1-b4)^4 $
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable               Function
Log likelihood function            .00000
Estimation based on N =      1, K =   4
Inf.Cr.AIC  =      8.0 AIC/N =    8.000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|-.20178D-05    .....(Fixed Parameter).....
      B2| .20178D-06    .....(Fixed Parameter).....
      B3|-.83743D-06    .....(Fixed Parameter).....
      B4|-.83743D-06    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

If you forget to set the sample, you can still get the right answer. But, it will take longer because you will be minimizing

$$F_N = \sum_{i=1}^{N} \mathrm{F}(\beta) = N\times\mathrm{F}(\beta)$$

where N is your current sample size. This would be (wastefully) computed by summation, so the function and derivatives would be computed N separate times to obtain the identical function. However, the program will not actually report the results in the normal output table. You will receive a message of the sort

```
Normal exit:  44 iterations. Status=0, F=    .2774660E-24
Error   143: Models - estimated variance matrix of estimates is singular
Error   447: Current estimated covariance matrix for slopes is singular
```

which is produced by the example above when we change **SAMPLE ; 1 $** to **SAMPLE ; 1-200 $**. The covariance matrix truly is singular. It is a 4×4 matrix that has rank 1, equal to 200 times the outer product of the derivative vector.

E66.7.2 Solutions to Equations

The preceding suggests a method of finding a solution to an equation or the set of solutions to a set of equations. If the equation in one variable can be written in the form '$f(x) = c$' then, one way to find the value of x is to use

```
SAMPLE      ; 1 $
MINIMIZE    ; Start = a guess such that f(x) is computable
            ; Labels = x
            ; Fcn = (f(x) -  c )^2  $
```

(where you insert the definition of the function in the last line). The solution to the equation occurs where the squared difference equals zero. (Note that this need not be unique.) If there is a second equation, '$g(z) = d$,' you could use

```
            ; Labels = x,z
            ; Fcn = (f(x) - c)^2 + (g(z) - d)^2 $
```

The extension to M equations is direct. The equations may also be simultaneous, as in the example below.

The following example appears in Greene (2011, p. 460). For random sampling from a gamma population with parameters λ and P and observations x_i, $i = 1,...,N$, $E[x_i] = P/\lambda$ and $E[1/x_i] = \lambda/(P-1)$. Thus, one (admittedly inefficient) way to estimate P and λ would be to equate these sample moments to their population counterparts. Thus, we wish to solve the two equations

$$(1/N)\sum_{i=1}^{N} x_i = P/\lambda$$

and

$$(1/N)\sum_{i=1}^{N} 1/x_i = \lambda/(P-1).$$

One might proceed as follows, assuming the variable x already exists:

```
CALC        ; Ran (12345) $
SAMPLE      ; 1-100 $
CREATE      ; x = Rni(5) $
CREATE      ; x1  = 1/x $
CALC        ; m1= Xbr(x) ; m0 = Xbr(x1) $
SAMPLE      ; 1 $
MINIMIZE    ; Labels = l,p
            ; Start = 1,10
            ; Fcn = (m1 - p/l)^2 + (m0 - l/(p-1))^2
            ; Output = 3 $
```

(Note, this is not the optimal way to solve this problem. The sufficient statistics based on the log likelihood are $\Sigma_i\, x_i$ and $\Sigma_i \log x_i$, so the efficient estimator will be a function of these two moments.)

```
Nonlinear Estimation of Model Parameters
Method=BFGS  ; Maximum iterations=100
Convergence criteria:gtHg   .0000D+00 chg.F   .0000D+00 max|dB|   .1000D-05
Nodes for quadrature: Laguerre=20;Hermite=64.
Replications for GHK simulator= 100
Start values:    .10000D+01   .10000D+02
1st derivs.    -.10071D+03   .10071D+02
Parameters:      .10000D+01   .10000D+02
Itr  1 F=  .2536D+02 gtHg=  .1012D+03 chg.F=  .2536D+02 max|db|=  .1007D+03
1st derivs.    -.71513D+01   .11996D+01
Parameters:      .16659D+01   .99334D+01
Itr  4 F=  .7276D-07 gtHg=  .1115D-02 chg.F=  .5612D-06 max|db|=  .3168D-01
1st derivs.    -.17561D-04   .35871D-05
Parameters:      .10307D+01   .51186D+01
Itr  5 F=  .4081D-10 gtHg=  .3857D-05 chg.F=  .7272D-07 max|db|=  .3320D-04
1st derivs.    -.76351D-07   .16354D-07
Parameters:      .10307D+01   .51185D+01
Itr  6 F=  .1703D-10 gtHg=  .5898D-07 chg.F=  .2378D-10 max|db|=  .6762D-06
                                                           * Converged
Normal exit:   6 iterations. Status=0, F=    .1703079E-10
Function=  .25360805597D+02, at entry,  .17030790377D-10 at exit

-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable             Function
Log likelihood function          .00000
Estimation based on N =      1, K =   2
Inf.Cr.AIC  =      4.0 AIC/N =    4.000
Model estimated: Aug 24, 2011, 15:05:17
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|    1.03071    .....(Fixed Parameter).....
       P|    5.11845    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

> **NOTE:** The **SOLVE** command described in Chapter E68 may also be used to search for the roots of an equation using a grid search rather than the type of minimization used here.

E66.7.3 Sum of Terms

Optimizing a sum of terms is identical to the preceding except that your **; Fcn** expression will involve one or more variables, and you will not reset the sample to just one observation. For example, the following sets up the log likelihood function for a probit model, where *y* is the dependent variable and *x*1 and *x*2 are the independent variables.

MAXIMIZE ; Labels = a1,a2,a3 ; Start = 0,0,0
; Fcn = Log(Phi((2*y-1)*(a1+a2*x1+a3*x2))) $

> **NOTE:** The number of observations in the current sample *always* controls the number of terms in a sum. Function values are summed over the sample, even if no variables actually appear in the function.

E66.7.4 Linear Programming

MINIMIZE and **MAXIMIZE** will also solve a linear programming problem. The general form of the problem is

MINIMIZE or **MAXIMIZE** **c’x**

Subject to **xl** $\leq$ **x** $\leq$ **xu**

and **bl** $\leq$ **Ax** $\leq$ **xu**

where all terms are vectors save for **A** which is a conformable matrix of coefficients (some of which may be zero). The necessary command is

```
MAXIMIZE   ; Alg  = Simplex
           ; Lhs = c
           ; Rhs = a
           ; Limits = xl,xu $
```

The vectors *c*, *xl* and *xu* and matrix *a* must be created with **MATRIX**. Any element in *a* can be zero. For one sided limits, you may use large values such as 1.D15 in *xl* or *xu*. Matrix *a* has the same number of columns as there are activities to be solved for. There is one row for each constraint. The first element in the *j*th row is *bl*(*j*). This is followed by the row of *a*. The last element in each row is *bu*(*j*). There are no other options for this procedure. An example appears below.

Results from this procedure are the solution itself and retained values

Matrices:	*b*	= the solution vector, **x**
	lpweight	= the vector **c**
Scalars:	*lpfunctn*	= the value of the criterion
	klp	= the number of activities
	exitcode	= 0 if the problem is solved
		3 if the problem cannot be solved
		5 if an error occurs in setting up the problem

For example, we solve the problem

Maximize $F = x1 + 3x2$

Subject to $0 \leq x1 \leq 1$

$0 \leq x2 \leq 1$

$x1 + x2 \leq 1.5$

$.5 \leq x1 + x2$

The command syntax is

```
MATRIX      ; xl  = [       0,     0        ]
            ; xu  = [       1,     1        ]
            ; c   = [       1,     3        ]
            ; a   = [-1.d15,    1,     1,        1.5  /
                         .5,      1,     1,     1.d15  ] $
MAXIMIZE    ; Lhs = c ; Rhs = a
            ; Limits = xl, xu
            ; Alg = Simplex $
```

The solution is $x1 = .5$, $x2 = 1$, *function* = 3.5

```
=====================================================
Linear Programming Solution by Simplex Method =======
Maximized function value =       3.50000      =======
(* indicates the constraint is binding)       =======
=====================================================
Activity               Lower Limit      Upper Limit
-----------------    ---------------  ---------------
X01        .50000           .00000          1.00000
X02       1.00000           .00000          1.00000 *
-----------------------------------------------------
Constraint             Lower Limit      Upper Limit
-----------------    ---------------  ---------------
Row 01               *************          1.50000 *
Row 02                      .50000    *************
```

E66.8 Applications

We show several applications that use **MAXIMIZE** or **MINIMIZE** to optimize a user defined function.

E66.8.1 Simple Function

The following set of commands demonstrates several features of the **MINIMIZE** program. The technical output from the program is omitted. We show only the final results of each command. The first is from Goldfeld and Quandt (1972).

$$F(\mathbf{c}) = (c_1+10c_2)^2 + 5(c_3-c_4)^2 + (c_2-2c_3)^4 + 10(c_1-c_4)^4.$$

The correct values of all four parameters are 0.0. The Fcn part is exactly as it is shown above. The unrestricted optimum is found using

```
SAMPLE      ; 1 $
MINIMIZE    ; Labels = c1,c2,c3,c4 ; Start = .1,-.1,.3,.05
            ; Fcn = (c1+10*c2)^2 + 5*(c3-c4)^2 + (c2-2*c3)^4 + 10*(c1-c4)^4 $
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function           .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
      B1|-.20178D-05    .....(Fixed Parameter).....
      B2| .20178D-06    .....(Fixed Parameter).....
      B3|-.83743D-06    .....(Fixed Parameter).....
      B4|-.83743D-06    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

We now repeat the preceding while holding two of the parameters fixed at the starting values.

MINIMIZE ; Labels = c1,c2,c3,c4 ; Start = .1,-.1,.3,.05 ; Fix = c2,c4
; Fcn = (c1+10*c2)^2 + 5*(c3-c4)^2 + (c2-2*c3)^4 + 10*(c1-c4)^4 $

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function           .50309
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
      C1|      .36642    .....(Fixed Parameter).....
      C2|     -.10000    .....(Fixed Parameter).....
      C3|      .04458    .....(Fixed Parameter).....
      C4|      .05000    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

E66.8.2 Sum of Terms

For the sum of terms functions, we create some data.

```
CALC      ; Ran (12345) $
SAMPLE    ; 1-25 $
CREATE    ; z1 = Rnn(0,1)                    ? correlated regressors
          ; z2 = .5*(z1+Rnn(0,1))
          ; z3 = (z1 + z2 + Rnn(0,1))/3
          ; ys =  z1 + z2 + z3 + Rnn(0,2)
          ; d = ys > 0                       ? probit dependent variable
          ; t  = (d=1) * ys $                ? tobit dependent variable
NAMES     ; z  = one,z1,z2,z3 $
```

We now estimate a tobit and a probit model. Starting values are based on OLS. We use Olsen's formulation for the tobit model.

```
REGRESS   ; Lhs = t ; Rhs = z $
CALC      ; thet = 1/s   $
MATRIX    ; beta = thet * b $
```

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=T       Mean                 =         .96086
            Standard deviation   =        1.29997
            No. of observations  =             25  Degrees of freedom
Regression  Sum of Squares       =        1.87929           3
Residual    Sum of Squares       =        38.6787          21
Total       Sum of Squares       =        40.5580          24
            Standard error of e  =        1.35715
Fit         R-squared            =         .04634  R-bar squared =  -.08990
Model test  F[  3,    21]        =         .34011  Prob F > F*   =   .79654
Diagnostic  Log likelihood       =      -40.92863  Akaike I.C.   =   .75641
            Restricted (b=0)     =      -41.52168  Bayes  I.C.   =   .95143
            Chi squared [  3]    =        1.18610  Prob C2 > C2* =   .75634
Model was estimated on Aug 24, 2011 at 03:15:22 PM
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       T|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|      .97289***      .27790     3.50  .0021      .42822   1.51757
      Z1|     -.20356         .55036     -.37  .7152    -1.28224    .87512
      Z2|      .26792         .50149      .53  .5988     -.71497   1.25081
      Z3|      .39646         .80823      .49  .6288    -1.18764   1.98057
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

