NLOGIT

VERSION 6

Reference 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
6/16 Carr Street
Waverton NSW 2060
Australia
Tel: +61 (0) 418-433-057
Email: hgroup@hensher.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.

Preface

NLOGIT is used for estimation of discrete multinomial choice models. The program is a superset of *LIMDEP* Version 11; *NLOGIT* 6 is *LIMDEP* 11 plus the **NLOGIT** command (and numerous variants) which invokes the multinomial choice estimators. The centerpiece of the estimation and analysis package is the multinomial logit model. The many variations include multinomial probit, latent class logit, and, most importantly, the many forms of the mixed (or, random parameters) logit model, including the most recently developed formulation, the generalized mixed logit model. No other program supports as wide a variety of multinomial choice model estimators and post estimation analysis tools.

NLOGIT has pioneered many of the methods described in this manual, such as several forms of the mixed logit model, the attribute nonattendance model and estimation of mixed models in WTP space. This version continues the ongoing collaboration of William Greene (Econometric Software, Inc.) and David Hensher (Econometric Software, Australia.) Recent developments, especially the random parameters and generalized mixed logit in its cross section and panel data variants have also benefited from the enthusiastic collaboration of John Rose. We note, the recent practitioner's guide, *Applied Choice Analysis,* 2nd Edition (Hensher, D., Rose, J. and Greene, W., Cambridge University Press, 2015). This is a wide ranging introduction to discrete choice modeling that contains numerous applications developed with *NLOGIT*. This book should provide a useful companion to the documentation for *NLOGIT*.

Econometric Software, Inc.
Plainview, NY 11803
2016

Table of Contents

What's New in Version 6?

NLOGIT 6 takes advantage of all the new features developed in *LIMDEP* 11. Several new models have been added in *NLOGIT* 6. We have also continued to add enhancements to give you greater flexibility in analyzing data and organizing results. The following will summarize the important new developments.

WN1 New Multinomial Choice Models

We have added several major model classes to the package. Some of these are extensions of the random parameters models that are at the forefront of current practice. We have also extended the latent class model. The following are new model classes (commands) in Version 6:

RPLOGIT = random parameters logit
ECLOGIT = error components logit
RRLOGIT = random regret logit
LCRRLOGIT = latent class random regret logit
RPRRLOGIT = random parameters random regret logit
BWLOGIT = best/worst logit
BWMNL = best/worst multinomial logit
BWRPL = best/worst random parameters logit
SEQLOGIT = sequential logit
REMLOGIT = random effects multinomial logit
FEMLOGIT = fixed effects multinomial logit
BLPLOGIT = Berry, Levinsohn and Pakes logit

WN1.1 Fixed Effects in Multinomial Logit Models

The fixed and random effects forms of the multinomial logit model will be useful for data from stated choice experiments in which the same individual is observed in a sequence of choice situations. The common effects model will appear

$$U_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \varepsilon_{ijt} + \alpha_{ij},\ j = 1,\ldots,J,\ t = 1,\ldots,T,\ i = 1,\ldots,n;$$

$$\text{Prob}(\text{choice}_{\text{ijt}}) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ijt} + \alpha_{ij})}{\Sigma_{m=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{x}_{imt} + \alpha_{im})}.$$

In the fixed effects form, it is assumed that the effects carry the unobserved attributes that might be correlated with observed attributes. An example might be a model for brand choice in which the observed data are on attributes, but price data are unavailable. This model builds off the Chamberlain/Rasch formulation of the binary logit model for panel data. Rather than use maximum likelihood estimation for the full system, which will take hours or days for even moderately sized problems, *NLOGIT* uses minimum distance estimation which will take seconds for comparably sized problems.

WN1.2 Random Effects Multinomial Logit Models

The random effects form of the model above adds an assumption whereby the common (fixed) effects are uncorrelated with the observed attributes. In this instance, the effects become random constant terms in the utility functions of the multinomial logit model. There are two forms of this model. If there is no variation across the observed utilities so that the utilities are determined only by the characteristics of the individual (age, gender, income, etc.) and the random term, then the model that applies is a 'multinomial logit,' estimated using the **MLOGIT ; Fcn = remnl** command. In the stated preference case, where there are attributes of the choices in the utility functions, then the model takes the form of a mixed logit model with random alternative specific constants, and is estimated using the new **REMLOGIT** command.

WN1.3 Random Regret Logit Model

The random regret model was introduced in Version 5 as an alternative to maximum random utility. The base case form substitutes the minimum random regret formulation for the familiar maximum random utility for all the attributes in the utility function. Three extensions of the random regret model have been added in *NLOGIT* 6:

- The MNL may be specified with a mixture of attributes valued by minimum regret and attributes valued by maximum utility.
- The mixed (random parameters) model may be specified as minimum regret or maximum utility.
- The latent class logit model may be specified as minimum regret or maximum utility.

WN1.4 Sequential Logit Model

The sequential logit model treats the outcome of a multinomial choice as a sequence of transitions. The outcomes are ordered by some construction (which might suggest an ordered choice model as a natural formulation.) Consider the outcome level of schooling, less than high school ($y = 0$), high school ($y =1$), college ($y = 2$) and advanced degree ($y = 3$). The sequential logit model treats the level as a sequence of transitions starting with transition from level 0, which everyone enters so $P_0 = 1$. The probability of stopping at level 0 is the probability of entering and not passing level 1, which is $P_0(1 - P_1)$. The probability of stopping at level 1 (high school) is the probability of leaving level 1 and not leaving level 2, $P_0P_1(1 - P_2)$. For level 2, the probability is, $P_0P_1P_2(1 - P_3)$ and finally for level 3, $P_0P_1P_2P_3$ This is treated as a sequence of conditional binary logit models;

$$\text{Prob(observed level} = j) = \text{Probability(exit levels } j = 1,\ldots,j) \times \text{Probability(no exit at level } j+1).$$

By this formulation, for the outcomes listed above,

$$\text{Prob(observed level} = j) = P_{0,1,\ldots,j}(1 - P_{J+1})$$

WN1.5 Best/Worst Outcome Data

Data for multinomial choice typically come in the form of the indicator of the most preferred outcome. In some survey situations, one might also have data on the least preferred outcome. *NLOGIT* 6 supports one form of this setting, in which data consist of a set of ranks for all of the alternatives. The new best/worst setting is a subset. Data might come in one of three forms. In the typical case, the individual simultaneously indicates his/her most and least preferred outcomes. We label this simply the best/worst case. Alternatively, the individual might indicate his/her most preferred alternative, then their least preferred choice among those that remain. We label this 'best then worst.' Probably less common would be a case in which the least favored alternative is chosen, then the most favored alternative is chosen among those that remain. Intuition suggests the three choice mechanisms would produce the same outcomes, but it is at least possible that some difference could occur.

WN1.6 Berry, Levinsohn and Pakes Random Parameters Logit Model

The Berry, Levinsohn and Pakes model is a random parameters logit model for market share data. The core of the specification is a multinomial logit model for the market shares, with fixed brand effects. The estimation alternates between a GMM estimator for $\boldsymbol{\beta}$ and an intermediate step to reconcile the fixed effects with the observed market share data. The multinomial logit model describes the utility of consumer i's choice of brand j in market period t;

$$U_{ijt} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \alpha_{jt} + \varepsilon_{ijt},$$
$$U_{i0t} = \varepsilon_{i0t} \quad \text{for the outside good.}$$

The assumptions of the model produce the conditional (on $\boldsymbol{\beta}_i$) probability,

$$\text{Prob(consumer } i \text{ chooses brand } j \text{ in market } t) = s_j(\mathbf{X}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_i) = \frac{\exp(\mathbf{x}_{jt}'\boldsymbol{\beta}_i + \alpha_{jt})}{1 + \Sigma_{m=1}^{J}\exp(\mathbf{x}_{mt}'\boldsymbol{\beta}_i + \alpha_{mt})}$$

This is a mixed logit model at this point, though it is based on market share data. Estimation of the model parameters is complicated by two factors:

1. Some attributes are endogenous due to omitted factors. In the BLP application, price is included in the model but features of the models that consumers respond to and which affect the price are not included.

2. The fixed brand effects must be estimated. The estimation procedure alternates between two steps, GMM conditioned on the fixed effects and the method of moments equating market shares to theoretical market shares to calibrate fixed effects.

Further technical details appear in Chapter N29.

WN.2 Model Extensions

A variety of improvements and extensions have been built into the existing models and tools for manipulating results post estimation.

WN2.1 Willingness to Pay

Willingness to pay (WTP) measures for models with variable coefficients such as mixed and generalized logit models are generally saved as the rows of a matrix *wtp_i* post estimation. They may now be saved directly as variables in the data area.

WN2.2 Model Simulation

After estimation, the simulator can be used to examine scenarios such as changes in attribute prices of values on the market shares predicted by the model. You may also plot the probabilities that result from your scenario. In the figure below, mode choice probabilities from a choice model are plotted for values of the cost of air ranging from 50 to 100.

Simulations can be based on a variety of different configurations, including setting an attribute at a specific value, setting several attributes, or changing one or more attributes by specific values. In the scenario below, the cost of air is fixed at 75 to examine the impact of this setting compared to the observed values in the sample:

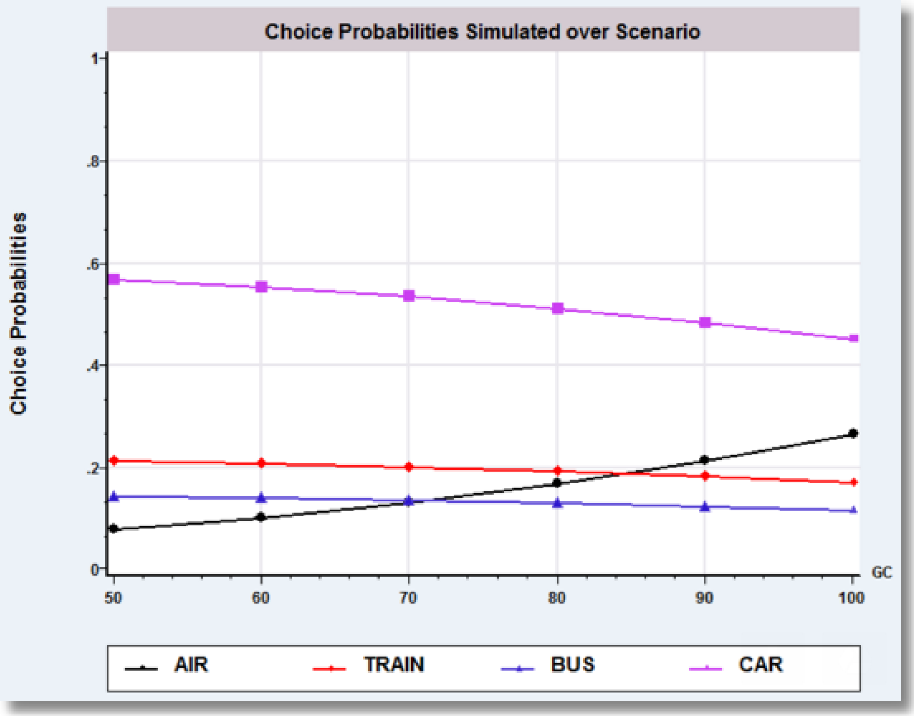


Figure WN.1 Predicted Choice Probabilities

```
------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------   ------------------- ---------
GC         AIR                              Fix base at new vlu    75.000
------------------------------------------------------------------------
The simulator located    210 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 | 11.820    25 |-15.799%      -33 |
|TRAIN     | 30.000    63 | 34.773    73 |  4.773%       10 |
|BUS       | 14.286    30 | 18.114    38 |  3.828%        8 |
|CAR       | 28.095    59 | 35.294    74 |  7.199%       15 |
|Total     |100.000   210 |100.000   210 |   .000%        0 |
+----------+--------------+--------------+------------------+
```

WN2.3 Estimated Elasticities and Partial Effects

There are numerous options for computing and displaying partial effects and elasticities in estimated choice models. New features added allow the user to export cross elasticity matrices to *Excel* and text editors. We have developed standard errors and confidence intervals for elasticities based on the method of Krinsky and Robb.

WN2.4 Robust Covariance Matrix

Robust covariance matrices may be computed for all multinomial choice model estimators. The 'cluster robust' covariance matrix may also be computed for stated choice data (though a formal estimator such as RPLOGIT is likely to be a preferable approach).

WN2.5 Random Number Generator

Simulation based calculations in *NLOGIT* use the Mersenne Twister. L'Ecuyer's generator used in earlier versions remains available. You may also modify the sampling procedure by combining the generator with shuffling, Modified Latin Hypercube sampling or antithetic draws.

WN2.6 Posterior Estimates from Latent Class Models

The latent class models create results for separate classes as well as conditional estimates for each individual in the sample. New **CREATE** and **MATRIX** commands provide access to the specific terms for simulations and other post estimation analyses. For example, the following fits a three class LC model to a stated choice experiment, then obtains the predicted probabilities for each choice situation based on the conditional (posterior) estimates:

```
NAMELIST    ; x = gc,invc,invt,ttme $
LCLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = x ; Pts = 3
            ; Pds = 3 ; Parameters $
CREATE      ; i =Trn(12,0) $
CREATE      ; utility = Mbx(beta_i,i,x) $
CREATE      ; mnlprob = mnl_probs(utility,Set=4) $
```

The same capability is provided for the random parameters models. (The new functions can be used for all of the choice models, but it would usually be needed only for these two.)

For latent class models, the conditional (posterior) class probabilities may also be saved in the data set in addition to the matrix *classp_i* as in previous versions.

Elasticities and other partial effects and simulations for latent class models are based on average coefficient vectors. Some applications involve analysis of the classes (based on the class specific parameter values). A convenient method of setting up the simulation for the class specific components of the LC model is provided.

WN2.7 Coefficients in RP Models

It can be inconvenient to specify random ASCs in an RP model as they take the name partly from the names given to the alternatives. To simplify the specification, the constants may be assigned the name of the alternative defined in the utility function. For example,

```
RPLOGIT ; …  ; Choices = air,train,bus,car
             ; … ; Rh2 = one
             ; Fcn = air(n),train(n),bus(n)
```

specifies what amounts to a random effects model. (As usual, one of the constants must be set to zero.) The lognormal distribution is used in RP models to constrain the sign of a coefficient. The specification will appear as in **; Fcn = price(l)**. However, as stated here, this forces the coefficient to be positive. To force it to be negative the well known 'trick' is to multiply the price variable by minus one, then force the coefficient to be positive as usual. Alternatively, the sign can be built into the coefficient by using **; Fcn = -price(l)**.

WN2.8 Simplified WALD Command

A simplified syntax is provided for the **WALD** command when it is based on the previously estimated model. In this case, the parameter values (**; Parameters = name**) and covariance matrix (**; Covariance = name**) may be omitted. The defaults *b* and *varb* will then be assumed. The labels are optional as well. If they are omitted, default labels, *b*1, *b*2, …, *bK* are used, where *K* is the number of elements in *B*. Labels may be provided if desired. For an example, the following two **WALD** commands are identical:

```
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car ; Rhs = gc,invc,invt,ttme $
WALD      ; Fn1 = wtp = -b3 / b1 $
WALD      ; Labels = bcost, bicost, btime, bttme ; Fn1 = wtp = -btime/bcost $
```

N1: Introduction to *NLOGIT* Version 6

N1.1 Introduction

NLOGIT is a package of programs for analyzing data on multinomial choice. The program, itself, consists of a special set of estimation and analysis routines, specifically for this class of models and style of analysis. *LIMDEP* provides the foundation for *NLOGIT*, including the full set of tools used for setting up the data, such as importing data files, transforming variables (e.g., **CREATE**), and so on. *NLOGIT* is created by adding a set of capabilities to *LIMDEP*. The notes below describe this connection in a bit more detail.

N1.2 The *NLOGIT* Program

NLOGIT adds one (very powerful) command to *LIMDEP*,

NLOGIT ; … specification of choice variable
; … specification of choice model behavioral equations
; … definition of choice modeling framework (e.g., nested logit)
; … other required and optional features $

The **NLOGIT** command is the gateway to the large set of features that are described in this *NLOGIT Reference Guide*. All other features and commands in *LIMDEP* are provided in the *NLOGIT* package as well.

The estimation results produced by *NLOGIT* look essentially the same as by *LIMDEP*, but at various points, there are differences that are characteristic of this type of modeling. For example, the standard data configuration for *NLOGIT* looks like a panel data set analyzed elsewhere in *LIMDEP*. This has implications for the way, for example, model predictions are handled. These differences are noted specifically in the descriptions to follow. But, at the same time, the estimation and post estimation tools provided for *LIMDEP*, such as matrix algebra and the hypothesis testing procedures, are all unchanged. That is, *NLOGIT* is *LIMDEP* with an additional special command.

N1.3 *NLOGIT* and *LIMDEP* Integration and Documentation

NLOGIT 6 is a suite of programs for estimating discrete choice models that are built around the logit and multinomial logit form. This is a superset of *LIMDEP*'s models – *NLOGIT* 6 is all of *LIMDEP* 11 plus the set of tools and estimators described in this guide. *LIMDEP* 11 contains the **CLOGIT** command and the estimator for the 'conditional logit' (or multinomial logit) model. **CLOGIT** is the same as the most basic form of the **NLOGIT** command described in Chapter N19.

The full set of features of *LIMDEP* 11 is part of this package. We assume that you will use the other parts of *LIMDEP* as part of your analysis. To use *NLOGIT*, you will need to be familiar with the *LIMDEP* platform. At various points in your operation of the program, you will encounter *LIMDEP*, rather than *NLOGIT* as the program name, for example in certain menus, dialog boxes, window headers, diagnostics, and so on. Once again, these result from the fact that in obtaining *NLOGIT*, you have installed *LIMDEP* plus some additional capabilities. If you are uncertain which program is actually installed on your computer, go to the About box in the main menu. It will clearly indicate which program you are operating.

This *NLOGIT Reference Guide* provides documentation for some aspects of discrete choice models in general but is primarily focused on the specialized tools and estimators in *NLOGIT* 6 that extend the multinomial logit model. These include, for example, extensions of the multinomial logit model such as the nested logit, random parameters logit, generalized mixed logit and multinomial probit models. This guide is primarily oriented to the commands added to *LIMDEP* that request the set of discrete choice estimators. However, in order to provide a more complete and useful package, Chapters N4-N17 in the *NLOGIT Reference Guide* describe common features of *LIMDEP* 11 and *NLOGIT* 6 that will be integral tools in your analysis of discrete choice data, as shown, for example, in many of the examples and applications in this manual.

Users will find the *LIMDEP* documentation, the *LIMDEP Reference Guide* and the *LIMDEP Econometric Modeling Guide*, essential for effective use of this program. It is assumed throughout that you are already a user of *LIMDEP*. The *NLOGIT Reference Guide*, by itself, will not be sufficient documentation for you to use *NLOGIT* unless you are already familiar with the program platform, *LIMDEP*, on which *NLOGIT* is placed.

The *LIMDEP* and *NLOGIT* documentation use the following format: The *LIMDEP Reference Guide* chapter numbers are preceded by the letter 'R.' The *LIMDEP Econometric Modeling Guide* chapter numbers are preceded by 'E,' and the *NLOGIT Reference Guide* chapter numbers are preceded by 'N.'

N1.4 Discrete Choice Modeling with *NLOGIT*

NLOGIT is a set of tools for building models of discrete choice among multiple alternatives. The essential building block that underlies the set of programs is the random utility model of choice,

U(choice 1) = f_1 (attributes of choice 1, characteristics of the chooser, ε_1,**v**,**w**)
...
U(choice J) = f_J (attributes of choice J, characteristics of the chooser, ε_J,**v**,**w**)

where the functions on the right hand side describe the utility to an individual decision maker of J possible choices, as functions of the attributes of the choices, the characteristics of the chooser, random choice specific elements of preferences, ε_j, that may be known to the chooser but are unobserved by the analyst, and random elements **v** and **w**, that will capture the unobservable heterogeneity across individuals. Finally, a crucial element of the underlying theory is the assumption of utility maximization,

The choice made is alternative j such that U(choice j) > U(choice q) $\forall\ q \neq j$.

The tools provided by *NLOGIT* are a complete suite of estimators beginning with the simplest binary logit model for choice between two alternatives and progressing through the most recently developed models for multiple choices, including random parameters, mixed logit models with individual specific random effects for repeated observation choice settings and the multinomial probit model.

Background theory and applications for the programs described here can be found in many sources. For a primer that develops the theory for multinomial choice modeling in detail and presents many examples and applications, all using *NLOGIT*, we suggest

Hensher, D., Rose, J. and Greene, W., *Applied Choice Analysis*, 2nd Edition, Cambridge University Press, 2015.

A general reference for ordered choice models, also based on *NLOGIT* is

Greene, W. and Hensher, D., *Modeling Ordered Choices*, Cambridge University Press, 2010.

It is not possible (nor desirable) to present all of the necessary econometric methodology in a manual of this sort. The econometric background needed for *Applied Choice Analysis* as well as for use of the tools to be described here can be found in many graduate econometrics books. One popular choice is

Greene, W., *Econometric Analysis*, 7th Edition, Prentice Hall, 2012.

N1.5 Types of Discrete Choice Models in *NLOGIT*

The order and organization of presentations in this manual are partly oriented to the types of models you will analyze and partly toward the types of data you will use. Chapters N2 and N3 describe discrete choice models including *NLOGIT* model and command summaries.

In Chapters N4-N15, we develop basic choice models that have occupied a large part of the econometrics literature for several decades. The situations are essentially those in which the characteristics of decision makers and the choices that they make form the observational base for the model building. The fundamental building block for all of these, as well as for the more elaborate models, is the *binary choice model*: The structural equations for a model of consumer choice based on a single alternative – either to choose an outcome or not to choose it – are

$$\begin{aligned} U(\mathit{choice}) &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \\ \text{Prob}(\mathit{choice}) &= \text{Prob}(U > 0) \\ &= F(\boldsymbol{\beta}'\mathbf{x}), \\ \text{Prob}(\mathit{not\ choice}) &= 1 - F(\boldsymbol{\beta}'\mathbf{x}), \end{aligned}$$

where $\mathbf{x}$ is a vector of characteristics of the consumer such as age, sex, education, income, and other sociodemographic variables, $\boldsymbol{\beta}$ is a vector of parameters and $F(.)$ is a suitable function that describes the model. The choice of vote for a political candidate or party is a natural application. Models for binary choice are developed at length in Chapters E26-E32 in the *LIMDEP Econometric Modeling Guide*. They will be briefly summarized in Chapters N4-N7 to provide the departure point for the models that follow. Useful extensions of the binary choice model presented in Chapters N8-N12 include models for more than one simultaneous binary choice (of the same type), including *bivariate binary choice models* and *simultaneous binary choice models* and a model for *multivariate binary choices* (up to 20).

The *ordered choice model* described in Chapters N13-N15 describe a censoring of the underlying utility in which consumers are able to provide more information about their preferences. In the binary choice model, decision makers reveal through their decisions that the utility from making the choice being modeled is greater than the utility of not making that choice. In the ordered choice case, consumers can reveal more about their preferences – we obtain a discretized version of their underlying utility. Thus, in survey data, voters might reveal their strength of preferences for a candidate or a food or drink product, from zero (strongly disapprove), one (somewhat disapprove) to, say, four (strongly approve).

The appropriate model might be

$$\begin{aligned} \text{Prob(strongly dislike)} &= \text{Prob}(U \le 0), \\ \text{Prob(dislike)} &= \text{Prob}(0 < U \le \mu_1), \\ \text{Prob(indifferent)} &= \text{Prob}(\mu_1 < U \le \mu_2), \end{aligned}$$

and so on.

We can also build extensions of the ordered choice model, such as a *bivariate ordered choice model* for two simultaneous choices and a sample selection model for nonrandomly selected samples.

The multinomial logit (MNL) model described in Chapters N16 and N17 is the original formulation of this model for the situations in which, as in the binary choice and ordered choice models already considered, we observe characteristics of the individual and the choices that they make. The classic applications are the Nerlove and Press (1973) and Schmidt and Strauss (1975) studies of labor markets and occupational choice. The model structure appears as follows:

$$\text{Prob}[y_i = j] = \frac{\exp\left(\boldsymbol{\beta}'_j \mathbf{x}_i\right)}{\sum_{q=1}^{J_i} \exp\left(\boldsymbol{\beta}'_q \mathbf{x}_i\right)}.$$

Note the signature feature, that the determinants of the outcome probability are the individual characteristics. This model represents a straightforward special case of the more general forms of the multinomial choice model described in Chapters N16 and N17 and in the extensions that follow in Chapters N23-N33.

Chapters N18-N22 document general aspects of operating *NLOGIT*. Chapter N18 describes the way that your data will be arranged for estimation of multinomial discrete choice models. Chapter N19 presents an overview of the command structure for *NLOGIT* models. The commands differ somewhat from one model to another, but there are many common elements that are needed to set up the essential modeling framework. Chapter N20 describes choice sets and utility functions. Chapter N21 describes results that are computed for the multinomial choice models beyond the coefficients and standard errors. Finally, Chapter N22 describes the model simulator. You will use this tool after fitting a model to analyze the effects of changes in the attributes of choices on the aggregate choices made by individuals in the sample.

The models developed in Chapters N23-N33 extend the binary choice case to situations in which decision makers choose among multiple alternatives. These settings involve richer data sets in which the attributes of the alternatives are also part of the observation, and more elaborate models of behavior. The broad modeling framework is the multinomial logit model. With a particular specification of the utility functions and distributions of the unobservable random components, we obtain the canonical form of the logit model,

$$\text{Prob}[y_i = j] = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{iq}\right)},$$

where y_i is the index of the choice made. This is the basic, core model of the set of estimators in *NLOGIT*. (This is the model described in Chapters N16 and N17.)

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 because we do 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 arguments that enter the utility functions, may be the same for all choices, or may be defined differently for each utility function. It is also possible to incorporate characteristics of the individual which do not vary across choices in the utility functions. The estimators described in this manual allow 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. *NLOGIT* allows specification of the model for estimation with ‘ranks data.’ The ranking might be incomplete – ranks data can include ties. An interesting extension of this possibility is ‘Best/Worst’ data in which the chooser indicates both their most and least favored alternatives. In addition, in some settings, the sample data might consist of aggregates for the choices, such as proportions (market shares) or frequency counts. *NLOGIT* will accommodate these cases as well.

The multinomial model has provided a mainstay of empirical research in this literature for decades. But, it does have limitations, notably the assumption of independence from irrelevant alternatives, which limit its generality. Recent research has produced many new, different formulations that have broadened the model. *NLOGIT* contains most of these, all of which remove the crucial IIA assumption of the multinomial logit (MNL) model. Chapters N23-N33 describe these frontier extensions of the multinomial logit model. In brief, these are as follows:

N1.5.1 Random Regret Logit Model

The random regret logit model is a variant of the basic conditional logit model. The form of the utility functions involves more direct comparisons of the attributes of the alternatives. Whereas in the essential MNL model, the utility functions enter the probability linearly in terms of the attributes, so the coefficients are marginal utilities, in the random regret model, the attributes enter the probabilities through the regret functions,

$$R_{ij}(m) = \log[1 + \exp(\beta_m(x_{jm} - x_{im}))]$$

which compare attribute m in alternative j to that attribute in alternative i.

N1.5.2 Scaled Multinomial Logit Model

The scaled multinomial logit model accommodates individual heterogeneity in choice structures through the scaling of the marginal utilities rather than in the location parameters. The coefficients in the scaled MNL model take the form

$$\boldsymbol{\beta}_i = \sigma_i \boldsymbol{\beta}$$

where

$$\sigma_i = \sigma \times \exp(\boldsymbol{\delta}'\mathbf{z}_i + \tau v_i).$$

This is a type of random parameters model; the scale parameter can vary systematically with the observables, $\mathbf{z}_i$ and randomly across individuals with v_i.

N1.5.3 Latent Class and Random Parameters LC Models

The latent class model is a semiparametric approximation to the random parameters multinomial logit model. It embodies many of the features of the RPL model. But, the parameters are modeled as having a discrete distribution with a small number of support points. An alternative interpretation is that individuals are intrinsically sorted into a small number of classes, and information about class membership is extracted from the sample along with class specific parameter vectors. The RP variant, which is new with this version of *NLOGIT*, provides a random parameters logit model (see Section N1.5.7) in each class.