We now fit the probit model using **MAXIMIZE** and **PROBIT** to verify the result.

```
PROBIT      ; Lhs = d ; Rhs = z $
MATRIX      ; bp = b $
CREATE      ; q = 2 * d - 1 $
MAXIMIZE    ; Start = 0,0,0,0
            ; Labels = b1,b2,b3,b4
            ; Fcn = Log(Phi(q*b1'z)) $
MATRIX      ; List ; check = b - bp  $
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Log likelihood function       -14.72268
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       D|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|      .22863         .27473      .83  .4053     -.30983    .76708
      Z1|      .31490         .54573      .58  .5639     -.75471   1.38452
      Z2|      .16081         .50783      .32  .7515     -.83452   1.15614
      Z3|      .42252         .81342      .52  .6035    -1.17175   2.01680
--------+--------------------------------------------------------------------
User Defined Optimization
Dependent variable               Function
Log likelihood function       -14.72268
```

```
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
UserFunc|  Coefficient      Error        z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
      B1|      .22863        .28778      .79  .4269     -.33541    .79266
      B2|      .31490        .63550      .50  .6202     -.93065   1.56045
      B3|      .16081        .59396      .27  .7866    -1.00334   1.32495
      B4|      .42252        .90251      .47  .6397    -1.34636   2.19140
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
   CHECK|            1
--------+-------------
       1|   .191626E-10
       2|  -.514560E-09
       3|   .151578E-09
       4|   .594567E-09
```

The tobit model is a little more complicated. Once again, the output from the internal tobit estimator is omitted. Vector *bt* scales the tobit coefficients.

```
TOBIT       ; Lhs = t ; Rhs = z  $
MATRIX      ; bt = 1/s * b $
MAXIMIZE    ; Start = beta, thet
            ; Labels = b1,b2,b3,b4,tt
            ; Fcn = bx = b1'z | (1-d)*Log(Phi(-bx))+d*Log(tt)-d/2*(tt*t-bx)^2 $
MATRIX      ; List ; check = bt - b(1:4) $
```

(The difference in the log likelihoods occurs because the **MAXIMIZE** function does not include -.5*log(2π) in the term multiplied by *d*.)

```
Normal exit:   5 iterations. Status=0, F=    37.44973
----------------------------------------------------------------------------
Limited Dependent Variable Model - CENSORED
Dependent variable                     T
Log likelihood function       -37.44973
--------+-------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       T|  Coefficient      Error        z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Primary Index Equation for Model
Constant|      .35849        .44082      .81  .4161     -.50549   1.22248
      Z1|     -.05698        .80711     -.07  .9437    -1.63889   1.52492
      Z2|      .50113        .73716      .68  .4966     -.94367   1.94593
      Z3|      .73428       1.18243      .62  .5346    -1.58325   3.05180
        |Disturbance standard deviation
   Sigma|     1.83412***     .36275     5.06  .0000     1.12315   2.54509
--------+-------------------------------------------------------------------
Normal exit:   9 iterations. Status=0, F=    23.66565
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function       -23.66565
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|     .19546         .27971      .70  .4847     -.35276    .74368
      B2|    -.03107         .57506     -.05  .9569    -1.15816   1.09603
      B3|     .27323         .52699      .52  .6041     -.75966   1.30611
      B4|     .40034         .83489      .48  .6316    -1.23600   2.03669
      TT|     .54522***      .11417     4.78  .0000      .32146    .76899
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
   CHECK|              1
--------+--------------
       1|   -.776470E-09
       2|    .255819E-09
       3|   -.553323E-09
       4|   -.794429E-10
```

E66.8.3 Model Estimator – Canonical NB Regression Model

The following uses **MAXIMIZE** to create a new count data model in *LIMDEP* that is not in the menu of supported, built in specifications.

Hilbe (2011) recommends an alternative form of the negative binomial that he labels the ‘canonical negative binomial’ model. The signature feature of the model is that it applies to a discrete random variable with a formal negative binomial distribution – it is not obtained by integrating heterogeneity out of a mixed distribution. Hence the name ‘canonical’ – it derives from first principles. The conditional (on x_i) density of the random variable is

$$f(y_i \mid \mathbf{x}_i) = \frac{\Gamma(y_i + \theta)\lambda_i^{y_i}(1-\lambda_i)^{\theta}}{\Gamma(y_i+1)\Gamma(\theta)}, \ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

The conditional mean function for this model is

$$E[y_i \mid \mathbf{x}_i] = -\theta\frac{\lambda_i}{\lambda_i - 1}.$$

The resemblance to the more familiar NB2 model is only superficial. It can be seen from the conditional mean function that the parameters in the models are very different. A more transparent way to examine the difference is to examine the partial effects. In the NB2 model,

$$\partial \mathrm{E}[y_i|\mathbf{x}_i]/\partial\mathbf{x}_i \ = \ \lambda_i\boldsymbol{\beta} \ = \ E[y_i|\mathbf{x}_i]\ \boldsymbol{\beta}.$$

In the CNB model,

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = -\theta\frac{-\lambda_i}{(\lambda_i-1)^2}\boldsymbol{\beta} = -E[y_i \mid \mathbf{x}_i]\left(\frac{\lambda_i}{\lambda_i-1}\right)\boldsymbol{\beta}.$$

This is a completely different scaling of the parameter vector. The implication seems likely to be that the parameters themselves from the two models will differ substantially if, as is common, the differences tend to even out in the partial effects. We will explore this in an example below.

The canonical NB model is not a built in procedure in *LIMDEP*. However, it is a very straightforward application of the **MAXIMIZE** command to obtain the estimates followed by **PARTIALS** and **SIMULATE** to obtain the partial effects and model simulations. The program below is written in the form of a template that requires only the specification of the dependent variable and the namelist containing the regressors. A substantive complication for this estimator is the starting values. The ordinary NB estimates might seem natural, but as the analysis above suggests, the parameters in the NB2 and the CNB models are likely to be quite different. Hilbe suggests - 1 for the constant term, zeros for the slopes, and 2 for θ (i.e., .5 for $\alpha = 1/\theta$).

The procedure is generic save for a single line that is modified for the specific application

```
PROC = CNBModel(y,x) $
CALC        ; k = Col(x) $
? MAXIMIZE Estimates the model parameters
MAXIMIZE    ; Start = -1,k_0,2
            ; Labels = b0,k_b,theta
            ; Fcn = bx = b0+b1'x          |
                    lambdai = Exp(bx)  |
                    y*bx + theta*Log(1-lambdai)
                    + Lgm(y+theta) - Lgm(y+1) - Lgm(theta) $
? PARTIALS computes the partial effects for the variables in the namelist
PARTIALS    ; Parameters = b
            ; Labels = b0,k_b,theta
            ; Covariance = varb
            ; Function = bx = b0+b1'x  |
                    lambdai = exp(bx)  |
                    -theta*lambdai/(lambdai-1)
            ; Effects: x ; Summary $
? We compare the results to the NB2 model. Partials are comparable APEs
NEGBIN      ; Lhs = y ; Rhs = one,x $
PARTIALS    ; Effects: x ; Summary $
ENDPROC $
```

To execute the procedure, we use the health care data, and commands

```
SAMPLE      ; All $
NAMELIST    ; x = age,educ,hhninc,female$
EXECUTE     ; Proc = CNBModel(docvis,x) $
```

The results are as follows: Notice that they begin with several warnings about the computation of the function. Unlike other models that we have examined thus far, this model does involve a computation that is quite likely to produce this result. One of the terms in the log likelihood is $\log(1-\lambda_i)$. The implication is that λ_i must be between zero and one. Since $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$, there is no constraint that can be placed on the parameters that will enforce this boundary. It is not unlikely that for some observations, this error will occur. The solver will draw the iterations on the parameters away from these values as it gets closer to a solution.

```
  Error   590: Obs.=     1 Cannot compute function: Logminus
Warning   137: Iterations: function not computable at crnt.trial estimates
  Error   590: Obs.=    96 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus
  Error   590: Obs.=    94 Cannot compute function: Logminus

Normal exit:  19 iterations. Status=0, F=    60207.36
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function     -60207.36401
Estimation based on N =  27326, K =   6
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.23013***      .00867   -26.54  .0000     -.24712   -.21313
      B1|     .00272***   .9967D-04    27.28  .0000      .00252    .00291
      B2|    -.00435***      .00063    -6.87  .0000     -.00559   -.00311
      B3|    -.06643***      .00747    -8.89  .0000     -.08107   -.05179
      B4|     .03904***      .00220    17.77  .0000      .03474    .04335
   THETA|     .52343***      .00545    96.04  .0000      .51275    .53411
--------+--------------------------------------------------------------------

---------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect       Error      |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .06788      .00324    20.97       .06153       .07422
      EDUC         -.10864      .01603     6.78      -.14007      -.07722
      HHNINC      -1.65817      .19003     8.73     -2.03061     -1.28572
   *  FEMALE         .91204      .05373    16.97       .80672      1.01736
---------------------------------------------------------------------
 (Intermediate results for Poisson regression omitted)
-----------------------------------------------------------------------------
Normal exit:  10 iterations. Status=0, F=    60164.22
-----------------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function     -60164.22014
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .62857***      .05457    11.52  .0000      .52162    .73553
     AGE|     .02042***      .00070    29.07  .0000      .01904    .02179
    EDUC|    -.03539***      .00378    -9.36  .0000     -.04281   -.02798
  HHNINC|    -.48779***      .04520   -10.79  .0000     -.57637   -.39921
  FEMALE|     .32673***      .01588    20.58  .0000      .29561    .35784
        |Dispersion parameter for count data model
   Alpha|    1.90309***      .01984    95.94  .0000     1.86421   1.94197
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------
Partial Effects for Loglinear, Exponential Mean
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
--------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
--------------------------------------------------------------------
      AGE           .06514     .00246   26.44      .06031      .06996
      EDUC        -.11290      .01213    9.31    -.13668     -.08913
      HHNINC     -1.55610      .14504   10.73   -1.84037    -1.27183
   *  FEMALE      1.03372      .05259   19.66      .93065     1.13680
--------------------------------------------------------------------
Maximum repetitions of PROC
```

We note, finally, a possible extension of the model. In the NB1 and NB2 formulations, we allow for heterogeneity in the scale parameter, θ. In particular, the generalized model specifies

$$\theta_i = \theta \exp(\boldsymbol{\delta}'\mathbf{z}_i).$$

It is straightforward to incorporate the same extension in the canonical model, as shown in the revised procedure below:

```
PROC = CNBModel(y,x,z) $
CALC  ; k = Col(x) ; m = Col(z) $
? MAXIMIZE Estimates the model parameters
MAXIMIZE    ; Start = -1,k_0,2, m_0
            ; Labels = b0,k_b,theta,m_d
            ; Fcn = bx = b0+b1'x         |
                    lambdai = Exp(bx)    |
                    vh = Exp(d1’z)       |
                    y*bx + theta*vh*Log(1-lambdai)
                    + Lgm(y+theta*vh) - Lgm(y+1) - Lgm(theta*vh) $
? PARTIALS computes the partial effects for the variables in the namelist
NAMELIST    ; xz = x,z $
PARTIALS    ; Parameters = b
            ; Labels = b0,k_b,theta,m_d
            ; Covariance = varb
            ; Function = bx = b0+b1'x   |
                    lambdai  = Exp(bx)  |
                    vh = Exp(d1’z)      |
                    -theta*vh*lambdai/(lambdai-1)
            ; Effects: xz ; Summary  $
ENDPROC$
SAMPLE      ; All $
NAMELIST    ; z = hhkids $
NAMELIST    ; x = age,educ,hhninc,female$
EXEC        ; Proc = CNBModel(docvis,x,z) $
```

The results of the computation of this extended model are shown below.

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable               Function
Log likelihood function    -60147.26561
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B0|    -.20971***      .00879   -23.86  .0000     -.22693   -.19248
      B1|     .00233***      .00010    22.58  .0000      .00213    .00253
      B2|    -.00459***      .00063    -7.23  .0000     -.00583   -.00335
      B3|    -.06695***      .00747    -8.96  .0000     -.08159   -.05231
      B4|     .03939***      .00220    17.91  .0000      .03508    .04370
   THETA|     .55914***      .00670    83.40  .0000      .54600    .57228
      D1|    -.17037***      .01569   -10.86  .0000     -.20112   -.13961
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
---------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
---------------------------------------------------------------------
                   Partial    Standard
(Delta method)     Effect      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
      AGE            .05720     .00306   18.71       .05121       .06320
      EDUC         -.11267     .01579    7.14      -.14362      -.08172
      HHNINC      -1.64338     .18613    8.83     -2.00818     -1.27858
   *  FEMALE         .90849     .05293   17.16       .80474      1.01224
   *  HHKIDS       -.52822     .04754   11.11      -.62139      -.43504
---------------------------------------------------------------------
```

E67: GMM Estimation

E67.1 Introduction

LIMDEP can be used for GMM estimation of econometric models. Although the methodology is common to all of them, we provide several approaches. The nonlinear least squares estimator presented in the Chapter E14 is based on the least squares criterion

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$$

which minimizes the simple sum of squares of a set of residuals. As noted earlier, different weighting schemes and the use of instrumental variables extends this to more general GMM interpretations. Thus, the more general estimation criterion,

$$M(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})$$

allows for instrumental variables and a weighting matrix. Depending on the choice of the weighting matrix, this will produce GMM estimators of various sorts. Section E21.5 and Chapter E25 extend this nonlinear least squares or instrumental variables approach to multiple equations. Finally, consider the less structured GMM criterion:

$$\text{q} = \overline{\mathbf{m}}'\boldsymbol{\Sigma}\,\overline{\mathbf{m}}$$

where $$\overline{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)$$

based on a set of L ‘orthogonality conditions,’

$$E[\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)] = \mathbf{0}.$$

E67.2 General Specifications of the GMM Estimator

The GMM estimation procedure departs from a set of ‘orthogonality conditions,’

$$E[m_{il}(\boldsymbol{\beta},\mathbf{x}_i)] = 0$$

where $\boldsymbol{\beta}$ is the vector of parameters to be estimated, $\mathbf{x}_i$ is a set of variables that is assumed to be in the set of information that defines the ‘moment condition,’ and $m_{il}(.)$ is one of L expectations that the model specifies to equal zero. The GMM estimator is obtained by finding the estimator, **b**, that makes the empirical moment,

$$\overline{m}_l = \frac{1}{n}\sum_{i=1}^{n} m_{il}(\mathbf{b},\mathbf{x}_i)$$

mimic the population expectation as closely as possible.

There are three possible cases:

- If there are L functionally independent conditions specified and $K = L$ parameters to be estimated, it will generally be possible to find a **b** that makes the empirical moments match the population expectations.

- If $L > K$, then it will generally not be possible to make the moments all equal zero, and we will have, instead, to minimize some criterion which makes the moments 'close' to zero. This is the GMM estimation problem.

- If $L < K$, then there are more parameters to be estimated than there are moment conditions specified, and, since they are functionally independent, the L moment conditions will not be sufficient to identify the parameters, and estimation will be impossible.

E67.3 GMM Estimation

Collect the L moment specifications in the column vector

$$\bar{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta},\mathbf{x}_i)\,.$$

The GMM estimator is the minimum distance estimator which minimizes the quadratic form

$$q = \bar{\mathbf{m}}'\boldsymbol{\Sigma}\,\bar{\mathbf{m}}$$

for some choice of positive definite matrix $\boldsymbol{\Sigma}$. Different choices of $\boldsymbol{\Sigma}$ will produce different estimators. At this point, we turn to formulating the command for the GMM estimator. A brief application will be shown next, then the remaining details of using the estimator will be given. Some technical details will follow.

The essential command structure for the GMM estimator is

GMME	**; Fn1**	=	definition of the first moment condition
	; Fn2	=	definition of the second moment condition
	; ...		up to 50 orthogonality conditions
	; Labels	=	the symbols used for the parameters,
	; Start	=	starting values for the optimization **$**

This basic command – note that $\boldsymbol{\Sigma}$ is not specified, requests minimization of the simple sum of squares. The default specification, therefore, is $\boldsymbol{\Sigma} = \mathbf{I}$. Notice that the number of parameters may not exceed the number of functions. The function definitions can make use of all the tools discussed earlier for specifying nonlinear regressions. They may also specify instrumental variables, as shown in the examples below.

Example 1:

Suppose $y_1...,y_n$ are a sample of n independent observations from the gamma distribution,

$$f(y) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda y} y^{P-1}, y \geq 0, \lambda, P > 0.$$

Then, the following expectations hold

$$\begin{aligned}
E[y] &= P/\lambda \\
E[y^2] &= P(P+1)/\lambda^2 \\
E[1/y] &= \lambda/(P-1), P > 1 \\
E[\log y] &= \Psi(P) - \log\lambda
\end{aligned}$$

where $\Psi(P)$ is the Psi function, $d\log\Gamma(P)/dP$. Any two moments could be used for estimation of the parameters. To use the two which, it turns out, define the maximum likelihood estimator, consider the first and the fourth. The command would be

```
GMME        ; Fn1 = y - p/lambda
            ; Fn2 = Log(y) - Psi(p) + Log(lambda)
            ; Start = ... the starting values
            ; Labels = p,lambda  $
```

Example 2: (This example is from Ruud (2000)

Hansen and Singleton's classic (1982) paper on consumption and asset pricing suggests the moment equations

$$E\left[z_{tj} \left\{ \left(\frac{1+r_t}{1+\delta}\right)\left(\frac{C_t}{C_{t-1}}\right)^{\gamma-1} - 1 \right\} \middle| t-1 \right] = 0$$

for a set of instrumental variables z_{tj} where t indexes periods, C_t is consumption, r_t is return, and δ and γ are the parameters to be estimated. Ruud suggests the instrumental variables obtained by differentiating the function in brackets with respect to $1/(1+\delta)$ and γ, which produces,

$$z_{t1} = (1+r_{t-1})\left(\frac{C_{t-1}}{C_{t-2}}\right)^{\gamma-1}$$

and

$$z_{t2} = z_{t1} \times \log\left(\frac{C_{t-1}}{C_{t-2}}\right)$$

We could set this up for estimation as follows:

```
SAMPLE    ; 1 - whatever is appropriate $
CREATE    ; ct1 = c / c[-1]
          ; lagct1 = ct1[-1]
          ; If(_obsno > 2) loglag = Log(lagct1)
          ; r1 = 1+r
          ; lagr1 = r1[-1] $
SAMPLE    ; 3 - whatever is appropriate
GMME      ; Labels = delta,gamma
          ; Start = 0,0
          ; Fn1 = (r/(1+delta) * ct1^(gamma-1) - 1) * lagr1 * ctl^(gamma-1)
          ; Fn2 = (r/(1+delta) * ct1^(gamma-1) - 1) * lagr1 * ctl^(gamma-1) * loglag $
```

We note, this can be made simpler to specify and to estimate by slightly reparameterizing the function. Let $\theta = 1/(1+\delta)$ and $\tau = \gamma - 1$. Making the substitutions, we would obtain the same results with

```
GMME      ; Labels = delta,gamma
          ; Start = 0,0
          ; Fn1 = ( r1 * theta * ct1^tau-1) * lagr1 * ctl^tau
          ; Fn2 = ( r1 * theta * ct1^tau-1) * lagr1 * ctl^tau * loglag $
WALD      ; Fn1 = 1/theta - 1
          ; Fn2 = tau + 1 $ (We do this to see our original parameters.)
```

E67.4 The Weighting Matrix

The GMM estimator defined earlier is consistent regardless of what matrix **Σ** is used in the minimization. (Indeed, if the problem is 'exactly identified,' that is, if there are the same number of equations as parameters), than, as has been widely documented elsewhere, the identical solution will be obtained for all matrices **Σ**. However, in terms of the efficiency of the estimator, not all choices are the same – in this discussion, we now consider only 'overidentified' problems, in which there are more equations than parameters. You may specify any matrix you like to be used in the optimization by adding

; Sigma = the name of the matrix

to the command. The name given must be that of a positive definite matrix with number of rows and columns equal to the number of moment equations.

The Optimal Weighting Matrix

As noted, you may specify any matrix you wish for the weighting in the criterion function. For GMM estimation, the 'optimal' weighting matrix is

$$\mathbf{\Sigma}^* = \{\text{Var}[\overline{\mathbf{m}}]\}^{-1}$$

This matrix can be estimated if one has in hand any consistent estimator of the model parameters. Thus, let **b** be that estimator. Then, the estimator would be

$$\mathbf{S}^* = \left(\frac{1}{n}\right)\left(\frac{1}{n}\right)\sum_{i=1}^{n}\mathbf{m}_i(\mathbf{b},\mathbf{x}_i)\mathbf{m}_i(\mathbf{b},\mathbf{x}_i)'$$

A natural way to proceed, then, would be to use two steps:

Step 1. Use the default $\mathbf{\Sigma} = \mathbf{I}$ to obtain the initial consistent estimates of the parameters,

Step 2. After computing **S***, redo the estimation while specifying $\mathbf{\Sigma}$ to be the inverse of this estimate.

When you use the **GMME** command, *LIMDEP* automatically saves **S*** for you as a matrix named *sigma*. So, to do the two steps, you would proceed as follows:

GMME	**; Fn1**	= definition of the first moment condition
	; Fn2	= definition of the second moment condition
	; ...	up to 20 orthogonality conditions
	; Labels	= the symbols used for the parameters,
	; Start	= starting values for the optimization **$**
MATRIX	**; optimalw = <sigma> $**	
GMME	**; Fn1**	= definition of the first moment condition
	; Fn2	= definition of the second moment condition
	; ...	up to 20 orthogonality conditions
	; Labels	= the symbols used for the parameters,
	; Start	= starting values for the optimization
	; Sigma	= **optimalw $**

E67.5 Output – Displayed Results

The **GMME** command is a particular form of **MINIMIZE**, so the results and displays are almost identical. The initial table of results will contain additional results that are specific to GMM estimation, as shown in the example below. The value of the GMM criterion is displayed as the function value. The 'degrees of freedom' is the difference between the number of moment equations specified and the number of parameters estimated. If this is positive, so that the model is overidentified, then a chi squared statistic can be computed to test the overidentifying restrictions – this equals the criterion function. This test is reported as part of the output.