N1.5.4 Heteroscedastic Extreme Value Model

In the base case, multinomial logit model, the assumption of equal variances produces great simplicity in the mathematical results, but at considerable cost in the generality of the model. In particular, if the assumption of equal variances is inappropriate, then the different scaling that is present in the variances will, instead, be forced on the coefficients in the utility functions, in ways that may distort the predictions of the model. The heteroscedastic extreme value model relaxes this assumption by allowing the disturbances in the utility functions each to have their own variance. An extension of this model allows these unequal variances to be dependent on characteristics of the individual as well. Thus, the heteroscedasticity assumption allows us to relax the assumption of equal variances across choices and to incorporate individual heterogeneity in the scaling as well as the 'locations' of the utility functions.

N1.5.5 Multinomial Probit Model

This model is much more general than the multinomial logit model, but until recently was largely impractical because of the multinormal integrals required for estimation. We include an implementation based on the GHK simulation method. The multinomial probit (MNP) model relaxes the assumptions of the MNL model by assuming joint normality for the random terms in the utility functions and by allowing (subject to some identification restrictions) the random terms to have different variances and unrestricted correlations.

N1.5.6 Nested Logit Models

The choice among alternatives could be viewed as taking place at more than one level. For instance, in an application developed in the chapters to follow, we consider transportation mode choice among four alternatives, *car, train, bus*, and *air*. One might view the choice among these four as first between *public* (*bus, train*) and *private* (*air, car*) transportation and then, within each of the two branches of the 'tree,' a second choice of specific mode. This sort of hierarchical choice is handled in the setting of 'nested logit models.' *NLOGIT* allows tree structures to have up to four levels. There are also several specific forms of the nested logit model that enforce the implications of utility maximization on the model parameters.

The nested logit (NL) model described in the previous paragraph is appropriately viewed as a relaxation of the strong IID structure of the multinomial logit model that implies the IIA assumption. In particular, the nested logit model allows for different variances for the groups of alternatives in the branches and for (equal) correlation across the alternatives within a branch. (The earlier interpretation of a decision structure is only superimposed on the nested logit model; it is not the statistical basis of the NL model. The 'decision' part of the model rests at the lowest level, among the alternatives.) The covariance heterogeneity model extends this model a bit further by allowing the variances to depend on variables in the model. The covariance heterogeneity model is a model of heteroscedasticity.

One of the weaker parts of the nested logit specification is the narrow specific assumption of which alternative appears in each branch of the tree. This is often not known with certainty. The generalized nested logit model allows alternatives to appear in more than one branch, in a probabilistic fashion.

N1.5.7 Random Parameters and Nonlinear RP Logit Model

This is the most general model contained in *NLOGIT*. As argued by McFadden and Train (2000), it may be the most flexible form of discrete choice model available generally, as they argue that any behavior pattern can be captured by this model form. The random parameters logit (RPL) model extends the MNL model by allowing its parameters to be random across individuals. The random parameters may have their own data dependent means, their own variances, and may be correlated. By this device, we obtain an extremely general, flexible model. The assumptions about the covariance matrix of the random parameters are transmitted to the random terms in the utility functions so that both the uncorrelatedness and equal variance assumptions are relaxed in the process. This model also allows a panel data treatment, with either random effects or an autoregressive pattern in the random terms. The error components logit model provides a method of incorporating a rich structure of individual specific random effects in the conditional logit and random parameters models. The nonlinear RP variant allows the utility functions in the probability model to be arbitrary nonlinear functions of the data and parameters.

N1.5.8 Error Components Logit and Fixed Effects Models

The error components logit model is essentially a random effects model for the MNL framework. The basic model structure for a repeated choice (panel data) setting would be

$$\text{Prob}[y_{it} = j \mid v_{i1},\ldots,v_{iM}) \;=\; \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ji} + \boldsymbol{\Sigma}_{s=1}^{M} d_{js} v_{is}\right)}{\sum_{q=1}^{J_i} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{iq} + \boldsymbol{\Sigma}_{s=1}^{M} d_{qs} v_{is}\right)},$$

where $v_{i1},\ldots,v_{iM}$ are M individual effects that appear in the J_i utility functions and d_{js} are binary variables that place specific effects in the different alternatives. Different sets of effects, or only particular ones, appear in each utility function, which allows a nested type of arrangement. A fixed effects version of the multinomial logit model would appear as

$$\text{Prob}[y_{it} = j \mid v_{i1},\ldots,v_{iM}) \;=\; \frac{\exp\left(\alpha_{ij} + \boldsymbol{\beta}'\mathbf{x}_{ji}\right)}{\sum_{q=1}^{J_i} \exp\left(\alpha_{iq} + \boldsymbol{\beta}'\mathbf{x}_{iq}\right)},$$

The fixed effects model is estimated using the conditional estimation method proposed in Chamberlain (1984) with a minimum distance estimator to extend the results to the multinomial outcome case.

N1.5.9 Generalized Mixed Logit Model

The generalized mixed logit model is an encompassing model for many of the specifications already noted, and a variety of new specifications as well. The model follows the random parameters model of Section N1.5.7, but adds several layers to the specification of the random parameters. Specifically,

$$\boldsymbol{\beta}_i \;=\; \sigma_i\boldsymbol{\beta} \;+\; [\gamma + (1 - \gamma)\sigma_i]\boldsymbol{\Gamma}\mathbf{v}_i,$$

where σ_i is the heterogeneous scale factor noted in Section N1.5.2, γ is a distribution parameter that moves emphasis to or away from the random part of the model, $\boldsymbol{\Gamma}$ is (essentially) the correlation matrix among the random parameters. As noted, several earlier specifications are special cases.

This form of the RP model allows a number of useful extensions, including estimation of the model in willingness to pay (WTP) space, rather than utility space.

N1.6 Functions of *NLOGIT*

The chapters to follow will describe the different features of *NLOGIT* and the various models it will estimate. The functionality of the program consists of these major features:

- Estimation programs. These are full information maximum likelihood estimators for the collection of models.
- Description and analysis. Model results are used to compute elasticities, marginal effects, and other descriptive measures.
- Hypothesis testing, including the IIA assumption and tests of model specification.
- Computation of probabilities, utility functions, and inclusive values for individuals in the sample.

Simulation of the model to predict the effects of changes in the values of attributes in the aggregate behavior of the individuals in the sample. For example, if *x*% of the sampled individuals choose a particular alternative, how would *x* change if a certain price in the model were assumed to be *p*% higher for all individuals.

N2: Discrete Choice Models

N2.1 Introduction

This chapter will provide a short, thumbnail sketch of the discrete choice models discussed in this manual. *NLOGIT* supports a large array of models for both discrete and continuous variables, including regression models, survival models, models for counts and, of relevance to this setting, models for discrete outcomes. The group of models described in this manual are those that arise naturally from a random utility framework, that is, those that arise from an individual choice setting in which the model is of an individual's selection among two or more alternatives. This includes several of the models described in the *LIMDEP* manual, such as the binary logit and probit models, but also excludes some others, including the models for count data and censored and truncated regression models, and some of the loglinear models such as the geometric regression model.

Two groups of models are considered. The first set are the binary, ordered and multivariate choice models that are documented at length in Chapters E26-E35 in the *LIMDEP Econometric Modeling Guide.* These form the basic building blocks for the *NLOGIT* extensions that are the main focus of this part of the program. Since they are developed in detail elsewhere, we will only provide the basic forms and only the essential documentation here. The second group of estimators are the multinomial logit models and extensions of them that form the group of tools specific to *NLOGIT*.

N2.2 Random Utility Models

The random utility framework starts with a structural model,

$$U(choice\ 1) \ = \ f_1\ (attributes\ of\ choice\ 1,\ characteristics\ of\ the\ consumer,\ \varepsilon_1,\mathbf{v},\mathbf{w}),$$
$$\ldots$$
$$U(choice\ J) \ = \ f_J\ (attributes\ of\ choice\ J,\ characteristics\ of\ the\ consumer,\ \varepsilon_J,\mathbf{v},\mathbf{w}),$$

where $\varepsilon_1,...,\varepsilon_J$ denote the random elements of the random utility functions and in our later treatments, **v** and **w** will represent the unobserved individual heterogeneity built into models such as the error components and random parameters (mixed logit) models. The assumption that the choice made is alternative j such that

$$U(choice\ j) > U(choice\ q)\ \forall\ q \neq j.$$

The observed outcome variable is then

$$y \ = \ \text{the index of the observed choice.}$$

The econometric model that describes the determination of y is then built around the assumptions about the random elements in the utility functions that endow the model with its stochastic characteristics. Thus, where Y is the random variable that will be the observed discrete outcome,

$$\text{Prob}(Y = j) \ = \ \text{Prob}(U(\text{choice } j) > U(\text{choice } q)\ \forall\ q \neq j).$$

The objects of estimation will be the parameters that are built into the utility functions including possibly those of the distributions of the random components and, with estimates of the parameters in hand, useful characteristics of consumer behavior that can be derived from the model, such as partial effects and measures of aggregate behavior.

To consider the simplest example, that will provide the starting point for our development, consider a consumer's random utility derived over a single choice situation, say whether to make a purchase. The two outcomes are 'make the purchase' and 'do not make the purchase.' The random utility model is simply

$$U(\textit{not purchase}) = \boldsymbol{\beta}_0'\mathbf{x}_0 + \varepsilon_0,$$
$$U(\textit{purchase}) = \boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1.$$

Assuming that ε_0 and ε_1 are random, the probability that the analyst will observe a purchase is

$$\begin{aligned}\text{Prob}(\textit{purchase}) &= \text{Prob}(U(\textit{purchase}) > U(\textit{not purchase}))\\ &= \text{Prob}(\boldsymbol{\beta}_1'\mathbf{x}_1 + \varepsilon_1 > \boldsymbol{\beta}_0'\mathbf{x}_0 + \varepsilon_0)\\ &= \text{Prob}(\varepsilon_1 - \varepsilon_0 < \boldsymbol{\beta}_1'\mathbf{x}_1 - \boldsymbol{\beta}_0'\mathbf{x}_0)\\ &= F(\boldsymbol{\beta}_1'\mathbf{x}_1 - \boldsymbol{\beta}_0'\mathbf{x}_0),\end{aligned}$$

where $F(z)$ is the CDF of the random variable $\varepsilon_1 - \varepsilon_0$. The model is completed and an estimator, generally maximum likelihood, is implied by an assumption about this probability distribution. For example, if ε_0 and ε_1 are assumed to be normally distributed, then the difference is also, and the familiar probit model emerges. (The probit model is developed in Chapters E26 and E27.)

The sections to follow will outline the models described in this manual in the context of this random utility model. The different models derive from different assumptions about the utility functions and the distributions of their random components.

N2.3 Binary Choice Models

Continuing the example in the previous section, the choice of alternative 1 (*purchase*) reveals that $U_1 > U_0$, or that

$$\varepsilon_0 - \varepsilon_1 < \boldsymbol{\beta}_1'\mathbf{x}_1 - \boldsymbol{\beta}_0'\mathbf{x}_0.$$

Let $\varepsilon = \varepsilon_1 - \varepsilon_0$ and $\boldsymbol{\beta}'\mathbf{x}$ represent the difference on the right hand side of the inequality – $\mathbf{x}$ 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, a binary choice model applies to the probability that $\varepsilon \leq \boldsymbol{\beta}'\mathbf{x}$, which is the familiar sort of model developed in Chapter E26. Two of the parametric model formulations in *NLOGIT* for binary choice models are the probit model based on the normal distribution:

$$F = \int_{-\infty}^{\beta'\mathbf{x}_i} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt = \Phi(\boldsymbol{\beta}'\mathbf{x}_i),$$

and the logit model based on the logistic distribution

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i).$$

Numerous variations on the model can be obtained. A model with multiplicative heteroscedasticity is obtained with the additional assumption

$$\varepsilon_i \sim \text{ normal or logistic with variance} \propto [\exp(\boldsymbol{\gamma}'\mathbf{z}_i)]^2,$$

where $\mathbf{z}_i$ is a set of observed characteristics of the individual. A model of sample selection can be extended to the probit and logit binary choice models. In both cases, we depart from

$$\text{Prob}(y_i = 1 \mid \mathbf{x}_i) = F(\boldsymbol{\beta}'\mathbf{x}_i),$$

where $F(t) = \Phi(t)$ for the probit model and $\Lambda(t)$ for the logit model,

$$d_i^* = \boldsymbol{\alpha}'\mathbf{z}_i + u_i,\ u_i \sim N[0,1],\ d_i = 1(d_i^* > 0),$$

$y_i, \mathbf{x}_i$ observed only when $d_i = 1$.

where $\mathbf{z}_i$ is a set of observed characteristics of the individual. 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

$$u_i, \varepsilon_i \sim N_2[(0,0),(1,\rho,1)].$$

(We use N_P to denote the P-variate normal distribution, with the mean vector followed by the definition of the covariance matrix in the succeeding brackets.) 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 $\mathbf{z}_i$ as stated above, the bivariate probability for y_i and $\mathbf{z}_i$ 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 Chapter E27.

There are several formulations for extensions of the binary choice models to panel data setting. These include

- **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:** $\text{Prob}(y_{it} = 1|\text{class } j) = F(\boldsymbol{\beta}_j'\mathbf{x}_{it})$,
  $\text{Prob}(\text{class} = j) = G_j(\boldsymbol{\theta},\mathbf{z}_i)$.

where $\mathbf{z}_i$ is a set of observed characteristics of the individual. Other variations include simultaneous equations models and semiparametric formulations.

N2.4 Bivariate and Multivariate Binary Choice Models

The bivariate probit model is a natural extension of the model above in which two decisions are made jointly;

$$y_{i1}^* = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1},\ \ y_{i1} = 1 \text{ if } y_{i1}^* > 0,\ y_{i1} = 0 \text{ otherwise},$$

$$y_{i2}^* = \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \varepsilon_{i2},\ \ y_{i2} = 1 \text{ if } y_{i2}^* > 0,\ y_{i2} = 0 \text{ otherwise},$$

$$[\varepsilon_{i1},\varepsilon_{i2}] \sim N_2[0,0,1,1,\rho],\ -1 < \rho < 1,$$

individual observations on y_1 and y_2 are available for all i.