```
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function             .00180
Estimation based on N =      20, K =   2
Inf.Cr.AIC  =      4.0 AIC/N =     .200
GMM Criterion function              .00180
Degrees of freedom = #eqn-#parms =    2
Significance level                  .99910
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
```

E67.6 Other Options

GMME is an optimization command that is largely the same as **NLSQ** and **MINIMIZE**. All other options that are available for the nonlinear optimization procedures, including output display and convergence are useable here as well. Moreover, the full range of specification options are available for defining the moment equations; that is, all functions, using quadrature, linear, bilinear, and quadratic forms, use of namelists, and so on, may all be used as they are in other optimization problems.

E67.7 Application

The following example appears in Chapter 18 of Greene (2011). It is based on 20 observations on a random variable '*y*' to which we fit a gamma distribution with parameters λ and P (see Example 1 above). The data are

y = 20.5, 31.5, 47.7, 26.2, 44, 8.28, 30.8, 17.2, 19.9, 9.96, 55.8, 25.2, 29, 85.5, 15.1, 28.5, 21.4, 17.7, 6.42, 84.9

We first obtain the maximum likelihood estimates by maximizing the log likelihood function directly:

```
MAXIMIZE    ; Fcn = p*Log(l) - Lgm(p) - l*y + (p-1)*Log(y)
            ; Labels = l,p
            ; Start = .1,2 $
```

The GMM estimator based on the first and fourth moments will replicate the maximum likelihood estimator.

```
GMME        ; Labels = l,p
            ; Start =.1,2
            ; Fn1 = p/l - y  ? We changed the sign of this, for convenience.
            ; Fn2 = Log(y) - Psi(p) + Log(l) $
```

Note, however, that the asymptotic covariance matrix will differ – a finite sample difference – because of the different formulas used to do the computations. It seems useful to pursue that difference here, as we can derive the results in full detail for this simple problem. We use the BHHH estimator for the asymptotic covariance matrix for the MLE. For the gamma model above,

$$\partial \log L/\partial\lambda \;=\; \Sigma_i\,(P/\lambda\; - \;y)$$

$$\partial \log L/\partial P \;=\; \Sigma_i\,(\log\lambda - \Psi(P) + \log y).$$

Note that the first order conditions for the MLE are $n\bar{\mathbf{m}} = \mathbf{0}$. Let $\mathbf{M}$ be the 20×2 matrix whose *i*th row is the derivative shown above for the *i*th observation. Then, the estimator of the asymptotic covariance matrix for the MLE is

$$\text{Est.Asy.Var[MLE]} \;=\; (\mathbf{M}'\mathbf{M})^{-1}.$$

For the GMM estimator, $\mathbf{\Sigma} = \mathbf{I}$ while $\mathbf{G}$ turns out to be a sum of constants, so the n disappears;

$$\mathbf{G} = \begin{bmatrix} -P/\lambda^2 & 1/\lambda \\ 1/\lambda & -\Psi'(P) \end{bmatrix}.$$

Inserting these in the formula for the asymptotic covariance matrix of the GMM estimator, we obtain after canceling

$$\text{Est.Asy.Var}[GMM] = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{M}'\mathbf{M}\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}.$$

As can be seen, this differs from the formula for the MLE. Since **G'G** and (1/n)**M′M** converge to the same matrix, we see that the difference is due to finite sample variation. Finally, we obtain the full GMM estimator, using all four moment equations, and two steps to obtain the efficient estimator at the second step.

This is the maximum likelihood estimator

```
MAXIMIZE    ; Fcn = p*Log(l) - Lgm(p) - l*y + (p-1)*Log(y)
            ; Labels = l,p
            ; Start = .1,2 $
```

```
Normal exit:   5 iterations. Status=0, F=    85.37567
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function        -85.37567
Estimation based on N =     20, K =   2
Inf.Cr.AIC  =    174.8 AIC/N =    8.738
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .07707***      .02708     2.85  .0044      .02400    .13014
       P|    2.41060***      .87683     2.75  .0060      .69206   4.12915
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This is the GMM estimator based on the same two moments as used by the maximum likelihood estimator.

```
GMME        ; Labels = l,p
            ; Start =.1,2
            ; Fn1 = p/l - y  ? We changed the sign of this, for convenience.
            ; Fn2 = Log(y) - Psi(p) + Log(l) $
```

```
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit:   5 iterations. Status=0, F=    .1203629E-14
```

```
--------------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function           .00000
Estimation based on N =      20, K =   2
Inf.Cr.AIC  =      4.0 AIC/N =     .200
GMM Criterion function            .00000
Degrees of freedom = #eqn-#parms =    0
Significance level               1.00000
Covariance matrix for moments kept as SIGMA
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
       L|     .07707***      .02555     3.02  .0026      .02698    .12716
       P|    2.41060***      .60848     3.96  .0001     1.21800   3.60321
--------+-----------------------------------------------------------------------
```

The following uses two different moments.

GMME ; Labels = l,p ; Start = .1,2
; Fn1 = y-p/l
; Fn2 = 1/y - l/(p-1) $

```
Normal exit:   7 iterations. Status=0, F=    .4554736E-17
--------------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function           .00000
Estimation based on N =      20, K =   2
Inf.Cr.AIC  =      4.0 AIC/N =     .200
GMM Criterion function            .00000
Degrees of freedom = #eqn-#parms =    0
Significance level               1.00000
Covariance matrix for moments kept as SIGMA
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
       L|     .08862***      .02677     3.31  .0009      .03616    .14109
       P|    2.77198***      .58599     4.73  .0000     1.62346   3.92050
--------+-----------------------------------------------------------------------
```

These are the GMM estimators based on all four moments. The first pass uses the identity matrix for the weighting matrix.

GMME ; Labels = l,p ; Start = .1,2
; Fn1 = y-p/l
; Fn2 = 1/y - l/(p-1)
; Fn3 = y^2 - p*(p+1)/l^2
; Fn4 = Log(y) - Psi(p) + Log(l) $

```
Normal exit:   6 iterations. Status=0, F=    .9017024E-03
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function          .00180
Estimation based on N =     20, K =   2
Inf.Cr.AIC  =      4.0 AIC/N =     .200
GMM Criterion function           .00180
Degrees of freedom = #eqn-#parms =    2
Significance level               .99910
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .06580***      .01890     3.48  .0005      .02876    .10284
       P|    2.05830***      .50345     4.09  .0000     1.07156   3.04504
--------+--------------------------------------------------------------------
```

For the second step of GMM, we use the optimal weighting matrix, based on the previous results.

```
MATRIX      ; optimalw = <sigma> $
GMME        ; Labels = l,p ; Start = .1,2 ; Sigma = optimalw
            ; Fn1 = y-p/l
            ; Fn2 = 1/y - l/(p-1)
            ; Fn3 = y^2 - p*(p+1)/l^2
            ; Fn4 = Log(y) - Psi(p) + Log(l) $
```

```
Normal exit:   9 iterations. Status=0, F=    .9876078
-----------------------------------------------------------------------------
User Defined Optimization
Generalized Method of Moments Estimator
Log likelihood function         1.97522
Estimation based on N =     20, K =   2
Inf.Cr.AIC  =       .0 AIC/N =     .002
GMM Criterion function          1.97522
Degrees of freedom = #eqn-#parms =    2
Significance level               .37247
Covariance matrix for moments kept as SIGMA
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       L|     .12449***      .03403     3.66  .0003      .05780    .19118
       P|    3.35894***      .64628     5.20  .0000     2.09225   4.62563
--------+--------------------------------------------------------------------
```

E67.8 Technical Details for the GMM Estimator

The underlying theory for the GMM estimator is well documented in the current literature, including the current textbooks such as Greene (2011), Ruud (2000), and Hayashi (2000), so it will be omitted here, and only final results will be shown.

The estimation criterion used is

$$q \quad = (1/2)\ \bar{\mathbf{m}}'\, \boldsymbol{\Sigma}\ \bar{\mathbf{m}}\,.$$

(The 1/2 is purely for convenience – it allows the ‘2’ to disappear from the derivatives.)

NOTE: The output displayed by the program reports $2q$, not q. That is, your final results will report the value of the quadratic form, not one half times it.

The first order conditions for minimizing q are

$$\frac{\partial q}{\partial \boldsymbol{\beta}} \ = \ \mathbf{G}'\, \boldsymbol{\Sigma}\ \overline{\mathbf{m}} \ = \ \mathbf{0},\ \text{where}\ \ \mathbf{G} = \frac{\partial \overline{\mathbf{m}}}{\partial \boldsymbol{\beta}'}\,.$$

Note that there are L equations and K parameters and $L \geq K$. Thus, $\mathbf{G}$ is an $L\times K$ matrix of partial derivatives. (Note, as well, that $\mathbf{G}$ is a sample mean.) If there are K moment equations used to identify the K parameters, then assuming that $\boldsymbol{\Sigma}$ is positive definite and that the moment equations are functionally independent so that $\mathbf{G}$ has an inverse, then we can premultiply the first order condition by $(\mathbf{G}'\ \boldsymbol{\Sigma})^{-1}$ and obtain the simpler necessary condition, $\overline{\mathbf{m}} = \mathbf{0}$. The solution to this is independent of $\boldsymbol{\Sigma}$, which establishes the earlier claim that $\boldsymbol{\Sigma}$ is irrelevant to the solution to an exactly identified problem.

The asymptotic covariance matrix is computed using the estimated parameters, and

$$\text{Est.Var}[\mathbf{b}] \ = \ \left[\mathbf{G}'\, \boldsymbol{\Sigma}\, \mathbf{G}\right]^{-1} \mathbf{G}'\, \boldsymbol{\Sigma}\, \mathbf{S}^*\, \boldsymbol{\Sigma}\, \mathbf{G} \left[\mathbf{G}'\, \boldsymbol{\Sigma}\, \mathbf{G}\right]^{-1}$$

where $\mathbf{S}^*$ was defined earlier. Note that if you have specified the optimal weighting matrix, $\boldsymbol{\Sigma} = (\mathbf{S}^*)^{-1}$, then the estimated variance reduces to the familiar result,

$$\text{Est.Var}[\mathbf{b}] \ = \ \left[\mathbf{G}'\, (\mathbf{S}^*)^{-1}\, \mathbf{G}\right]^{-1}.$$

If the model is exactly identified, then q is minimized at zero. (See the example above.) If not, then q will be positive. The theoretical result that $2q$ will have a limiting chi squared distribution with degrees of freedom equal to the number of overidentifying restrictions (equations minus parameters) can be used to test restrictions in this framework. (The multiplier, 2, appears because in our formulation of the problem, we initially divided by 2.) For two nested models, with $q0$ being the unrestricted one and $q1$ embodying the restrictions, $2(q1 - q0)$ can be used to test the restrictions – refer this statistic to the chi squared table with degrees of freedom equal to the number of restrictions.

E68: Numerical Analysis

E68.1 Introduction

This chapter describes some features for examining nonlinear functions. The six commands described here are:

WALD	for obtaining variances and covariances for nonlinear functions,
FPLOT	for plotting a nonlinear function,
MINIMIZE	for computing first and second derivatives of a function,
FINTEGRAL	for obtaining the integral of a function
SOLVE	for finding the zeros of a function,
FUNCTION	for evaluating a function.

All of these features are modifications of the **MINIMIZE** command. Relevant information which you should examine before using these features is given in Chapter E66 (the **MINIMIZE** and **MAXIMIZE** commands), Chapter E14 (the **NLSQ** command and most of the information needed to use **MINIMIZE**), and in Section R14.4 which describes the **WALD** command in full detail.