This model extends the binary choice model to two different, but related outcomes. One might, for example, model y_1 = home ownership (vs. renting) and y_2 = automobile purchase (vs. leasing). The two decisions are obviously correlated (and possibly even jointly determined).

A special case of the bivariate probit model is useful for formulating the correlation between two binary variables. The tetrachoric correlation coefficient is equivalent to the correlation coefficient in the following bivariate probit model:

$$y_{i1}^* = \mu + \varepsilon_{i1},\ \ y_{i1} = 1(y_{i1}^* > 0),$$

$$y_{i2}^* = \mu + \varepsilon_{i2},\ \ y_{i2} = 1(y_{i2}^* > 0),$$

$$(\varepsilon_{i1},\varepsilon_{i2}) \sim N_2[(0,0),(1,1,\rho)].$$

The bivariate probit model has been extended to the random parameters form of the panel data models. For example, a true random effects model for a bivariate probit outcome can be formulated as follows: Each equation has its own random effect, and the two are correlated.

The model structure is

$$y_{it1}^* = \boldsymbol{\beta}_1'\mathbf{x}_{it1} + \varepsilon_{it1} + u_{i1},\ y_{it1} = 1 \text{ if } y_{it1}^* > 0,\ y_{it1} = 0 \text{ otherwise},$$

$$y_{it2}^* = \boldsymbol{\beta}_2'\mathbf{x}_{it2} + \varepsilon_{it2} + u_{i2},\ y_{it2} = 1 \text{ if } y_{it2}^* > 0,\ y_{it2} = 0 \text{ otherwise},$$

$$[\varepsilon_{it1},\varepsilon_{it2}] \sim N_2[0,0,1,1,\rho],\ -1 < \rho < 1,$$

$$[u_{i1}\ ,\ u_{i2}] \sim N_2[0,0,1,1,\theta],\ -1 < \theta < 1.$$

Individual observations on y_{i1} and y_{i2} 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 multivariate probit model is the extension to M equations of the bivariate probit model

$$y_{im}^* = \boldsymbol{\beta}_m'\mathbf{x}_{im} + \varepsilon_{im},\ m = 1,\ldots,M$$

$$y_{im} = 1 \text{ if } y_{im}^* > 0, \text{ and } 0 \text{ otherwise},$$

$$\varepsilon_{im},\ m = 1,\ldots,M \sim N_M[\mathbf{0},\mathbf{R}],$$

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.

N2.5 Ordered Choice Models

The basic ordered choice model can be cast in an analog to our random utility specification. We suppose that preferences over a given outcome are reflected as earlier, in the random utility function:

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

$$\varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ \boldsymbol{\theta} = \text{a vector of parameters},$$

$$E[\varepsilon_i|\mathbf{x}_i] = 0,$$

$$\text{Var}[\varepsilon_i|\mathbf{x}_i] = 1.$$

The consumers are asked to reveal the strength of their preferences over the outcome, but are given only a discrete, ordinal scale, 0,1,...,J. The observed response represents a complete censoring of the latent utility 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 model as stated does embody the strong assumption that the threshold values are the same for all individuals. We will relax that assumption below.) 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 ordered logit model arises if ε_i is assumed to have a logistic distribution rather than a normal. The variance of ε_i is assumed to be the standard, one for the probit model and $\pi^2/6$ for the logit model, since as long as y_i*, $\boldsymbol{\beta}$, and ε_i are all 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 also relax that assumption.) Since the μs are free parameters, there is no significance to the unit distance between the set of observed values of y_i. They merely provide the coding. Estimates are obtained by maximum likelihood. The probabilities which enter the log likelihood function are

$$\text{Prob}(y_i = j) \;=\; \text{Prob}(y_i^* \text{ is in the } j\text{th range}).$$

The model may be estimated either with individual data, with $y_i = 0, 1, 2, ...$ or with grouped data, in which case each observation consists of a full set of $J + 1$ proportions, $p_{i0},...,p_{iJ}$.

There are many variants of the ordered probit model. A model with multiplicative heteroscedasticity of the same form as in the binary choice models is

$$\text{Var}[\varepsilon_i] \;=\; [\exp(\boldsymbol{\gamma}'\mathbf{z}_i)]^2.$$

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 that was given above and, second, a selection equation, a univariate probit model,

$$d_i^* \;=\; \boldsymbol{\alpha}'\mathbf{z}_i + u_i,$$

$$d_i \;=\; 1 \text{ if } d_i^* > 0 \text{ and } 0 \text{ otherwise.}$$

The observation mechanism is

$$[y_i,\mathbf{x}_i] \quad \text{is observed if and only if } d_i \;=\; 1,$$

$$\varepsilon_i,u_i \;\sim\; N_2[0,0,1,1,\rho]; \text{ there is 'selectivity' if } \rho \text{ is not equal to zero.}$$

The general set of panel data formulations is also available for the ordered probit and logit models.

- **Fixed effects:** $\text{Prob}(y_{it} = j) \;=\; F[\mu_j - (\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)] - F[\mu_{j-1} - (\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i)]$,
  α_i correlated with $\mathbf{x}_{it}$.
- **Random effects:** $\text{Prob}(y_{it} = j) \;=\; F[\mu_j - (\boldsymbol{\beta}'\mathbf{x}_{it} + u_i)] - F[\mu_{j-1} - (\boldsymbol{\beta}'\mathbf{x}_{it} + u_i)]$,
  u_i uncorrelated with $\mathbf{x}_{it}$.
- **Random parameters:** $\text{Prob}(y_{it} = j) \;=\; F(\mu_j - \boldsymbol{\beta}_i'\mathbf{x}_{it}) - F(-\mu_{j-1}\boldsymbol{\beta}_i'\mathbf{x}_{it})$,
  $\boldsymbol{\beta}_i \mid i \;\sim\; h(\boldsymbol{\beta}|i)$ with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$.
- **Latent class:** $\text{Prob}(y_{it} = j|\text{class } c) \;=\; F(\mu_{j,c} - \boldsymbol{\beta}_c'\mathbf{x}_{it}) - F(\mu_{j-1,c} - \boldsymbol{\beta}_c'\mathbf{x}_{it})$,
  $\text{Prob}(\text{class} = c) \;=\; G_c(\boldsymbol{\theta},\mathbf{z}_i)$.

The hierarchical ordered probit model, or generalized ordered probit model, relaxes the assumption that the threshold parameters are the same for all individuals. Two forms of the model are provided.

$$\text{Form 1:}\quad \mu_{ij} = \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z}_i),$$

$$\text{Form 2:}\quad \mu_{ij} = \exp(\theta_j + \boldsymbol{\delta}_j'\mathbf{z}_i).$$

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.

Harris and Zhao (2004, 2007) have developed a zero inflated ordered probit (ZIOP) counterpart to the zero inflated Poisson model. The ZIOP formulation would appear

$$\begin{aligned} d_i^* &= \boldsymbol{\alpha}'\mathbf{z}_i + u_i, \quad d_i = 1\,(d_i^* > 0),\\ y_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \quad y_i = 0 \text{ if } y_i^* \le 0 \text{ or } d_i = 0,\\ &\qquad 1 \text{ if } 0 < y_i^* \le \mu_1 \text{ and } d_i = 1,\\ &\qquad 2 \text{ if } \mu_1 < y_i^* \le \mu_2 \text{ and } d_i = 1,\\ &\qquad \text{and so on.} \end{aligned}$$

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_i and ε_i 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_i and ε_i is an optional specification that analysts might want to test. The zero inflation model may also be combined with the hierarchical (generalized) model given above.

The bivariate ordered probit model is analogous to the seemingly unrelated regressions model for the ordered probit case:

$$\begin{aligned} y_{ij}^* &= \boldsymbol{\beta}_j'\mathbf{x}_{ji} + \varepsilon_{ij},\\ y_{ij} &= 0 \text{ if } y_{ij}^* \le 0,\\ &\quad 1 \text{ if } 0 < y_{ij}^* < \mu_1,\\ &\quad 2, \ldots \text{ and so on, } j = 1,2, \end{aligned}$$

for a pair of ordered probit models that are linked by $\text{Cor}(\varepsilon_{i1},\varepsilon_{i2}) = \rho$. 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. Either variable (but not both) may be binary. (If both are binary, the bivariate probit model should be used.) 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 $y_1 = 0,1,\ldots,J_1$ and $y_2 = 0,1,\ldots,J_2$. (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.) For the case noted, the polychoric correlation is the correlation in the bivariate ordered probit model, so it can be estimated just by specifying a bivariate ordered choice model in which both right hand sides contain only a constant term.

N2.6 Multinomial Logit Model

The canonical random utility model suggested by the structure of Section N2.2 is as follows:

$U(\text{alternative } 0) = \boldsymbol{\beta}_0'\mathbf{x}_{i0} + \varepsilon_{i0},$

$U(\text{alternative } 1) = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1},$

...

$U(\text{alternative } J) = \boldsymbol{\beta}_J'\mathbf{x}_{iJ} + \varepsilon_{iJ},$

Observed y_i = choice j if $U_i(\text{alternative } j) > U_i(\text{alternative } q)\ \forall\ q \neq j$.

The 'disturbances' in this framework (individual heterogeneity terms) are assumed to be independently and identically distributed with identical type 1extreme value distribution; the CDF is

$$F(\varepsilon_j) = \exp(-\exp(-\varepsilon_j)).$$

Based on this specification, the choice probabilities are

$$\text{Prob}(\text{choice } j) = \text{Prob}(U_j > U_q),\ \forall\ q \neq j$$

$$= \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_{ij})}{\sum_{q=0}^{J}\exp(\boldsymbol{\beta}_q'\mathbf{x}_{iq})},\ j = 0,\ldots,J.$$

At this point we make a purely semantic distinction between two cases of the model. When the observed data consist of individual choices and (only) data on the characteristics of the individual, identification of the model parameters will require that the parameter vectors differ across the utility functions, as they do above. The study on labor market decisions by Schmidt and Strauss (1975) is a classic example. For the moment, we will call this the *multinomial logit model*. When the data also include attributes of the choices that differ across the alternatives, then the forms of the utility functions can change slightly – and the coefficients can be generic, that is the same across alternatives. Again, only for the present, we will call this the *conditional logit model*. (It will emerge that the multinomial logit is a special case of the conditional logit model, though the reverse is not true.) The conditional logit model is defined in Section N2.7.

The general form of the *multinomial logit* model is

$$\text{Prob}(\text{choice } j) = \frac{\exp(\boldsymbol{\beta}_j'\mathbf{x}_i)}{\sum_{q=0}^{J}\exp(\boldsymbol{\beta}_q'\mathbf{x}_i)},\ j = 0,\ldots,J.$$

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. The data will appear as follows:

- Individual data: y_i coded 0, 1, ..., J,
- Grouped data: $y_{i0}, y_{i1},\ldots,y_{iJ}$ give proportions or shares,

The sequential logit model (command **SEQLOGIT**) is an alternative interpretation of the multinomial outcome of the choice process. In this specification, the values 0, 1, …, J represent levels, such as schooling, in which the outcome represents the highest level reached.

N2.6.1 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 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

$$P_{ijt} = \frac{\exp(\boldsymbol{\beta}'_j\mathbf{x}_{it})}{\sum_{q=0}^{J}\exp(\boldsymbol{\beta}'_q\mathbf{x}_{it})}.$$

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

$$U_{ijt} = \boldsymbol{\beta}_j'\mathbf{x}_{it} + u_{ij} + \varepsilon_{ijt},\ t = 1,...,T_i,\ j = 0,1,...,J, i=1,...,N.$$

The resulting choice probabilities, conditioned on the random effects, are

$$P_{ijt} \mid \alpha_{i1},...,\alpha_{iJ} = \frac{\exp(\boldsymbol{\beta}'_j\mathbf{x}_{it} + u_{ij})}{\sum_{q=0}^{J}\exp(\boldsymbol{\beta}'_q\mathbf{x}_{it} + u_{iq})}.$$

To complete the model, we assume that the heterogeneity is normally distributed with zero means and $(J+1)\times(J+1)$ covariance matrix, $\boldsymbol{\Sigma}$. For identification purposes, one of the coefficient vectors, $\boldsymbol{\beta}_q$, must be normalized to zero and one of the u_{iq}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 us as

$$\mathbf{u}_i = \boldsymbol{\Gamma}\mathbf{v}_i$$

where $\boldsymbol{\Gamma}$ is a lower triangular matrix to be estimated and $\mathbf{v}_i$ is a standard normally distributed (mean vector $\mathbf{0}$, covariance matrix, $\mathbf{I}$) vector.

N2.6.2 A Dynamic Multinomial Logit Model

The preceding random effects model can be modified to produce the dynamic multinomial logit model proposed in Gong, van Soest and Villagomez (2000). The choice probabilities are

$$P_{ijt} \mid u_{i1},...,u_{iJ} = \frac{\exp(\boldsymbol{\beta}'_j \mathbf{x}_{it} + \boldsymbol{\gamma}'_j \mathbf{z}_{it} + u_{ij})}{\sum_{q=1}^{J} \exp(\boldsymbol{\beta}'_q \mathbf{x}_{it} + \boldsymbol{\gamma}'_q \mathbf{z}_{it} + u_{iq})} \quad t = 1,...,T_i,\ j = 0,1,...,J, i=1,...,N,$$

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. The core of the treatment is to treat the first period as an equilibrium, with no lagged effects,

$$P_{ij0} \mid \theta_{i1},...,\theta_{iJ} = \frac{\exp(\boldsymbol{\delta}'_j \mathbf{x}_{i0} + \theta_{ij})}{\sum_{q=1}^{J} \exp(\boldsymbol{\delta}'_q \mathbf{x}_{i0} + \theta_{iq})}, \ t = 0,\ j = 0,1,...,J, i=1,...,N,$$

where the vector of effects, $\boldsymbol{\theta}$, is built from the same primitives as $\mathbf{u}$ in the later choice probabilities. Thus, $\mathbf{u}_i = \boldsymbol{\Gamma}\mathbf{v}_i$ and $\boldsymbol{\theta}_i = \boldsymbol{\Phi}\,\mathbf{v}_i$, 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, $(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_J,\boldsymbol{\gamma}_1,...,\boldsymbol{\gamma}_J,\boldsymbol{\delta}_1,...,\boldsymbol{\delta}_J,\boldsymbol{\Gamma},\Phi)$ 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 $1,2,...,T_i$ probabilities defined earlier.

N2.7 Conditional Logit Model

If the utility functions are conditioned on observed individual, choice invariant characteristics, $\mathbf{z}_i$, as well as the attributes of the choices, $\mathbf{x}_{ij}$, then we write

$$U(\textit{choice } j \textit{ for individual } i) = U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i + \varepsilon_{ij},\ j = 1,...,J_i.$$

(For this model, which uses a different part of *NLOGIT*, we number the alternatives $1,...,J_i$ rather than $0,...,J_i$. There is no substantive significance to this – it is purely for convenience in the context of the model development for the program commands.) 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 t chooses alternative j is

$$\text{Prob}(U_{ij} > U_{iq}) \text{ for all } q \neq j.$$

It has been shown that for independent type 1 extreme value distributions, as above, this probability is

$$\text{Prob}(y_i = j) \;=\; \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{iq} + \boldsymbol{\gamma}_q'\mathbf{z}_i\right)}$$

where y_i is the index of the choice made. 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. This is functionally identical to the multinomial logit model of Section N2.6. Indeed, the earlier model emerges by the simple restriction $\boldsymbol{\gamma}_j = \mathbf{0}$. We have distinguished it in this fashion because the nature of the data suggests a different arrangement than for the multinomial logit model and, second, the models in the section to follow are formulated as extensions of this one. Data for the choice variable in this context may come in several different forms that require explicit treatment in the estimation and analysis of the model results:

- Individual data: y_i coded 0 (not chosen) or 1 (for the one chosen alternative)
- Market shares: y_i coded as $0 < s_{ij} < 1$ for J alternatives such that $\Sigma_j s_{ij} = 1$
- Frequencies: y_i coded as $F_{ij} > 0$ for J alternatives
- Ranks: y_i coded 1,2,…, J (with possible ties for last place)
- Best/worst: y_i coded 0 (not best or worst), 1 (best) or 2 (worst)

N2.7.1 Fixed Effects

There are several forms of the multinomial choice models that are suitable for choice experiments, which are essentially panel data sets. The error components logit model (**ECLOGIT** command) noted in Section N2.8 is an example that encompasses *random effects* specifications. The *fixed effects* multinomial logit model (**FEMLOGIT** command) is

$$\text{Prob}(y_{it} = j) \;=\; \frac{\exp\left(\alpha_{jt} + \boldsymbol{\beta}'\mathbf{x}_{itj}\right)}{\sum_{q=1}^{J_i}\exp\left(\alpha_{qt} + \boldsymbol{\beta}'\mathbf{x}_{itq}\right)}.$$

The estimator for this model builds on Chamberlain's conditional logit estimator for binary choice, but uses a far faster (extremely so) algorithm based on the minimum distance estimator.

N2.7.2 Random Regret Logit and Hybrid Utility Models

We consider two direct extensions of the conditional logit model, one related to the forms of the utility functions and a second related to the treatment of heterogeneity.

The random utility form of the model is based on linear utility functions of the alternatives,

$$U_{ijt} \;=\; \boldsymbol{\beta}'\mathbf{x}_{it} \;+\; \varepsilon_{ijt},\ t = 1,...,T_i,\ j = 0,1,...,J, i=1,...,N.$$

The random regret form bases the choices at least partly on attribute level regret functions,

$$R_{ij}(k) = \log[1+\exp(\beta_k(x_{jk} - x_{ik}))]$$

where k denotes the specific attribute and i and j denote association with alternatives i and j, respectively. (See Chorus (2010) and Chorus, Greene and Hensher (2013).) The systematic regret of choice i can then be written

$$R_i = \sum_{j=1}^{J} \sum_{k=1}^{K} \log[1+\exp(\beta_k(x_{jk} - x_{ik}))].$$

The random regret form of the choice model is then

$$P_j = \frac{\exp(-R_j)}{\sum_{j=1}^{J}\exp(-R_j)}$$

This model does not impose the IIA assumptions. The model may also be specified with only a subset of the attributes treated in the random regret format. This hybrid model is

$$P_j = \frac{\exp(-R_j + \boldsymbol{\beta}'\mathbf{x}_{ij})}{\sum_{j=1}^{J}\exp(-R_j + \boldsymbol{\beta}'\mathbf{x}_{ij})}$$

The random regret model is also extended to the latent class framework.

N2.7.3 Scaled MNL Model

The scaled multinomial logit model allows the model to accommodate broad heterogeneity across individuals, for example when two or more data sets from different groups are combined. This is a special case of the generalized mixed logit model described in Section N2.11.2. The general form of the scaled MNL model is

$$\text{Prob}(y_i = j) = \frac{\exp\left(\sigma_i \boldsymbol{\beta}'\mathbf{x}_{ij}\right)}{\sum_{q=1}^{J_i}\exp\left(\sigma_i \boldsymbol{\beta}'\mathbf{x}_{iq}\right)}$$

where $\sigma_i = \exp(\boldsymbol{\delta}'\mathbf{z}_i + \tau v_i)$

The scaling factor, σ_i differs across individuals, but not across choices. It has a deterministic component, $\exp(\boldsymbol{\delta}'\mathbf{z}_i)$, and a random component, $\exp(\tau v_i)$. Either (or both) may equal 1.0, that is, either or both restrictions $\boldsymbol{\delta} = \mathbf{0}$ or $\tau = 0$. For example, a simple nonstochastic scaling differential between two groups would result if $\tau = 0$ and if $\mathbf{z}_i$ were simply a dummy variable that identifies the two groups. Other forms of scaling heterogeneity can be produced by different variables in $\mathbf{z}_i$. The scaling may also be random through the term τv_i. In this instance, v_i is a random term (usually, but not necessarily normally distributed). With $\boldsymbol{\delta} = \mathbf{0}$ and $\tau \neq 0$, we obtain a randomly scaled multinomial logit model.

N2.8 Error Components Logit Model

When the sample consists of a 'panel' of data, that is, when individuals are observed in more than one choice situation, the conditional logit model can be augmented with individual effects, similar to the use of common effects models in regression and other single equation cases. A 'panel data' form of this model that is a counterpart to the random effects model is what we label the 'error components model.' (This has been called the 'kernel logit model' in some treatments in the literature.) The model arises by introducing M up to $\max_i J_i$ alternative and individual specific random terms in the utility functions as in

$$U(\textit{choice } j \textit{ for individual } i \textit{ in choice setting } t)$$

$$= U_{ijt}$$

$$= \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i + \varepsilon_{ij} + \Sigma_{m=1}^{M} d_{jm}\sigma_m u_{im}\,,\ j = 1,\ldots,J_i,\ t = 1,\ldots,T_i.$$

where d_{jm} = 1 if effect m appears in utility function j, 0 if not,

σ_m = the standard deviation of effect m (to be estimated),

v_{im} = effect m for individual i.

The M random individual specifics are $\sigma_m u_{im}$. They are distributed as normal with zero means and variances σ_m^2. The constants d_{jm} equal one if random effect m appears in the utility function for alternative j, and zero otherwise. The error components account for unobserved, alternative specific variation. With this device, the sets of random effects in different utility functions can overlap, so as to accommodate correlation in the unobservables across choices. The random effects may also be heteroscedastic, with

$$\sigma_{m,i}^2 = \sigma_m^2 \exp(\boldsymbol{\theta}_m'\mathbf{z}_i).$$

The probabilities attached to the choices are now

$$\text{Prob}(y_i = j) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i + \Sigma_{m=1}^{M} d_{jm}\sigma_m u_{im}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{iq} + \boldsymbol{\gamma}_q'\mathbf{z}_i \Sigma_{m=1}^{M} d_{qm}\sigma_m u_{im}\right)}.$$

This is precisely an analog to the random effects model for single equation models. Given the patterns of d_{jm}, this can provide a nesting structure as well. Examples in Chapter N30 will demonstrate.

N2.9 Heteroscedastic Extreme Value Model

In the conditional logit model,

$$U(choice\ j\ for\ individual\ i) \;=\; U_{ij} \;=\; \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i + \varepsilon_{ij},\ j = 1,...,J_i,$$

$$\text{Prob}(y_i = j) \;=\; \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i\right)}{\sum_{m=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{im} + \boldsymbol{\gamma}_m'\mathbf{z}_i\right)},$$

an implicit assumption is that the variances of ε_{ji} are the same. With the type 1 extreme value distribution assumption, this common value is $\pi^2/6$. This assumption is a strong one, and it is not necessary for identification or estimation. The heteroscedastic extreme value model relaxes this assumption. We assume, instead, that

$$F(\varepsilon_{ij}) \;=\; \exp(-\exp(-\theta_j\varepsilon_{ij})],$$

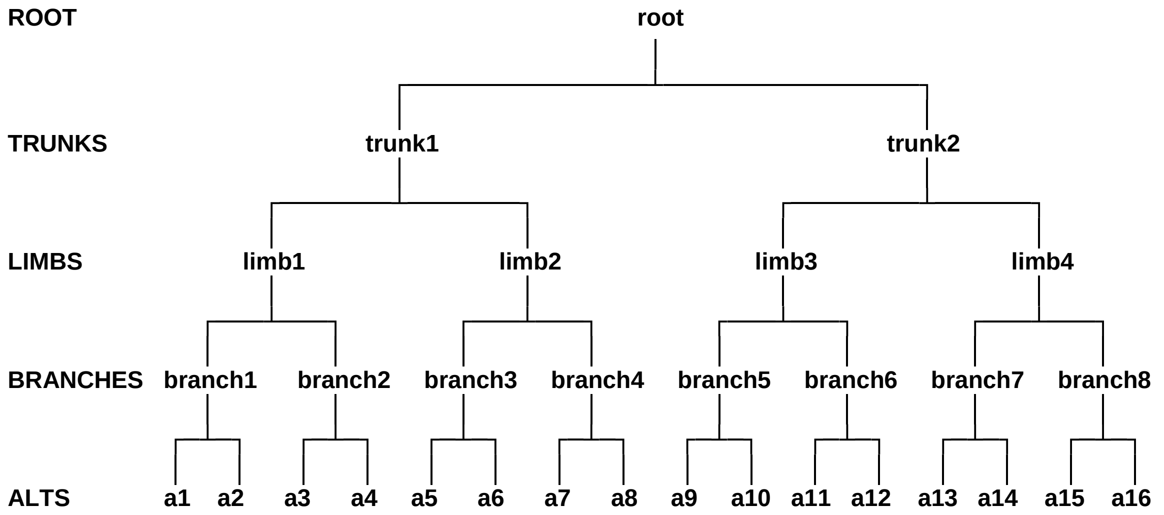
$$\text{Var}[\varepsilon_{ij}] = \sigma_j^2\ (\pi^2/6)\ \text{ where }\ \sigma_j^2 = 1/\theta_j^2,$$

with one of the variance parameters normalized to one for identification. (Technical details for this model including a statement of the probabilities appears in Chapter N26.) A further extension of this model allows the variance parameters to be heterogeneous, in the standard fashion,

$$\sigma_{ij}^2 \;=\; \sigma_j^2\ \exp(\boldsymbol{\gamma}'\mathbf{z}_i).$$

N2.10 Nested and Generalized Nested Logit Models

The nested logit model is an extension of the conditional logit model. The models supported by *NLOGIT* are based on variations of a four level tree structure such as the following:



The choice probability under the assumption of the nested logit model is defined to be the conditional probability of alternative j in branch b, limb l, and trunk r, $j|b,l,r$:

$$P(j|b,l,r) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r})}{\sum_{q|b,l,r}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l,r})} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r})}{\exp(J_{b|l,r})},$$

where $J_{b|l,r}$ is the inclusive value for branch b in limb l, trunk r, $J_{b|l,r} = \log \Sigma_{q|b,l,r}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l,r})$. At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l, trunk r,

$$P(b|l,r) = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \tau_{b|l,r}J_{b|l,r})}{\sum_{s|l,r}\exp(\boldsymbol{\alpha}'\mathbf{y}_{s|l,r} + \tau_{s|l,r}J_{s|l,r})} = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \tau_{b|l,r}J_{b|l,r})}{\exp(I_{l|r})},$$

where $I_{l|r}$ is the inclusive value for limb l in trunk r, $I_{l|r} = \log \Sigma_{s|l,r}\exp(\boldsymbol{\alpha}'\mathbf{y}_{s|l,r} + \tau_{s|l,r}J_{s|l,r})$. The probability of choosing limb l in trunk r is

$$P(l|r) = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_{l|r} + \sigma_{l|r}I_{l|r})}{\sum_{s|r}\exp(\boldsymbol{\delta}'\mathbf{z}_{q|r} + \sigma_{s|r}I_{s|r})} = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_{l|r} + \sigma_{l|r}I_{l|r})}{\exp(H_r)},$$

where H_r is the inclusive value for trunk r, $H_r = \log \Sigma_{s|l}\exp(\boldsymbol{\delta}'\mathbf{z}_{s|r} + \sigma_{s|r}I_{s|r})$.

Finally, the probability of choosing a particular limb is

$$P(r) = \frac{\exp(\boldsymbol{\theta}'\mathbf{h}_r + \phi_r H_r)}{\sum_s \exp(\boldsymbol{\theta}'\mathbf{h}_s + \phi_s H_s)}.$$

By the laws of probability, the unconditional probability of the observed choice made by an individual is

$$P(j,b,l,r) = P(j|b,l,r) \times P(b|l,r) \times P(l|r) \times P(r).$$

This is the contribution of an individual observation to the likelihood function for the sample.

The 'nested logit' aspect of the model arises when any of the $\tau_{b|l,r}$ or $\sigma_{l|r}$ or ϕ_r differ from 1.0. If all of these deep parameters are set equal to 1.0, the unconditional probability reduces to

$$P(j,b,l,r) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r} + \boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \boldsymbol{\delta}'\mathbf{z}_{l|r} + \boldsymbol{\theta}'\mathbf{h}_r)}{\sum_r \sum_l \sum_b \sum_j \exp(\boldsymbol{\beta}'\mathbf{x}_{j,b,l,r} + \boldsymbol{\alpha}'\mathbf{y}_{b,l,r} + \boldsymbol{\delta}'\mathbf{z}_{l,r} + \boldsymbol{\theta}'\mathbf{h}_r)},$$

which is the probability for a one level conditional (multinomial) logit model.