NOTE: After the estimation programs, we would expect the **WALD** command to be one of the most useful tools that you will find in *LIMDEP*. Analysts often devote large amounts of time and effort to obtaining standard errors and confidence intervals for functions of things that they estimate (such as parameters). **WALD** automates this entire procedure. In particular, you will not have to obtain and program any derivatives for applying the delta method. This is all done automatically.

E68.2 Variances for Nonlinear Functions

The **WALD** command for analyzing nonlinear functions of parameters is described in full in Chapter R14. It is provided primarily for testing hypotheses, but you can use **WALD** simply to compute the variances and covariances of a set of functions (or, just the functions themselves). To avoid duplication, we will merely reproduce the essential format of the command here:

```
WALD        ; Fn1 =  first nonlinear function
            ; Fn2 = second function
            ; ... up to 50 functions of 100 parameters
            ; Labels = list of labels
            ; Start = values of underlying parameters
            ; Var = covariance matrix for parameters  $
```

WALD computes the sample averages of the functions if the command contains

```
            ; Average
```

You would use this, for example, for analyzing average partial effects. (You might use **PARTIALS** instead for this function, however **WALD** might be simpler for a general function of the model parameters that is not a conditional mean.)

Output from this command is the set of function values, estimates of their standard errors, asymptotic ‘t ratios’ and the probability that a standard normal variable would exceed that value in absolute value. *LIMDEP* will also attempt to compute a Wald statistic for the null hypothesis that all the functions are zero, which would be

$$Wald = \mathbf{f}' [\mathbf{Var(f)}]^{-1}\mathbf{f}$$

where **f** is the set of functions, and the covariance matrix is that of the estimated functions. If the matrix is not positive definite, a warning will be issued. This can be ignored if your only intent is to compute a set of nonlinear functions. In this case, there is no requirement that the functions be independent, so singularity of the estimated covariance matrix should not be taken as indicative of any problem. In addition to the display, *LIMDEP* retains matrices *waldfns* containing the matrix of functions and *varwald* with the estimated asymptotic covariance matrix. If the Wald statistic is computable, a scalar named *wald* will contain the value.

Kmenta’s (1967) method of estimating the CES production provides a useful example. The production function is

$$\log y = \log \gamma - (\nu/\rho)\log[\delta K^{-\rho} + (1-\delta)L^{-\rho}].$$

Kmenta derives the approximation

$$\text{Log} y \approx \beta_1 + \beta_2 \log K + \beta_3 \log L + \beta_4 \log^2(K/L) + \varepsilon,$$

where

$$\beta_1 = \log\gamma, \qquad \gamma = e^{\beta_1},$$

$$\beta_2 = \nu\delta, \qquad \delta = \beta_2 / (\beta_2 + \beta_3),$$

$$\beta_3 = \nu(1 - \delta), \qquad \nu = \beta_2 + \beta_3,$$

$$\beta_4 = -\rho\nu\delta(1-\delta)/2, \qquad \rho = -2\beta_4(\beta_2 + \beta_3) / (\beta_2\beta_3).$$

The results below show the application of Kmenta’s estimator to Greene’s (Table A7.1) data on the SIC 33, primary metals. Note that **WALD** requires nothing more than the formulas for the nonlinear functions. Everything else needed for the computation is found internally.

The commands are:

IMPORT $

```
 i,y,k,l
 1    657.29   162.31    279.99
 2    935.93   214.43    542.50
 3   1110.65   186.44    721.51
 4   1200.89   245.83   1167.68
 5   1052.68   211.40    811.77
 6   3406.02   690.61   4558.02
 7   2427.89   452.79   3069.91
 8   4257.46   714.20   5585.01
 9   1625.19   320.54   1618.75
10   1272.05   253.17   1562.08
11   1004.45   236.44    662.04
12    598.87   140.73    875.37
13    853.10   145.04   1696.98
14   1165.63   240.27   1078.79
15   1917.55   536.73   2109.34
16   9849.17  1564.83  13989.55
17   1088.27   214.62    884.24
18   8095.63  1083.10   9119.70
19   3175.39   521.74   5686.99
20   1653.38   304.85   1701.06
21   5159.31   835.69   5206.36
22   3378.40   284.00   3288.72
23    592.85   150.77    357.32
24   1601.98   259.91   2031.93
25   2065.85   497.60   2492.98
26   2293.87   275.20   1711.74
27    745.67   137.00    768.59
```

CREATE ; ly = Log(y) ; lk = Log(k) ; ll = Log(l) ; lkl = (Log(k/l))^2 $

Compute the regression coefficients.

REGRESS ; Lhs = ly ; Rhs = one,lk,ll,lkl $

There are two ways to specify the command. The first is the generic method, in which you supply all the information needed to do the analysis:

WALD ; Fn1 = gamma = Exp(b1) **?** Functions to analyze
; Fn2 = delta = b2/(b2+b3)
; Fn3 = nu = b2+b3
; Fn4 = rho_kl = -2*b4*(b2+b3)/(b2*b3)
; Start = b **?** Values of the parameters
; Var = varb **?** Asymptotic covariance matrix
; Labels = b1,b2,b3,b4 $ **?** Labels for the parameters.

If the analysis is based on the most recently fit model, then all the information needed has been saved as *b*, *varb*, and the *last model* names. All you must provide is the functions.

```
WALD    ; Fn1 = gamma = Exp(b_one)
        ; Fn2 = delta = b_lk / (b_lk+b_ll)
        ; Fn3 = nu = b_lk + b_ll
        ; Fn4 = rho_kl = -2 * b_lkl * (b_lk + b_ll) / (b_lk * b_ll) $
```

The second set of results is identical to the first, so it is not presented.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LY       Mean                 =       7.44363
             Standard deviation   =        .76115
             No. of observations  =            27  Degrees of freedom
Regression   Sum of Squares       =       14.2614            3
Residual     Sum of Squares       =       .801802           23
Total        Sum of Squares       =       15.0632           26
             Standard error of e  =        .18671
Fit          R-squared            =        .94677  R-bar squared =   .93983
Model test   F[  3,    23]        =     136.36447  Prob F > F*   =   .00000
Diagnostic   Log likelihood       =       9.16451  Akaike I.C.   = -3.22043
             Restricted (b=0)     =     -30.43298  Bayes  I.C.   = -3.02846
             Chi squared [  3]    =      79.19498  Prob C2 > C2* =   .00000
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LY|  Coefficient      Error       t    |t|>T*          Interval
--------+--------------------------------------------------------------------
Constant|    1.46773***      .40823     3.60  .0015      .66761   2.26784
      LK|    1.10023**       .43422     2.53  .0186      .24917   1.95128
      LL|    -.11150         .41620     -.27  .7912    -.92723     .70423
     LKL|     .15225         .12734     1.20  .2440    -.09734     .40184
--------+--------------------------------------------------------------------
```

Note that the 'Wald' statistic in the listing below should be ignored, as we are not interested in the joint hypothesis that all four functions are zero.

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
Wald Statistic            =  93486.36534
Prob. from Chi-squared[ 4] =       .00000
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
WaldFcns|  Coefficient      Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
   GAMMA|    4.33936**     1.77146     2.45  .0143      .86736   7.81135
   DELTA|    1.11277***     .41944     2.65  .0080      .29069   1.93485
      NU|     .98872***     .06259    15.80  .0000      .86605   1.11139
  RHO_KL|    2.45416       8.08604      .30  .7615  -13.39419  18.30251
--------+--------------------------------------------------------------------
```

E68.2.1 The Delta Method

Suppose $\mathbf{b}$ is an estimator of parameter vector $\boldsymbol{\beta}$, with asymptotic covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{g}(\mathbf{b})$ is an estimator of J continuous, differentiable functions of $\boldsymbol{\beta}$. We make the following assumptions about the estimators:

- The estimator $\mathbf{b}$ is asymptotically normally distributed with mean vector $\boldsymbol{\beta}$ and asymptotic covariance matrix $\boldsymbol{\Sigma}$,
- The vector function $\mathbf{g}(\boldsymbol{\beta})$ is continuous and continuously differentiable,
- The vector function $\mathbf{g}(\boldsymbol{\beta})$ is not a function of the sample size,
- The conditions underlying the Slutsky theorem are met so that plim $\mathbf{g}(\mathbf{b}) = \mathbf{g}(\boldsymbol{\beta})$.

Then, the vector $\mathbf{g}(\mathbf{b})$ is asymptotically normally distributed with mean $\mathbf{g}(\boldsymbol{\beta})$ and asymptotically normally distributed with asymptotic covariance matrix

$$\text{Asy.Var}[\mathbf{g}(\mathbf{b})] = \boldsymbol{\Gamma}(\boldsymbol{\beta})\,\boldsymbol{\Sigma}\,\boldsymbol{\Gamma}(\boldsymbol{\beta})'$$

where $$\boldsymbol{\Gamma}(\boldsymbol{\beta}) = \partial\mathbf{g}(\boldsymbol{\beta}) / \partial\boldsymbol{\beta}'.$$

The empirical estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are $\mathbf{b}$ and $\mathbf{S}$, the results of estimation. The Jacobian, $\boldsymbol{\Gamma}(\mathbf{b})$ is estimated with $\boldsymbol{\Gamma}(\mathbf{b})$, the matrix of derivatives computed at the estimator of $\boldsymbol{\beta}$. This is precisely the set of computations done with **WALD** described in the preceding section. Your **; Start = b** provides the estimated parameters, **; Var = varb** provides **S** and the list of functions defines $\mathbf{g}(\mathbf{b})$. The matrix of derivatives is estimated by two sided numerical derivatives, which we compute using

$$\frac{\partial g_j}{\partial b_k} = \frac{g_j(\mathbf{b}+\delta_k\mathbf{e}_k) - g_j(\mathbf{b}-\delta_k\mathbf{e}_k)}{2\delta_k}$$

where $$\delta_k = \max(.0001, .00001|\mathrm{b}_k|)$$

and $\mathbf{e}_k$ is a vector with zeros save for a one in the kth position.

E68.2.2 Krinsky and Robb Simulation Method

The method of Krinsky and Robb (1986, 1990) is based on the claimed asymptotic normality of the estimator. The technique involves simply simulating draws from the distribution of the structural parameters, then using the empirical variance of the functions of the draws to estimate the desired variances. Thus, the Krinsky and Robb estimator is

$$\text{Est.Asy.Var}[g(b)] = \frac{1}{R}\sum_{r=1}^{R} [\mathbf{g}(\mathbf{b}_r) - \mathbf{g}(\mathbf{b})][\mathbf{g}(\mathbf{b}_r) - \mathbf{g}(\mathbf{b})]'.$$

We simulate the draws by drawing R primitive K variate standard normal vectors, $\mathbf{v}_r$, then computing

$$\mathbf{b}_r = \text{b} + \mathbf{C}\mathbf{v}_r$$

where **C** is the Cholesky square root of **S**.

Request this with the following addition to the **WALD** command:

; K&R ; Pts = number of draws

The **; Pts** specification is optional. The default is 1,000 draws. Replicability of Krinsky and Robb results can be obtained by setting the seed for the random number generator before the WALD command with

CALC ; Ran (value) $

We repeated the computations of the previous example using the Krinsky and Robb method. The results are below. In two of the cases, the standard errors change substantially. (We note for users of this technique, the second, apparently lesser known Krinsky and Robb paper cited above reports that the large differences found in the first paper could be mostly attributed to a programming error. As the estimates below show, differences do arise. Whether they are substantive enough to warrant reconsideration of the delta method remains to be settled.)