N2.10.1 Alternative Normalizations of the Nested Logit Model

The formulation of the nested logit model imposes no restrictions on the inclusive value parameters. However, the assumption of utility maximization and the stochastic underpinnings of the model do imply certain restrictions. For the former, in principle, the inclusive value parameters must be between zero and one. For the latter, the restrictions are implied by the way that the random terms in the utility functions are constructed. In particular, the nesting aspect of the model is obtained by writing

$$\varepsilon_{j|b,l,r} = u_{j|b,l,r} + v_{b|l,r}.$$

That is, within a branch, the random terms are viewed as the sum of a unique component, $u_{j|b,l,r}$, and a common component, $v_{b|l,r}$. This has certain implications for the structure of the scale parameters in the model. *NLOGIT* provides a method of imposing the restrictions implied by the underlying theory.

There are three possible normalizations of the inclusive value parameters which will produce the desired results. These are provided in this estimator for two and three level models only. This includes most of the received applications. We will detail the first two of these forms here and describe how to estimate all of them in Chapter N28. For convenience, we label these random utility formulations RU1, RU2 and RU3. (RU3 is just a variant of RU2.)

RU1

The first form is

$$P(j|b,l) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\sum_{q|b,l}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|j,l})} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\exp(J_{b|l})},$$

where $J_{b|l}$ is the inclusive value for branch b in limb l,

$$J_{b|l} = \log \Sigma_{q|b,l} \exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l}).$$

At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l,

$$P(b|l) = \frac{\exp\left[\lambda_{b|l}(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l})\right]}{\sum_{s|l}\exp\left[\lambda_{s|l}(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l})\right]} = \frac{\exp\left[\lambda_{b|l}(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l})\right]}{\exp(I_l)},$$

where I_l is the inclusive value for limb l,

$$I_l = \log \Sigma_{s|l} \exp[\lambda_{s|l} (\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l})].$$

The probability of choosing limb l is

$$P(l) = \frac{\exp[\gamma_l(\boldsymbol{\delta}'\mathbf{z}_{l|} + I_l)]}{\sum_s \exp[\gamma_s(\boldsymbol{\delta}'\mathbf{z}_s + I_s)]} = \frac{\exp[\gamma_l(\boldsymbol{\delta}'\mathbf{z}_l + I_l)]}{\exp(H)}.$$

Note that this the same as the familiar normalization used earlier; this form just makes the scaling explicit at each level.

RU2

The second form moves the scaling down to the twig level, rather than at the branch level. Here it is made explicit that within a branch, the scaling must be the same for alternatives.

$$P(j|b,l) = \frac{\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\sum_{q|b,l}\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{q|b,l})\right]} = \frac{\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\exp(J_{b|l})}.$$

Note in the summation in the inclusive value that the scaling parameter is not varying with the summation index. It is the same for all twigs in the branch. Now, $J_{b|l}$ is the inclusive value for branch j in limb l,

$$J_{b|l} = \log \Sigma_{q|b,l} \exp[\mu_{b|l} (\boldsymbol{\beta}'\mathbf{x}_{q|b,l})].$$

At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l,

$$P(b|l) = \frac{\exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + (1/\mu_{b|l})J_{b|l}\right)\right]}{\sum_s \exp\left[\gamma_s\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + (1/\mu_{s|l})J_{s|l}\right)\right]} = \frac{\exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + (1/\mu_{b|l})J_{b|l}\right)\right]}{\exp(I_l)},$$

where I_l is the inclusive value for limb l,

$$I_l = \log\sum_{s|l}\exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + (1/\mu_{s|l})J_{s|l}\right)\right].$$

Finally, the probability of choosing limb l is

$$P(l) = \frac{\exp\left[\boldsymbol{\delta}'\mathbf{z}_l + (1/\gamma_l)I_l\right]}{\sum_s \exp\left[\boldsymbol{\delta}'\mathbf{z}_s + (1/\gamma_s)I_s\right]} = \frac{\exp\left[\boldsymbol{\delta}'\mathbf{z}_l + (1/\gamma_l)I_l\right]}{\exp(H)},$$

where the log sum for the full model is

$$H = \log\sum_s \exp\left[\boldsymbol{\delta}'\mathbf{z}_s + (1/\gamma_s)I_s\right].$$

N2.10.2 A Model of Covariance Heterogeneity

This is a modification of the two level nested logit model. The base case for the model is

$$P(j\,|\,b) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b})}{\sum_{q=1}^{J|b}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b})}.$$

Denote the logsum, the log of the denominator, as J_b = inclusive value for branch b = IV(b). Then,

$$P(b) = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_b + \tau_b J_b)}{\sum_{s=1}^{B}\exp(\boldsymbol{\alpha}'\mathbf{y}_s + \tau_s J_s)}.$$

The covariance heterogeneity model allows the τ_b inclusive value parameters to be functions of a set of attributes, $\mathbf{v}_b$, in the form

$$\tau_b^* = \tau_b \times \exp(\boldsymbol{\delta}'\mathbf{v}_b),$$

where $\boldsymbol{\delta}$ is a new vector of parameters to be estimated. Since the inclusive parameter is a scaling parameter for a common random component in the alternatives within a branch, this is equivalent to a model of heteroscedasticity.

N2.10.3 Generalized Nested Logit Model

The generalized nested logit model is an extension of the nested logit model in which alternatives may appear in more than one branch. Alternatives that appear in more than one branch are allocated across branches probabilistically. The model estimated includes the usual nested logit framework (only two levels are supported in this framework), as well as the matrix of allocation parameters. The only difference between this and the more basic nested logit model is the specification of the tree. For the allocations of choices to branches, a multinomial logit form is used,

$$\pi_{j,b} \;=\; \text{Prob}(\textit{alternative } j \textit{ is in branch } b) \;=\; \exp(\theta_{j,b}) \,/\, \Sigma_s \exp(\theta_{j,s}),$$

where the parameters θ are estimated by the program. Note the denominator summation is over branches that the alternative appears in. The probabilities sum to one. The identification rule that one of the θs for each alternative modeled equals one is imposed. These allocations may depend on an individual characteristic (not a choice attribute), such as income. In this instance, the multinomial logit probabilities become functions of this variable,

$$\pi_{j,b} \;=\; \text{Prob}(\textit{alternative } j \textit{ is in branch } b) \;=\; \exp(\theta_{j,b} + \gamma_{j,b}z_i) \,/\, \Sigma_s \exp(\theta_{j,s} + \gamma_{j,s}z_i).$$

Now, to achieve identification, one of the θs is set equal to zero and one of the γs is set equal to zero. It is convenient to form the matrix $\Pi = [\pi_{j,b}]$. This is a $J \times B$ matrix of allocation parameters. The rows sum to one, and note that some values in the matrix are zero. But, no rows have all zeros – every alternative appears in at least one branch, and no columns have all zeros – every branch contains at least one alternative.

The probabilities for the observed choices are formed as

$$\text{Prob}(\textit{alternative}, \textit{branch}) = P(j,b) = P(j|b) \times P(b)$$

where
$$P(j\,|\,b) = \frac{[\pi_{j,b}U_j]^{\sigma_b}}{\sum_{s=1}^{B}[\pi_{j,s}U_s]^{\sigma_s}}$$

(the denominator summation is over the alternatives in that branch)

and
$$P(b) = \frac{\left[\sum_{j|b}[\pi_{j,b}U_j]^{\sigma_b}\right]^{1/\sigma_b}}{\sum_{b=1}^{B}\left[\sum_{j|b}[\pi_{j,b}U_j]^{\sigma_b}\right]^{1/\sigma_b}}.$$

N2.10.4 Box-Cox Nested Logit

The Box-Cox form of the nested logit model automates a model specification that was already in *NLOGIT* 4. This form can replace the function transformation BCX(variable) in the utility functions.

N2.11 Random Parameters Logit Models

In its most general form, we write the multinomial logit probability as

$$P(j\,|\,\mathbf{v}_i) = \frac{\exp(\alpha_{ji} + \boldsymbol{\theta}_j'\mathbf{z}_i + \boldsymbol{\phi}_j'\mathbf{f}_{ji} + \boldsymbol{\beta}_{ji}'\mathbf{x}_{ji})}{\sum_{q=1}^{J}\exp(\alpha_{qi} + \boldsymbol{\theta}_q'\mathbf{z}_i + \boldsymbol{\phi}_q'\mathbf{f}_{qi} + \boldsymbol{\beta}_{qi}'\mathbf{x}_{qi})},$$

where $U(j,i) = \alpha_{ji} + \boldsymbol{\theta}_j'\mathbf{z}_i + \boldsymbol{\phi}_j'\mathbf{f}_{ji} + \boldsymbol{\beta}_{ji}'\mathbf{x}_{ji}$, $j = 1,...,J_i$ alternatives in individual i's choice set

α_{ji} is an alternative specific constant which may be fixed or random, $\alpha_{Ji} = 0$,

$\boldsymbol{\theta}_j$ is a vector of nonrandom (fixed) coefficients, $\boldsymbol{\theta}_{Ji} = \mathbf{0}$,

$\boldsymbol{\phi}_j$ is a vector of nonrandom (fixed) coefficients,

$\boldsymbol{\beta}_{ji}$ is a coefficient vector that is randomly distributed across individuals; $\mathbf{v}_i$ enters $\boldsymbol{\beta}_{ji}$,

$\mathbf{z}_i$ is a set of choice invariant individual characteristics such as age or income,

$\mathbf{f}_{ji}$ is a vector of M individual and choice varying attributes of choices, multiplied by $\boldsymbol{\phi}_j$,

$\mathbf{x}_{ji}$ is a vector of L individual and choice varying attributes of choices, multiplied by $\boldsymbol{\beta}_{ji}$.

The term 'mixed logit' is often used in the literature (e.g., Revelt and Train (1998)) for this model. The choice specific constants, α_{ji} and the elements of $\boldsymbol{\beta}_{ji}$ are distributed randomly across individuals such that for each random coefficient, ρ_{ki} = any (not necessarily all of) α_{ji} or β_{jki}, the coefficient on attribute x_{jik}, $k = 1,...,K$,

$$\rho_{jki} = \alpha_{ji} \text{ or } \beta_{jki} = \rho_{jk} + \boldsymbol{\delta}_k'\mathbf{w}_i + \sigma_k v_{ki},$$

or

$$\rho_{jki} = \alpha_{ji} \text{ or } \beta_{jki} = \exp(\rho_{jk} + \boldsymbol{\delta}_{jk}'\mathbf{w}_i + \sigma_{jk} v_{jki}).$$

The vector $\mathbf{w}_i$ (which does not include *one*) is a set of choice invariant characteristics that produce individual heterogeneity in the means of the randomly distributed coefficients; ρ_{jk} is the constant term and $\boldsymbol{\delta}_{jk}$ is a vector of 'deep' coefficients which produce an individual specific mean. The random term, v_{jki} is normally distributed (or distributed with some other distribution) with mean 0 and standard deviation 1, so σ_{jk} is the standard deviation of the marginal distribution of ρ_{jki}. The v_{jki}s are individual and choice specific, unobserved random disturbances – the source of the heterogeneity. Thus, as stated above, in the population

$$\alpha_{ji} \text{ or } \beta_{jki} \sim \text{Normal or Lognormal } [\rho_{jk} + \boldsymbol{\delta}_{jk}'\mathbf{w}_i, \sigma_{jk}^2].$$

(Other distributions may be specified.) For the full vector of K random coefficients in the model, we may write

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Delta}\mathbf{w}_i + \boldsymbol{\Gamma}\mathbf{v}_i$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix which contains σ_k on its diagonal. A nondiagonal $\boldsymbol{\Gamma}$ allows the random parameters to be correlated. Then, the full covariance matrix of the random coefficients is $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$. The standard case of uncorrelated coefficients has $\boldsymbol{\Gamma} = \text{diag}(\sigma_1,\sigma_2 ,\ldots,\sigma_k)$. If the coefficients are freely correlated, $\boldsymbol{\Gamma}$ is a full, unrestricted, *lower triangular* matrix and $\boldsymbol{\Sigma}$ will have nonzero off diagonal elements. An additional level of flexibility is obtained by allowing the distributions of the random parameters to be heteroscedastic,

$$\sigma_{ijk}^2 = \sigma_{jk}^2 \times \exp(\boldsymbol{\gamma}_{jk}'\mathbf{h}_i).$$

This is now built into the model by specifying

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Delta}\mathbf{w}_i + \boldsymbol{\Gamma}\,\boldsymbol{\Omega}_i\,\mathbf{v}_i$$

where

$$\boldsymbol{\Omega}_i = \text{diag}[\sigma_{ijk}^2]$$

and now, $\boldsymbol{\Gamma}$ is a lower triangular matrix of constants with ones on the diagonal. Finally, autocorrelation can also be incorporated by allowing the random components of the random parameters to obey an autoregressive process,

$$v_{ki,t} = \tau_{ki}\, v_{ki,t-1} + c_{ki,t}$$

where $c_{ki,t}$ is now the random element driving the random parameter.

This produces, then, the full random parameters logit model

$$P(j \mid \mathbf{v}_i) = \frac{\exp(\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{ji})}{\sum_{m=1}^{J} \exp(\alpha_{mi} + \boldsymbol{\beta}_i'\mathbf{x}_{mi})},$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\boldsymbol{\Omega}_i \mathbf{v}_i$$

$\mathbf{v}_i$ ~ with mean vector **0** and covariance matrix **I**.

The specific distributions may vary from one parameter to the next. We also allow the parameters to be lognormally distributed so that the preceding specification applies to the logarithm of the specific parameter.

N2.11.1 Nonlinear Utility RP Model

In the NLRP model, the model parameters may be specified as in the model above. But, the utility functions need not be linear in the attributes and characteristics. This more general model is

$$P(j \mid \mathbf{v}_i) = \frac{\exp[U_j(\boldsymbol{\beta}_i', \mathbf{x}_{ji})]}{\sum_{m=1}^{J} \exp[U_j(\boldsymbol{\beta}_i', \mathbf{x}_{ji})]},$$

where $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\boldsymbol{\Omega}_i \mathbf{v}_i$

$\mathbf{v}_i$ ~ with mean vector **0** and covariance matrix **I**.

and $U_j(\boldsymbol{\beta}_i', \mathbf{x}_{ji})$ is any nonlinear function of the data and parameters.