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors
Wald Statistic              =   1333.98675
Prob. from Chi-squared[ 4] =       .00000
Krinsky-Robb method used with 1000 draws
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
WaldFcns|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
   GAMMA|    4.71091**      2.12590     2.22  .0267      .54421   8.87760
   DELTA|    1.09630***      .41582     2.64  .0084      .28131   1.91130
      NU|     .98757***      .06337    15.58  .0000      .86336   1.11178
  RHO_KL|     .07203       20.12783      .00  .9971   -39.37780  39.52185
--------+--------------------------------------------------------------------
```

E68.3 Plotting a Function

The **FPLOT** command is used to examine a general function G(**x,β**). The function may be a simple function (i.e., not a sum of terms), the minimand of a **MINIMIZE** command, or some other kind of sum of terms. **FPLOT** is used to plot the function when one of the parameters varies and the remaining parameters stay fixed at the preset values. One use of this feature might be to examine the slope of a likelihood surface.

Setup for this command is identical to **MINIMIZE** or **NLSQ**. Specify the function to be plotted exactly as if it were to be optimized with one of these commands. (Note, though, that it is not necessary that this actually be a minimization problem, or even that the function you define have a minimum.) With the problem fully specified, add:

; Plot(one of the parameters in ; Labels)
; Limits = range of the parameter over which to plot the function
; Pts = number of points to plot

Two values must be specified for **; Limits**. Starting values must be given for all parameters of the function which appear in the **; Labels** list. The **; Limits** values must bracket the starting value for the variable being plotted. The function will be evaluated at the starting values to ensure that this is possible.

We illustrate with two examples. In the first, we plot the normal CDF over -4 to 4 using 100 points. Since the function does not involve any data, and is not a sum, we set the sample to one observation for the plot. (Otherwise, the function is $\Sigma_i f_i = Nf$ as f does not vary with i.)

```
SAMPLE   ; 1 $
FPLOT    ; Fcn = Phi(x) ; Title = Standard Normal CDF
         ; Labels = x ; Start = 0
         ; Plot(x) ; Limits = -4,4 ; Pts = 100 $
```

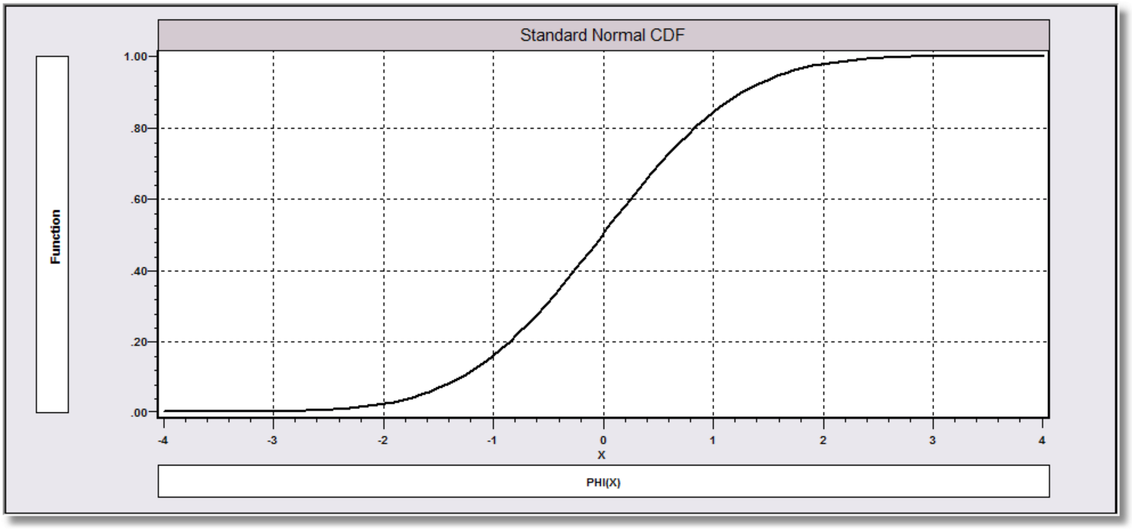


Figure E68.1 Normal CDF Produced with FPLOT

For the second example, we plot the sum of squared residuals from a regression after **REGRESS**, varying one of the slopes between -2 and +2 estimated standard errors of that slope. Note in the experiment below, the sum of squares is computed varying *b*1, the slope on *x*1. The function plotted is the sum of squares for the 100 observations – the **; Fcn** specification defines a computation that is summed over the current sample.

```
CALC        ; Ran (123579) $
SAMPLE      ; 1-100 $
CREATE      ; x1 = Rnn(0,1) ; x2 = x1 + Rnn(0,1)
            ; x3 = .5*x1 + 1.5*x2 + Rnn(0,1) $
CREATE      ; y = x1 - x2 + .5*x3 + Rnn(0,3) $
REGRESS     ; Quietly ; Lhs = y ; Rhs= x1,x2,x3,one $
CALC        ; Lower = b(1) - 2*Sqr(varb(1,1))
            ; Upper = b(1) + 2*Sqr(varb(1,1)) $
FPLOT       ; Fcn = b1 * x1 + b2 * x2 + b3 * x3 + b4
            ; Lhs = y
            ; Start = b ; Labels = b1,b2,b3,b4
            ; Plot(b1) ; Pts = 200 ; Limits = Lower,Upper $
```

NOTE: Specifying a function and an Lhs variable requests a sum of squared residuals. When there is an Lhs variable specified, the function is computed as $\Sigma_i (y_i - function_i)^2$. The function need not be linear; it can be any function definition you like.

Also, the preceding does not recompute the least squares slopes with the new values of *b*1. That is, the sum of squares is not the minimum sum of squares with each *b*1 at the set value and the others recomputed. It is the sum of squares which results when *b*1 is varied and the other slopes remain at their original least squares values. Thus, the minimum in the figure below occurs where *b*1 equals the original least squares value.

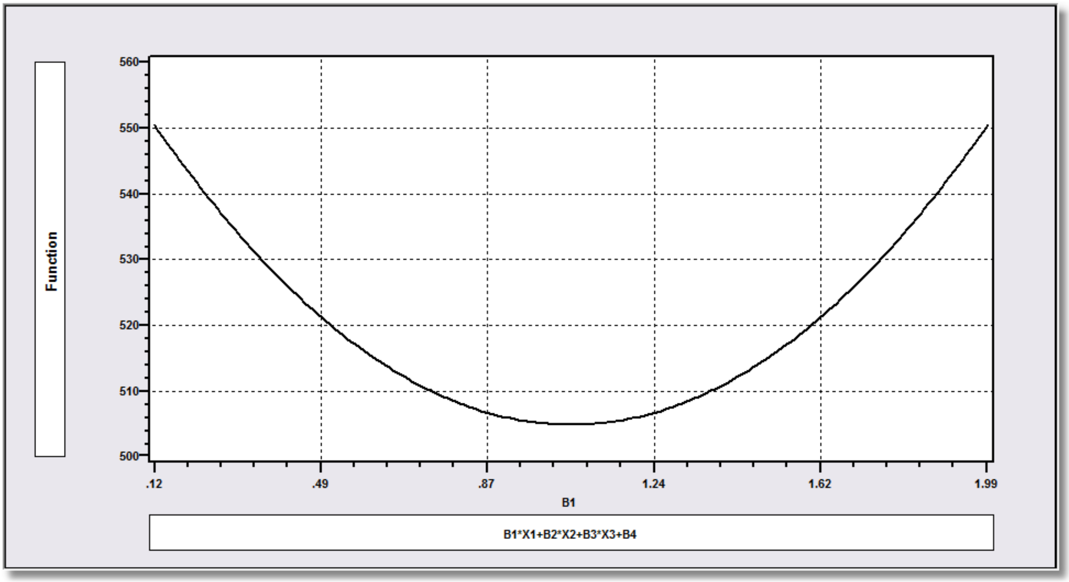


Figure E68.2 Function Plot of OLS Sum of Squared Residuals

E68.3.1 Retaining the Results from FPLOT

Two options for keeping the results are

; Keep = name to retain the function values
and **; Res = name** to keep the values of the changing variable.

E68.3.2 Application – Plotting a Log Likelihood Function

For another illustration of **FPLOT**, we use a small data set based on the Poisson model.

IMPORT $

```
y,x1,x2,x3
1  -0.545   0.160   0.033
0   0.892   0.125   1.476
2   1.647   0.619  -0.262
2   1.749  -1.446   0.310
2   0.362  -0.589  -1.404
0   0.531  -0.606   0.777
2   0.003  -0.800  -0.897
0   0.260   0.597  -0.640
3   1.502  -0.309   0.112
0   0.613   0.273  -0.845
0  -1.028  -0.307  -1.170
2   0.155  -0.262  -0.534
1  -1.795  -2.051  -0.398
0  -1.007   1.974   0.189
1   0.596  -0.493  -1.369
ENDDATA
```

NAMELIST ; x = one,x1,x2,x3 $

The log likelihood for the Poisson model is

$$\log L = \Sigma_i -\log\Gamma(y_i + 1) - \lambda_i + y_i\, \boldsymbol{\beta}'\mathbf{x}_i,\ \ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i),\ \Gamma(y_i + 1) = y_i!.$$

The log likelihood function is maximized. We then plot the function for various values of β_4. The maximizing value of $b4$ is -.46, which is at the top of the hill below.

```
MAXIMIZE    ; Start = 0,0,0,0
            ; Labels = b1,b2,b3,b4
            ; Alg = N
            ; Fcn = -Lgm(y+1) - Exp(b1'x) + y*b1'x $
FPLOT       ; Start = b
            ; Labels = b1,b2,b3,b4
            ; Plot(b4)
            ; Limits =  -1,0
            ; Pts = 100
            ; Fcn = -Lgm(y+1) -Exp(b1'x) + y*b1'x
            ; Title = Poisson Log Likelihood $
```

```
-----------------------------------------------------------------------------
User Defined Optimization
Dependent variable              Function
Log likelihood function        -16.92002
Estimation based on N =      15, K =   4
Inf.Cr.AIC  =     41.8 AIC/N =    2.789
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient     Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
      B1|    -.59537        .45909    -1.30  .1947    -1.49518     .30444
      B2|     .58869*       .30551     1.93  .0540     -.01010    1.18748
      B3|    -.53499        .35309    -1.52  .1297    -1.22703     .15705
      B4|    -.45956        .37586    -1.22  .2214    -1.19623     .27711
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

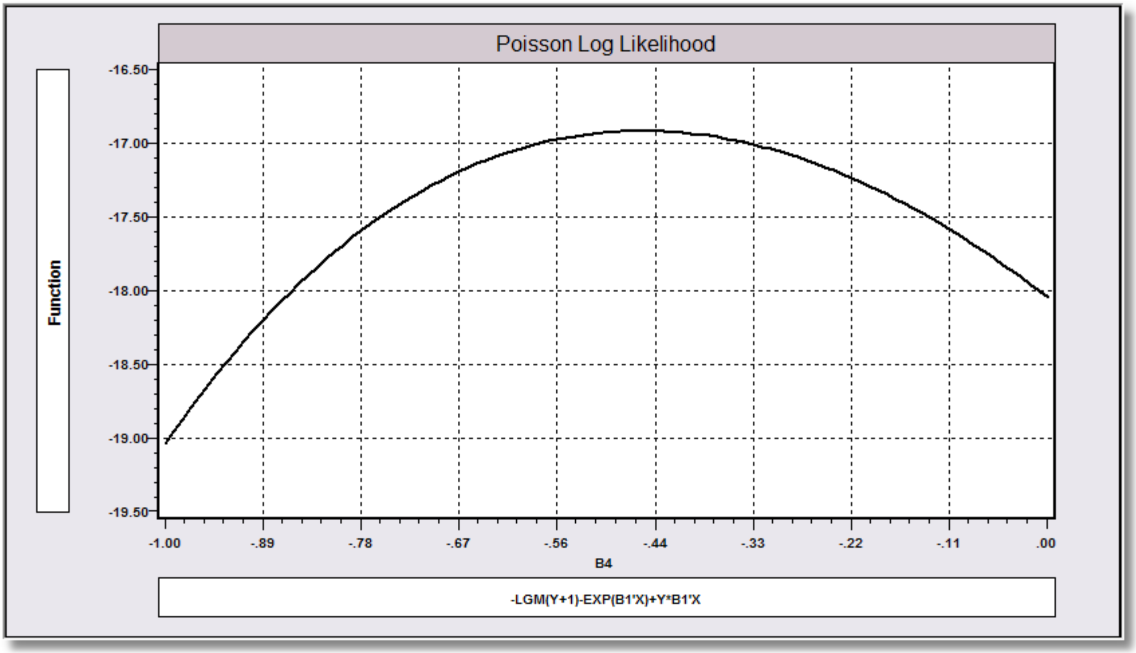


Figure E68.3 Poisson Log Likelihood Profile

E68.4 Evaluating a Function

The command

```
FUNCTION   ; Labels = list of labels
           ; Start = list of values to use to evaluate function
           ; Fcn = the function definition
           ; Keep = variable name $
```

will evaluate the function you define and store the function values in the variable named. You can also save the derivatives of the function with respect to the named parameters as follows: Before the **FUNCTION** command, define the place to store the derivatives with

```
NAMELIST   ; name = a list of variables to store the derivatives in $
```

The namelist must contain the same number of variables as there are parameters in the **; Labels** definition in the **FUNCTION** command. Then, you can add

```
           ; Derivatives = namelist
```

to store the derivatives.

As an example, we use the function in the previous application.

```
MAXIMIZE   ; Start = 0,0,0,0 ; Labels = b1,b2,b3,b4 ; Alg = N
           ; Fcn = -Lgm(y+1) - Exp(b1'x) + y*b1'x $
CREATE     ; dfdb1 = 0 ; dfdb2 = 0 ; dfdb3 = 0 ; dfdb4 = 0 $
NAMELIST   ; dfdb = dfdb1, dfdb2, dfdb3, dfdb4 $
FUNCTION   ; Labels = b1,b2,b3,b4 ; Start = b ; Fcn = Exp(b1'x)
           ; Keep = condmean ; Derivatives = dfdb $
```

This produces the following results:

```
==== [Evaluation of User Specified Function] ====
Function values were saved in variable  CONDMEAN
Derivatives were saved in variable list DFDB
     Parameter  Derivative
   1 B1          DFDB1
   2 B2          DFDB2
   3 B3          DFDB3
   4 B4          DFDB4
```

NOTE: This procedure produces the same results as **SIMULATE** for the function evaluation, and has some overlap with the differentiation routine described for **MAXIMIZE** in the next section. The difference here is that the derivatives saved here are observation specific whereas the next section shows how to obtain the gradient of the function which will usually be a sum of terms.

E68.5 Function Differentiation

If you wish only to obtain the derivatives of a function, set up **MINIMIZE** or **MAXIMIZE** exactly as if you were going to optimize that function *even if the function is not one which could be optimized*. You can set it up as a simple function or as a sum of terms. The function will be evaluated at the points you specify with **; Start = list**. To obtain the first and second derivatives, add

```
; Fix
; Alg = N
```

with no other specification to the command. Note that **; Fix all** and **; Fix = list** are alternative forms that request different computations. With **; Fix** as above, the output from this command will be as follows: The usual output for **MINIMIZE** or **MAXIMIZE** will be given. With the exception of the reported function value, this output can be ignored. The function and derivatives are found in scalar *logl* which contains the function value, matrix *gradient* which contains the vector of first derivatives, and *varb* which contains the Hessian, not a covariance matrix.

For example, to obtain the derivatives of the sum of squares for a nonlinear regression model, you might proceed as follows:

```
MINIMIZE    ; Fcn = (y - the function) ^ 2
            ; Labels = ...
            ; Start = ... list
            ; Fix
            ; Alg = N $
```

Another example to illustrate the computation is shown below for a Poisson regression model in the previous section. The function is first maximized. Then, the first and second derivatives are computed at the maximizing values. The results confirm what we would know about the derivatives – they are zero at the maximum. The function and Hessian are also listed. The log likelihood is maximized first.

```
MINIMIZE    ; Start = 4_0 ; Labels = b1,b2,b3,b4 ; Alg = N
            ; Fcn = Exp(b1'x) - y*b1'x $
```

The results of this step appear above Figure E68.3 above. The next command computes the function, gradient and Hessian. Since *b* is already the MLE, the main results are identical to those shown above.

```
MINIMIZE    ; Start = b ; Labels = b1,b2,b3,b4 ; Alg = N
            ; Fcn = Exp(b1'x) - y*b1'x ; Fix $
```

We now display the first derivatives (essentially zero) and the Hessian. The **; Stat** command below verifies that *varb* is the inverse of the matrix used in the first output to obtain the standard errors of the estimated parameters.

```
MATRIX      ; List ; gradient ; varb $
```

```
GRADIENT|              1
--------+--------------
       1|    .478281E-09
       2|    .190823E-08
       3|    .547061E-09
       4|    .435193E-09


    VARB|              1              2              3              4
--------+------------------------------------------------------------
       1|        16.0000        10.5946       -8.26667       -6.97172
       2|        10.5946        20.4281       -5.22819        .228849
       3|       -8.26667       -5.22819        12.5434        2.34921
       4|       -6.97172        .228849        2.34921        12.0844
```

MATRIX ; Stat (b,<varb>,x) $

```
-----------------------------------------------------------------------------
Number of observations in current sample =      15
Number of parameters computed here       =       4
Number of degrees of freedom             =      11
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Matrix|  Coefficient       Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|    -.59537         .45909    -1.30  .1947     -1.49518     .30444
      X1|     .58869*        .30551     1.93  .0540      -.01010    1.18748
      X2|    -.53499         .35309    -1.52  .1297     -1.22703     .15705
      X3|    -.45956         .37586    -1.22  .2214     -1.19623     .27711
--------+--------------------------------------------------------------------
```

E68.6 Integration

You can compute integrals of the form

$$F = \int_L^U f(x)\mathrm{d}x.$$

The function may be any function that you can set up with the **MINIMIZE/MAXIMIZE** command, which would include

$$F = \int_L^U f(\theta_1, \theta_2, ..., x)\mathrm{d}x,$$

where we pull one of the variables out of the list and name it x, or, even

$$F = \int_L^U \Sigma_i w_i F(\mathbf{x}_i, \theta_1, \theta_2, ..., z)\mathrm{d}z.$$

In this case, we simply set up the sum of terms, and name one of the variables in the function ‘z.’

Integration is set up essentially the same as **FPLOT**. The command is

FINTEGRATE ; Fcn = the function
; Labels = ... the full set of labels for function arguments
; Start = the values of all the parameters
; Limits = L,U
; Pts = number of points in grid (see below)
; Vary (variable) $

The last part specifies the variable of integration. This must be one of the variables in your **; Labels** list. The starting values given provide values for all the variables in the function. One of them will be an interior value for the variable being varied. Suppose, for example, your function is over $x1$,$x2$ and $x3$, with $x1 = 9$, $x2$ ranging from 0 to 1 and $x3 = -.7$. Then you might use **; Start = 9,.5,-.7**. The value for the changing variable is given so that it can be assessed before the computation is done whether the function evaluation is possible at all.

E68.6.1 The Trapezoid Rule

The method used is the trapezoid rule with first order correction (which is also the Newton-Cotes method). The integral is approximated as follows: We divide the range from L to U into N-1 equal length intervals. The interval length is

$$\Delta = (U - L) / (N\text{-}1).$$

Then,

$$F = \int_L^U f(x)\mathrm{d}x \approx \frac{1}{2}\Delta[f(L) + f(U)] + \sum_{i=2}^{N-1} \Delta f(L + i\Delta)$$

For example, a moderately accurate method of obtaining probabilities for the normal distribution would be

SAMPLE ; 1 $
FINTEGRATE ; Fcn = N01(x)
; Labels = x
; Start = some intermediate value
; Limits = Lower,Upper ; Vary(x)
; Pts = 100 $

The result of the calculation will be displayed in your output and kept in a scalar named *integral*.

The accuracy of the approximation improves with the number of points. However, the amount of computation does as well. If the function is a simple function which is not a sum of terms, several hundred points may not be overly time consuming. But, if you are summing functions of many observations, you may want to limit the number of points.

To continue the example above, using

CALC ; List ; Phi(1) - Phi(-1) $

produces an answer of .68269. Using the method above with 100 points and limits of -1,1 produces a value of .68267.

The commands are:

```
SAMPLE       ; 1 $
FINTEGRATE   ; Fcn = N01(t)
             ; Labels = t
             ; Start = 0
             ; Limits = -1,1 ; Vary(t)
             ; Pts = 100 $
```

```
+----------------------------------------------------+
| Function integration:                              |
| Grid is  100 points in [    -1.000 to      1.000] |
| Value of the integral is          .68267           |
+----------------------------------------------------+
```

E68.6.2 Quadrature

For some integrals which are improper in both tails and of the form

$$\int_{-\infty}^{+\infty} f(x)e^{-x^2}dx$$

the value can be well approximated by Hermite quadrature:

$$\int_{-\infty}^{+\infty} f(x)e^{-x^2}dx \approx \sum_{s=-1,+1}\sum_{m=1}^{K} w(m)[f(s\times z(m))]$$

where $w(m)$ is a weight and $z(m)$ is the abscissa of the Hermite polynomial. (See Abramovitz and Stegun(1971).) A 20 point quadrature is used (K=10) by default. For integrals which can be written

$$\int_{0}^{\infty} f(x)e^{-x}dx$$

we can use Gauss-Laguerre quadrature, instead,

$$\int_{0}^{+\infty} f(x)e^{-x}dx \approx \sum_{m=1}^{K} w(m)f(z(m)) .$$

For this procedure, a 40 point quadrature is used. To request these procedures, the commands are

```
FINTEGRAL  ; Fcn = ... as before
           ; Start = ... as before
           ; Labels = ... as before
           ; Vary (variable that is varying)
           ; [integral type] $
```

where [integral type] is **HR1** for the Hermite quadrature or **GL1** for the Gauss-Laguerre quadrature.

You can set the number of points for the Hermite quadrature by including

; Hpt = n

where *n* is one of (4,6,8,10,12,14,16,18,20,24,32,40,64,96). You can reset the number of points for the Gauss-Laguerre quadrature to one of (2,3,4,5,6,7,8,9,10,12,15,20,40,68) by using

; Lpt = n

More points provides a more accurate approximation, but takes longer to compute. If you are not summing over a sample to compute the function, the difference will be trivial, and you should use a large value.