N2.11.2 Generalized Mixed Logit Model

The second major extension of the random parameters model is the generalized mixed logit model developed by Fiebig, Keane, Louviere and Wasi (2010). The extension of the random parameters model is

$$\beta_i = \sigma_i\beta + \gamma\Gamma v_i + (1 - \gamma)\sigma_i\Gamma v_i$$

The generalized mixed logit model embodies several different forms of heterogeneity in the random parameters and random scaling, as well as the distribution parameter, γ, which allocates the influence of the parameter heterogeneity and the scaling heterogeneity. Several interesting model forms are produced by different restrictions on the parameters. For example, if $\gamma = 0$ and $\Gamma = 0$, we obtain the scaled MNL model in Section N2.7.3. A variety of other special cases are also provided. One nonlinear normalization in particular allows the model to be transformed from a specification in 'utility space' as above to 'willingness to pay space' by analyzing an implicit ratio of coefficients.

N2.12 Latent Class Logit Models

In the latent class formulation, parameter heterogeneity across individuals is modeled with a discrete distribution, or set of 'classes.' The situation can be viewed as one in which the individual resides in a 'latent' class, *c*, which is not revealed to the analyst. There are a fixed number of classes, *C*. Estimates consist of the class specific parameters and for each person, a set of probabilities defined over the classes. Individual *i*'s choice among *J* alternatives at choice situation *t* given that individual *i* is in class *c* is the one with maximum utility, where the utility functions are

$$U_{jit|c} = \boldsymbol{\beta}_c'\mathbf{x}_{jit} + \varepsilon_{jit}$$

where U_{jit} = utility of alternative *j* to individual *i* in choice situation *t*

$\mathbf{x}_{jit}$ = union of all attributes that appear in all utility functions. For some alternatives, $x_{jit,k}$ may be zero by construction for some attribute *k* which does not enter their utility function for alternative *j*.

ε_{jit} = unobserved heterogeneity for individual *i* and alternative *j* in choice situation *t*.

$\boldsymbol{\beta}_c$ = class specific parameter vector.

Within the class, choice probabilities are assumed to be generated by the multinomial logit model

$$\text{Prob}[y_{it} = j \mid \text{class} = c] = \frac{\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}{\sum_{j=1}^{J_i}\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}.$$

As noted, the class is not observed. Class probabilities are specified by the multinomial logit form,

$$\text{Prob}[\text{class} = c] = Q_{ic} = \frac{\exp(\boldsymbol{\theta}_c'\mathbf{z}_i)}{\sum_{c=1}^{C}\exp(\boldsymbol{\theta}_c'\mathbf{z}_i)},\ \boldsymbol{\theta}_C = \mathbf{0}.$$

where $\mathbf{z}_i$ is an optional set of person, situation invariant characteristics. The class specific probabilities may be a set of fixed constants if no such characteristics are observed. In this case, the class probabilities are simply functions of *C* parameters, θ_c, the last of which is fixed at zero. This model does not impose the IIA property on the observed probabilities.

For a given individual, the model's estimate of the probability of a specific choice is the expected value (over classes) of the class specific probabilities. Thus,

$$\text{Prob}(y_{it} = j) = E_c\left[\frac{\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}{\sum_{j=1}^{J_i}\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}\right]$$

$$= \sum_{c=1}^{C}\text{Prob}(class = c)\left[\frac{\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}{\sum_{j=1}^{J_i}\exp(\boldsymbol{\beta}_c'\mathbf{x}_{jit})}\right].$$

N2.12.1 2^K Latent Class Model for Attribute Nonattendance

NLOGIT accommodates attribute 'nonattendance' by the '-888' feature described in Chapter N18. In particular, in some choice analyses, some, but not all individuals indicate that they did not pay attention to certain attributes. The appropriate model building strategy is to impose zero restrictions on the utility parameters, $\boldsymbol{\beta}$, for these specific individuals. *NLOGIT* provides this capability throughout the estimation suite – all models are fit with this capability. (This feature is unique to *NLOGIT*.) This feature accommodates cases in which individuals explicitly reveal the form of their utility functions. The model noted here is usable when the sorting of individuals in this way is latent – there is no observed indicator. Consider a model with four attributes, $x1$, $x2$, $x3$, $x4$. All individuals attend to $x1$ and $x2$. Some ignore $x3$, some ignore $x4$, and some ignore both $x3$ and $x4$ (and some attend both). Thus, in terms of the possible utility functions, there are four types of individuals in the population, distinguished by the type of utility function that is appropriate:

($x3$ and $x4$)	$U_{ij} = \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \beta_4 x4 + \varepsilon$
($x3$ only)	$U_{ij} = \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \varepsilon$
($x4$ only)	$U_{ij} = \beta_1 x1 + \beta_2 x2 + \beta_4 x4 + \varepsilon$
(Neither)	$U_{ij} = \beta_1 x1 + \beta_2 x2 + \varepsilon$

The difference that is built into this model form is that the analyst does not know which individual is in which group. This can be treated as a latent class model. The number of classes is 2^K where K is the number of attributes that treated by the latent class specification.

N2.12.2 Latent Class – Random Parameters Model

The LCRP model is a combination of the latent class model described above and the random parameters model in Section N2.11. This is a latent class model in which a random parameters model applies within each class.

N2.13 Multinomial Probit Model

In this model, the individual's choice among J alternatives is the one with maximum utility, where the utility functions are

$$U_{ji} = \boldsymbol{\beta}'\mathbf{x}_{ji} + \varepsilon_{ji}$$

where U_{ji} = utility of alternative j to individual i

$\mathbf{x}_{jit}$ = union of all attributes that appear in all utility functions. For some alternatives, $x_{jit,k}$ may be zero by construction for some attribute k which does not enter their utility function for alternative j.

The multinomial logit model specifies that ε_{ji} are draws from independent extreme value distributions (which induces the IIA condition). In the multinomial probit model, we assume that ε_{ji} are normally distributed with standard deviations Sdv[ε_{ji}] = σ_j and correlations Cor[ε_{ji}, ε_{qi}] = ρ_{jq} (the same for all individuals). Observations are independent, so Cor[ε_{ji},ε_{qs}] = 0 if i is not equal to s, for all j and q. A variation of the model allows the standard deviations and covariances to be scaled by a function of the data, which allows some heteroscedasticity across individuals.

The correlations ρ_{jq} are restricted to $-1 < \rho_{jq} < 1$, but they are otherwise unrestricted save for a necessary normalization. The correlations in the last row of the correlation matrix must be fixed at zero. The standard deviations are unrestricted with the exception of a normalization – two standard deviations are fixed at 1.0 – *NLOGIT* fixes the last two.

This model may also be fit with panel data. In this case, the utility function is modified as follows:

$$U_{ji,t} \;=\; \boldsymbol{\beta}'\mathbf{x}_{ji,t} + \varepsilon_{ji,t} + v_{ji,t}$$

where ‘t’ indexes the periods or replications. There are two formulations for $v_{ji,t}$,

Random effects	$v_{ji,t} = v_{ji,t}$ (the same in all periods)
First order autoregressive	$v_{ji,t} = \alpha_j v_{ji,t-1} + a_{ji,t}$.

It is assumed that you have a total of T_i observations (choice situations) for person i. Two situations might lend themselves to this treatment. If the individual is faced with a set of choice situations that are similar and occur close together in time, then the random effects formulation is likely to be appropriate. However, if the choice situations are fairly far apart in time, or if habits or knowledge accumulation are likely to influence the latter choices, then the autoregressive model might be the better one.

You can also add a form of individual heterogeneity to the disturbance covariance matrix. The model extension is

$$\text{Var}[\boldsymbol{\varepsilon}_i] \;=\; \exp[\boldsymbol{\gamma}'\mathbf{h}_i] \times \boldsymbol{\Sigma}$$

where $\boldsymbol{\Sigma}$ is the matrix defined earlier (the same for all individuals), and $\mathbf{h}_i$ is an individual (not alternative) specific set of variables not including a constant.

N3: Model and Command Summary for Discrete Choice Models

N3.1 Introduction

The chapters to follow will provide details on the various discrete choice models you can estimate with *NLOGIT* and on the model commands you will use to request the estimates. This chapter will provide a brief summary listing of the models and model commands. The variety of logit models now use a set of specific names, rather than qualifiers to more general model classes as in earlier versions. For example, the model name **OLOGIT** can be used instead of **ORDERED ; Logit**. The earlier formats remain available, but the newer ones may prove more convenient. The full listing of these commands is also given below. The commands below specify the essential parts needed to fit the model. The numerous options and different forms are discussed in the chapters to follow (and, were noted in the *LIMDEP Econometric Modeling Guide* as well).

N3.2 Model Dimensions

The descriptions below present the different discrete choice models that are the main feature of *NLOGIT*. *NLOGIT* contains all of *LIMDEP*, so all of the models documented in the *LIMDEP Econometric Modeling Guide*, including the regression models, limited dependent variable models, generalized linear models, sample selection models, and so on are supported in *NLOGIT*, as well as the ancillary tools including **MATRIX**, etc.

There are various built in limits in the estimators. These are noted at the specific points below where necessary. The following lists the most important internal constraints on the estimators:

	maximum numbers of:
• Multinomial choice model estimators in *NLOGIT*:	
◦ Alternatives	500
◦ Attributes	300
◦ Branches in nested logit models	25
◦ Limbs in nested logit models	10
◦ Random error components	10
• Maximum number of choices in the MLOGIT form of the model	25
• Heteroscedasticity models, maximum number of variables	75
• Ordered choice models: maximum number of outcomes	25
• Unconditional fixed effects models, number of individuals	100,000
• Random parameters models, maximum number of RPs	25
• Latent class models, maximum number of classes	30

N3.3 Basic Discrete Choice Models

The binomial probit and logit models and the ordered probit and logit models are the primary model frameworks for single equation, single decision, discrete choice models. The ordered choice and the bivariate and multivariate probit models are multivariate extensions of the simple probit model.

N3.3.1 Binary Choice Models

There are six binary choice models, probit, logit, complementary log log, Gompertz, Burr, and arctangent documented in Chapter E27. The ones that interest us here are the binary probit and logit models. The probit model is requested with

```
PROBIT      ; Lhs  = dependent variable
            ; Rhs  = independent variables $
```

The binary logit model may be invoked with

```
BLOGIT      ; Lhs  = dependent variable
            ; Rhs  = independent variables $
```

In earlier versions, you would use the **LOGIT** command, which is still useable. **LOGIT** is the same as **BLOGIT** when the data on the dependent variable are either binary (zeros and ones) or proportions (strictly between zero and one). Chapters E26-E29 document numerous extensions of these models. Chapters E30-E32 consider semiparametric and nonparametric approaches and extensions of the binary choice models for panel data.

N3.3.2 Bivariate Binary Choices

The command for the bivariate probit model is

```
BVPROBIT    ; Lhs  = variable 1, variable 2
            ; Rh1 = independent variables for equation 1
            ; Rh2 = independent variables for equation 2 $
```

In this form, the Lhs specifies two binary dependent variables. You may use proportions data instead, in which case, you will provide four proportions variables, in order, *p*00, *p*01, *p*10, *p*11. This command is the same as **BIVARIATE PROBIT** in earlier versions. (You may still use **BIVARIATE PROBIT**.)

N3.3.3 Multivariate Binary Choice Models

The multivariate probit model is specified with

```
MVPROBIT    ; Lhs  = y1, y2, ..., yM
            ; Eq1 = Rhs variables for equation 1
            ; Eq2 = Rhs variables for equation 2
            ...
            ; EqM = Rhs variables for equation M $
```

Data for this model must be individual. The Lhs specifies a set of binary dependent variables. This command is the same as **MPROBIT** (which may still be used) in earlier versions.

N3.3.4 Ordered Choice Models

Chapter E34 describes five forms for the ordered choice model, probit, logit, complementary log log, Gompertz and arctangent. The first two interest us here. The ordered probit model is requested with

OPROBIT ; Lhs = dependent variable
; Rhs = independent variables $

This is the same as the **ORDERED PROBIT** command, which may still be used. In this model, the dependent variable is integer valued, taking the values 0, 1, ..., J. All J+1 values must appear in the data set, including zero. You may supply a set of J+1 proportions variables instead. Proportions will sum to 1.0 for every observation. Chapter E35 documents a bivariate version of the ordered probit model for two joint ordered outcomes, and a sample selection model.

The ordered logit model is requested with

OLOGIT ; Lhs = dependent variable
; Rhs = independent variables $

The same arrangement for the dependent variables as for the ordered probit model is assumed. This command is the same as **ORDERED ; Logit** in earlier versions.

N3.4 Multinomial Logit Models

The 'multinomial logit model' is a special case of the conditional logit model, which, itself, is the gateway model to the main model extensions described in Chapter N2.

N3.4.1 Multinomial Logit

The multinomial logit model described in Section N2.6 and Chapter E37 is invoked with

MLOGIT ; Lhs = dependent variable
; Rhs = independent variables $

Data for the MLOGIT model consist of an integer valued variable taking the values 0, 1, ..., J. This model may also be fit with proportions data. In that case, you will provide the names of J+1 Lhs variables that will be strictly between zero and one, and will sum to one at every observation. The **MLOGIT** command is the same as **LOGIT**. The program inspects the command (Lhs) and the data, and determines internally whether **BLOGIT** or **MLOGIT** is appropriate. Note, on proportions data, if you want to fit a binary logit model with proportions data, you will supply a single proportions variable, not two. (What would be the second one is just one minus the first.) If you want to fit a multinomial logit model with proportions data with three or more outcomes, you must provide the full set of proportions. Thus, you would never supply two Lhs variables in a **LOGIT**, **BLOGIT** or **MLOGIT** command. Three other forms of this canonical model are the sequential logit model, **SEQLOGIT**, and two forms for panel data, **REMLOGIT** for random effects and **FEMLOGIT** for fixed effects.