To illustrate the procedure, we will approximate the gamma function over the range 0.5 to 2.5 with a 40 point Gauss-Laguerre quadrature. We compute the function at 21 points, then list and plot the results. For comparison, the value of the function computed with the internal program is given as well. (Note that for small P, the function is not very well approximated.)

```
SAMPLE        ; 1 $
CALC          ; i = 0 $
MATRIX        ; gammafn = Init(21,1,0) ; p = gammafn $
PROCEDURE
CALC          ; ap = .5 + i * .1 $
FINTEGRAL     ; Fcn = x^(ap - 1) ; Labels = x ; Start = 1 ; GL1 ; Vary(x) $
CALC          ; i = i + 1 ; List ; ap ; integral ; Gma(ap) $
MATRIX        ; p(i) = ap ; gammafn(i) = integral $
ENDPROCEDURE
EXECUTE       ; n = 21 $
MPLOT         ; Lhs = p ; Rhs = gammafn ; Grid ; Fill
              ; Title = Gamma Function Values $
```

Each execution of the procedure produces a result such as the following: This is the first one.

```
+-----------------------------------------------------+
| Function integration:                               |
| Laguerre quadrature; 40 pts from    0 to +inf       |
| Value of the integral is         1.63524            |
+-----------------------------------------------------+
+--------------------------------------+
| Listed Calculator Results            |
+--------------------------------------+
 AP      =       .500000
 INTEGRAL=      1.635238
 Result  =      1.772454
```

This is an interior value.

```
+-----------------------------------------------------+
| Function integration:                               |
| Laguerre quadrature; 40 pts from    0 to +inf       |
| Value of the integral is          .88788            |
+-----------------------------------------------------+
+-------------------------------------+
| Listed Calculator Results           |
+-------------------------------------+
 AP       =      1.400000
 INTEGRAL=        .887879
 Result  =        .887264
```

For brevity, the displays are omitted below. Following each integration report are the reported values of the quadrature approximation and the (better) internal function value for the gamma function.

```
[CALC] AP      =       .5000000 [CALC] INTEGRAL=      1.5790359
[CALC] *Result*=      1.7724539
[CALC] AP      =       .6000000 [CALC] INTEGRAL=      1.3979428
[CALC] *Result*=      1.4891922
[CALC] AP      =       .7000000 [CALC] INTEGRAL=      1.2568597
[CALC] *Result*=      1.2980553
[CALC] AP      =       .8000000 [CALC] INTEGRAL=      1.1475006
[CALC] *Result*=      1.1642297
[CALC] AP      =       .9000000 [CALC] INTEGRAL=      1.0635047
[CALC] *Result*=      1.0686287
[CALC] AP      =      1.0000000 [CALC] INTEGRAL=      1.0000000
[CALC] *Result*=      1.0000000
[CALC] AP      =      1.1000000 [CALC] INTEGRAL=       .9532685
[CALC] *Result*=       .9513507
[CALC] AP      =      1.2000000 [CALC] INTEGRAL=       .9204909
[CALC] *Result*=       .9181688
[CALC] AP      =      1.3000000 [CALC] INTEGRAL=       .8995511
[CALC] *Result*=       .8974707
[CALC] AP      =      1.4000000 [CALC] INTEGRAL=       .8888876
[CALC] *Result*=       .8872638
[CALC] AP      =      1.5000000 [CALC] INTEGRAL=       .8873809
[CALC] *Result*=       .8862270
[CALC] AP      =      1.6000000 [CALC] INTEGRAL=       .8942694
[CALC] *Result*=       .8935153
[CALC] AP      =      1.7000000 [CALC] INTEGRAL=       .9090861
[CALC] *Result*=       .9086387
[CALC] AP      =      1.8000000 [CALC] INTEGRAL=       .9316136
[CALC] *Result*=       .9313838
[CALC] AP      =      1.9000000 [CALC] INTEGRAL=       .9618524
[CALC] *Result*=       .9617658
[CALC] AP      =      2.0000000 [CALC] INTEGRAL=      1.0000000
[CALC] *Result*=      1.0000000
[CALC] AP      =      2.1000000 [CALC] INTEGRAL=      1.0464398
[CALC] *Result*=      1.0464858
[CALC] AP      =      2.2000000 [CALC] INTEGRAL=      1.1017376
[CALC] *Result*=      1.1018025
[CALC] AP      =      2.3000000 [CALC] INTEGRAL=      1.1666450
[CALC] *Result*=      1.1667119
[CALC] AP      =      2.4000000 [CALC] INTEGRAL=      1.2421099
[CALC] *Result*=      1.2421693
[CALC] AP      =      2.5000000 [CALC] INTEGRAL=      1.3292927
[CALC] *Result*=      1.3293404
Maximum repetitions of PROC
```

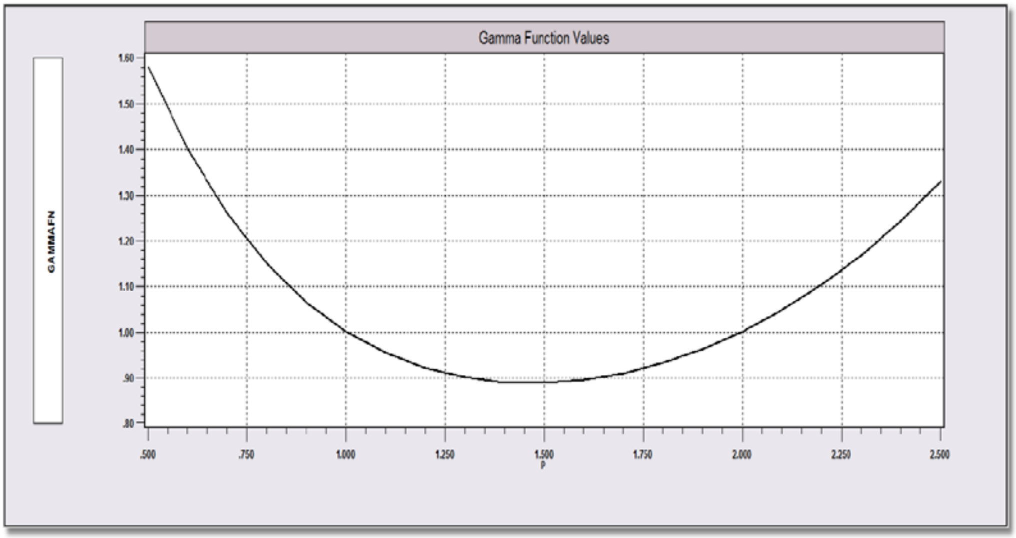


Figure E68.4 Gamma Function

E68.6.3 Monte Carlo Integration

Consider an integral that is of the form of an expected value of a function

$$f(.) = \int_u g(u)\, f(u) d\iota$$

where $f(u)$ is the density of the random variable u. In particular, a case typically of interest arises when u has a normal distribution with mean zero and variance σ^2, in which case the preceding can be written

$$f(.) = \int_u g(\sigma v) \phi(v) dv$$

where $v \sim N[0,1]$. The 'Monte Carlo' integration method amounts to replacing the integration above with averaging over a set of draws from the assumed population;

$$\hat{f}(.) = \frac{1}{R} \sum\nolimits_{r=1}^{R} g(u_r).$$

So long as the function is smooth and well behaved, an appeal to the law of large numbers will produce plim $\hat{f}(.) = f(.)$. You can use this method of integration by using

; Simulation (or just ; Sim)
; Pts = the number of draws, R

in your **FINTEGRAL** command. The number of points may be from 10 to 10,000. Since you are not integrating over a range, you may omit the **; Limits = Lower, Upper** from the command when you do Monte Carlo integration.

As an example, consider the integral

$$f(u) = \int_{-\infty}^{\infty} \Phi(u)\phi(u)du\ .$$

If we randomly draw values of u, the values of $\Phi(u)$ should vary randomly in the interval from zero to one, with center at one half. The expected value should be one half. (That is the exact theoretical result.) We use the command

```
SAMPLE      ; 1 $
CALC        ; Ran (12347) $
FINTEGRAL   ; Labels = z ; Start = 0 ; Vary(z) ; Simulation
            ; Pts = 3000 ; Fcn = Phi(z) $
```

```
+---------------------------------------------------+
| Function integration:                             |
| Monte Carlo integral with 3000 draws              |
| Value of the integral is          .50013          |
+---------------------------------------------------+
```

E68.7 Finding the Roots of a Function

To find the zeros of a function, use

```
SOLVE       ; Labels =  list of labels for parameters
            ; Start  =  an interior point of the function for nonvarying parameters
            ; Fcn    =  the function definition
            ; Vary      (the label of the parameter that varies)
            ; Limits =  low, high = the range over which the function is computed
            ; Pts    =  the number of points in the range to evaluate
            ; Plot      this is optional  $
```

The last specification will produce a plot of the function in the dimension of the label of the varying parameter holding the others fixed at their start values. This procedure uses a simple grid search to search the range from low to high for the values at which f(x) = 0. When two points in the grid search bracket a zero value, several Newton iterations are used to find the actual x at which f(x)=0.

The following builds on the earlier example to create a function to analyze:

```
MAXIMIZE    ; Start = 0,0,0,0
            ; Labels = b1,b2,b3,b4
            ; Alg = N
            ; Fcn = -Lgm(y+1) - Exp(b1'x) + y*b1'x $
SOLVE       ; Labels = b1,b2,b3,b4
            ; Start = b
            ; Fcn =  -Lgm(y+1) -Exp(b1'x) + y*b1'x + 18/15
            ; Vary (b1)  ; Limits = -1.2,.1 ; Pts  = 500 ; Plot $
```

The function is the one plotted in Figure E68.3 plus 18/15 which produces two zeros in the range specified. (Your function will be better defined – we have put this together to produce an uncomplicated example.) The search locates two roots in the range specified.

```
Newton iterations to search for any root near    -.595370
Iteration  B1        Function   Newton Step
Iteration  B1        Function   Newton Step
Found  2 roots in the range B1      =   -1.2000 to      .1000
    -.986736    -.249137
```

The figure shows the results of the search.

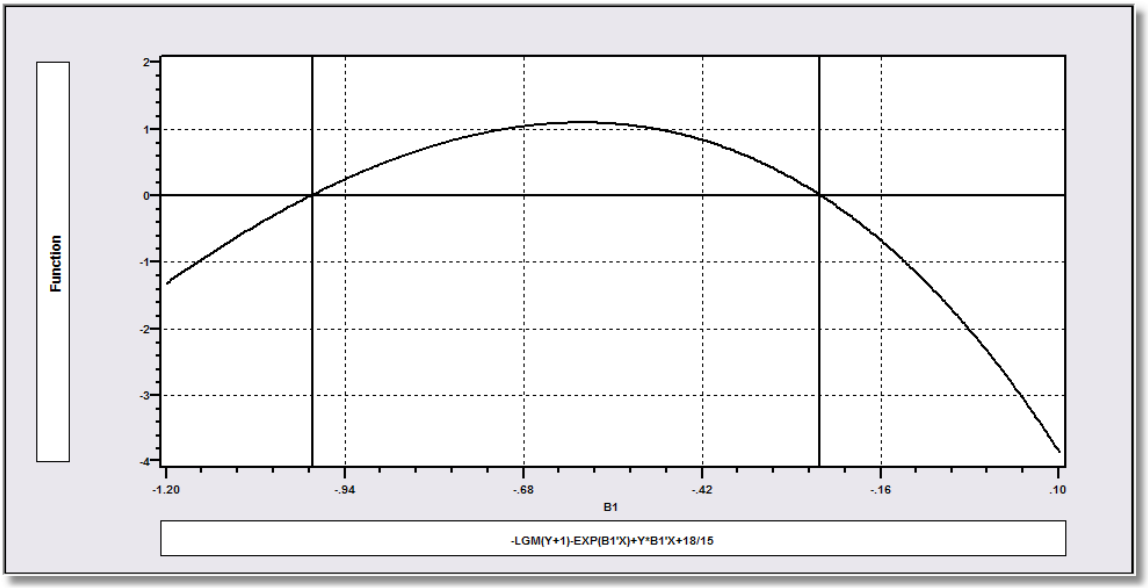


Figure E68.5 Roots of a Function

LIMDEP 10 *Econometric Modeling Guide* Index