N3.4.2 Conditional Logit

The command for the conditional model, and the commands in the sections to follow, are variants of the **NLOGIT** command. This is a full class of estimators based on the conditional logit form. There are several forms of the essential command for fitting the conditional logit model with *NLOGIT*. The simpler one is

CLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

As discussed in Chapter N20 and in Section E38.3, the data for this estimator consist of a set of *J* observations, one for each alternative. (The observation resembles a group in a panel data set.) The command just given assumes that every individual in the sample chooses from the same size choice set, *J*. The choice sets may have different numbers of choices, in which case, the command is changed to

; Lhs = dependent variable, choice set size variable

The second Lhs variable is structured exactly the same as a **; Pds** variable for a panel data estimator. In the second form of the model command, the utility functions are specified directly, symbolically. The **; Rhs** and **; Rh2** specifications can be replaced with

; Model: ... specification of the utility functions

This is discussed in Chapter N21 and Chapter E39.

The **CLOGIT** command is the same as **DISCRETE CHOICE**. It is also the same as **NLOGIT** when the only information given in the command is that specified above, that is when none of the specifications that invoke the model extensions that are described in the sections to follow are provided.

N3.5 NLOGIT Extensions of Conditional Logit

The conditional logit model provides the basic framework for a very large number of extensions that are provided by *NLOGIT*. The following lists the basic commands for most of these. Each model form is developed in greater detail in one of the chapters that follow. Each model may be specified with a variety of options and different specifications for numerous variants. The following shows the essential command for the most basic form of the model.

N3.5.1 Random Regret Logit

The random regret form of the model is specified with

RRLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

The command is otherwise the same as **CLOGIT**, with the same formats for variable choice set sizes, etc. The utility functions must be specified as above, not using **; Model: …,** owing to the particular form of the utility functions in the random regret format. The random regret logit model can be framed in several forms, including latent class, random parameters and best/worst multinomial logit.

N3.5.2 Scaled Multinomial Logit

The scaled multinomial logit model is a randomly scaled MNL, with $\beta_i = \sigma_i\beta$, where σ_i is a heterogeneous scalar. The model command is

SMNLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

N3.5.3 Heteroscedastic Extreme Value

The heteroscedastic extreme value model is requested with the command

HLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

The command is otherwise the same as the **CLOGIT** command, with the same formats for variable choice set sizes and utility function specifications. The **HLOGIT** command is the same as

NLOGIT ; Heteroscedasticity
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

as used in earlier versions of *NLOGIT*. (This may still be used if desired.)

N3.5.4 Error Components Logit and Fixed Effects

The error components model is described in Section N2.8 and in Chapter N30. The model command is

ECLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; ECM = specification of the tree structure for the error components $

This command is the same as **NLOGIT ; ECM = specification ... $** The error components model may also be specified as a part of the random parameters model. Thus, your **RPLOGIT** command may also contain the **; ECM = specification**. The error components logit model is a random effects model if each alternative (save for one) has a separate choice situation invariant random term. The default specification is consistent with a random effects model.

A 'fixed effects' specification,

$$U_{ijt} = \alpha_{ij} + \mathbf{x}_{ijt}'\boldsymbol{\beta} + \varepsilon_{ijt}$$

is also provided as the **FEMLOGIT** command.

N3.5.5 Nested and Generalized Nested Logit

The nested logit model is the default form of the **NLOGIT** command. Request the nested logit model with

NLOGIT ; Tree = specification of the tree structure
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

The generalized nested logit model command is

GNLOGIT ; Tree = specification of the tree structure
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

The **GNLOGIT** command in place of the **NLOGIT** command tells *NLOGIT* that the tree structure may have overlapping branch specifications. (You may also use **NLOGIT ; GNL**.) If you specify that alternatives appear in more than one branch in the **NLOGIT** command, this will produce an error message. The option is available only for the **GNLOGIT** command. The specification of variable choice set sizes and utility functions is the same as for the **CLOGIT** command.

N3.5.6 Random Parameters Logit

The random parameters logit model (mixed logit model) is requested by specifying a conditional logit model, and adding the specification of the random parameters. The model command is

RPLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Fcn = the specifications of the random parameters
; ... other specifications for the random parameters model $

Once again, variable choice set sizes and utility function specifications are specified as in the **CLOGIT** command. This command is the same as

NLOGIT ; RPL
; ... the rest of the command $

There is one modification that might be necessary. If you are providing variables that affect the means of the random parameters, you would generally use

NLOGIT ; RPL = the list of variables
; ... the rest of the command $

The RPL specification may still be used this way. The command can be **NLOGIT** as above, or

RPLOGIT ; RPL = the list of variables
; ... the rest of the command $

These are identical.

The random parameters model may also include an error components specification defined in the next section. The command will be

RPLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Fcn = the specifications of the random parameters
; ... other specifications for the random parameters model
; ECM = specification $

There are many variants of the random parameters logit model supported in *NLOGIT*. The most flexible is the generalized mixed logit model described in the next section. Three somewhat narrower, specific forms are

Best/worst multinomial logit:	**BWRPLOGIT ; …**
Random regret multinomial logit:	**RPRRLOGIT**
Berry, Levinsohn, Pakes (BLP) logit:	**BLPLOGIT**

The logit model of Berry, Levinsohn and Pakes is used for aggregate, market share data. *NLOGIT* uses a new algorithm (see Lee and Seo (2015)) for estimation of the BLP model that promises to be much faster than the conventional contraction mapping approach.

N3.5.7 Generalized Mixed Logit

The generalized mixed logit model is an extension of the random parameters model. The command has several parts that produce the various model types. The essential command is

GMXLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Fcn = specification of the random parameters $

N3.5.8 Nonlinear Random Parameters Logit

This command extends the random parameters model by allowing the utility functions to be any nonlinear that you specify. There are numerous variants of this model. The essential command is

NLRPLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Labels = the labels used for the model parameters
; Start = starting values for iterations
; Fn1 = specification of a nonlinear function
; … up to 50 nonlinear function specifications
; Model: U(name…) = one of the nonlinear functions defined /
U(name…) = another one of the functions, etc.
; Fcn = specifications of the random parameters $

The model is set up by defining the choice variable and a set of nonlinear functions that will be combined to make the utility functions. The functions may be arbitrarily complex

N3.5.9 Latent Class Logit

The essential form of the command for the latent class model is

LCLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Pts = the number of classes $

Like the **RPLOGIT** command, you need to modify this command if you are providing variables that affect the class probabilities. You would generally use

NLOGIT ; LCM = the list of variables
; ... the rest of the command $

The LCM specification may still be used this way. The command can be **NLOGIT** as above, or identically,

LCLOGIT ; LCM = the list of variables
; ... the rest of the command $

The default framework for the latent class model is random utility. The model may be changed to random regret with model command **LCRRLOGIT**. Another variant of the latent class model may have a mix of random regret based classes and maximum random utility classes.

N3.5.10 2^K Latent Class Logit

The 2^K model is a particular latent class model in which there are simple constraints across the classes, but only one parameter vector used for the whole model. The model is set up as a latent class model with an additional specification:

LCLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Pts = the number of classes $

In this form of the model, the number of points is specified as 102, 103, or 104, corresponding to whether the first 2, 3, or 4 variables in the RHS list are given the special treatment that defines the model.

N3.5.11 Latent Class Random Parameters

The latent class random parameters model extends the latent class model. The essential command is

LCRPLOGIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics
; Fcn = definition of the random parameters part
; Pts = the number of classes $

N3.5.12 Multinomial Probit

The multinomial probit model is described in Chapter N27 and Section N2.13. The essential command is

MNPROBIT ; Lhs = dependent variable
; Choices = the names of the J alternatives
; Rhs = list of choice specific attributes
; Rh2 = list of choice invariant individual characteristics $

Variable choice set sizes and utility function specifications are specified as in the **CLOGIT** command. This command is the same as

NLOGIT ; MNP
; ... the rest of the command $

N3.6 Command Summary

The following lists the current and where applicable, alternative forms of the discrete choice model commands. The two sets of commands are identical, and for each model, in *NLOGIT* 6, either command may be used for that model.

Models	Command	Alternative Command Form
Binary Choice Models		
Binary Probit	**PROBIT**	**PROBIT**
Binary Logit	**BLOGIT**	**LOGIT**
Bivariate Probit	**BVPROBIT**	**BIVARIATE PROBIT**
Multivariate Probit	**MVPROBIT**	**MPROBIT**
Ordered Choice Models		
Ordered Probit	**OPROBIT**	**ORDERED PROBIT**
Ordered Logit	**OLOGIT**	**ORDERED ; Logit**
Multinomial Logit Models		
Multinomial Logit	**MLOGIT**	**LOGIT**
Conditional Logit	**CLOGIT**	**DISCRETE CHOICE**
Conditional Logit Extensions		
Conditional Logit	**CLOGIT**	**CLOGIT**
Multinomial Logit	**NLOGIT**	**NLOGIT** (same as **CLOGIT**)
Scaled Multinomial Logit	**SMNLOGIT**	**GMXLOGIT ; SMNL**
Best/Worst Model	**BWLOGIT**	
Best/Worst Multinomial Logit	**BWMNLOGIT**	
Best/Worst Random Parameters	**BWRPLOGIT**	
Sequential Logit	**SEQLOGIT**	
Random Effects Multinomial Logit	**REMLOGIT**	
Fixed Effects Multinomial Logit	**FEMLOGIT**	
Random Regret Multinomial Logit	**RRLOGIT**	
Latent Class Random Regret Model	**LCRRLOGIT**	
Error Components Logit	**ECLOGIT**	**NLOGIT ; ECM = ...**
Heteroscedastic Extreme Value	**HLOGIT**	**NLOGIT ; Het**
Nested Logit	**NLOGIT ; Tree = ...**	**NLOGIT ; Tree = ...**
Generalized Nested Logit	**GNLOGIT ; Tree = ...**	**NLOGIT ; GNL ; Tree = ...**
Random Parameters Logit	**RPLOGIT**	**NLOGIT ; RPL**
Berry, Levinsohn, Pakes RP Logit	**BLPLOGIT**	
Latent Class Random Parameters	**RPRRLOGIT**	
Generalized Mixed Logit	**GMXLOGIT**	
Nonlinear Random Parameters	**NLRPLOGIT**	
Latent Class Logit	**LCLOGIT**	**NLOGIT ; LCM**
2^K Latent Class	**LCLOGIT**	
Random Parameters Latent Class	**LCRPLOGIT**	
Multinomial Probit	**MNPROBIT**	**NLOGIT ; MNP**

NLOGIT contains an additional command that is used for a specific purpose:

NLCONVERT ; Lhs = ... ; Rhs = ... ; Other parameters $

This command is used to reconfigure a data set from a one line format to a multiple line format that is more convenient in *NLOGIT*. **NLCONVERT** is described in Chapter N18.

N3.7 Subcommand Summary

The following subcommands are used in *NLOGIT* model commands. The **BLOGIT**, **BPROBIT**, **BVPROBIT**, **MVPROBIT**, **OLOGIT** and **OPROBIT** commands have additional specifications that are documented in the *LIMDEP Econometric Modeling Guide* for these specific models. The specifications below are those that may appear in the **NLOGIT** command or the conditional logit extensions described above.

General Model Specification and Data Setup

Data on Dependent Variable

; Ranks indicates that data are in the form of ranks, possibly ties at last place.
; Shares indicates that data are in the form of proportions or shares.
; Frequencies indicates that data are in the form of frequencies or counts.
; Checkdata checks validity of the data before estimation.
; Wts = name specifies a weighting variable. (Noscale is not used here.)
; Scale (list of variables) = values for scaling loop specifies scaling of certain variables during iterations.
; Pds = spec indicates multiple choice situations for individuals. Used by RPL, LCM, ECM, MNP and by binary choice models to indicate a panel data set.

Specification of the Dependent Variable

; Lhs = names specifies model dependent variable(s).
Second Lhs variable indicates variable choice set size.
Third Lhs variable indicates specific choices in a universal choice set.
First Lhs variable is a set of utilities if **; MCS** is used.
; MCS requests data generated by Monte Carlo simulation.
; Choices = list lists names for alternatives.

Specification of Utility Functions

; Rhs = names lists choice varying attribute variables.
; Rh2 = names lists choice invariant characteristic variables.
; Model: alternative way to specify utility functions, followed by definitions of utility functions.
; Fix = list lists names of and values for coefficients that are to be fixed.
; Uset (list of alternatives) = list of values or [list of values] alternative method of specifying starting values or fixed coefficients.
; Lambda = value specifies coefficient to use for Box-Cox transformation.
; Attr = list lists names for attributes used in one line entry format.

Output Control

List and Retain Variables and Results

; Prob = name	keeps predicted probabilities from estimated model as variable.
; Keep = name	keeps predicted values from estimated model as variable. Used by **PROBIT** and **BLOGIT** only.
; Utility = name	keeps predicted utilities as variable.
; List	lists predicted probabilities and predicted outcomes with model results.
; Parameters	retains additional parameters as matrices. With RPL and LCM, keeps matrices of individual specific parameter means.
; WTP = list	lists specifications to retain computations of willingness to pay.

Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown),
; Robust	computes robust sandwich estimator for asymptotic covariance matrix.
; Cluster = spec	computes robust cluster corrected asymptotic covariance matrix.

Display of Estimation Results

; Show	displays model specification and tree structure.
; Describe	lists descriptive statistics for attributes by alternative.
; Odds	includes odds ratios in estimation results. Used only by **BLOGIT**.
; Crosstab	includes cross tabulation of predicted and actual outcomes.
; Table = name	adds model results to stored tables.

Marginal Effects

; Effects: spec	displays estimated marginal effects. Used by **NLOGIT**.
; Partial Effects	displays marginal effects, same as **; Marginal Effects**. Used by **PROBIT**, **BLOGIT**, **BVPROBIT**, **MVPROBIT**, **OLOGIT**, **OPROBIT**.
; Means	computes marginal effects using data means. Uses average partial effects if this is not specified.
; Pwt	uses probability weights to compute average partial effects.

Hypothesis Testing

; Test: spec	defines a Wald test of linear restrictions.
; Wald: spec	defines a Wald test of linear restrictions, same as **; Test: spec**.
; IAS = list	lists choices used with **CLOGIT** to test IIA assumption.

Optimization

Iterations Controls

; Alg = name	specifies optimization method.
; Maxit = n	sets the maximum iterations.
; 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.
; Set	keeps current setting of optimization parameters as permanent.
; Output = n	requests technical output during iterations; the level '**n**' is 1, 2, 3 or 4.

Starting Values

; Start = list	provides starting values for all model parameters.
; PR0 = list	provides starting values for free parameters only. (Generally not used.)

Constrained Estimation

; CML: spec	defines a constrained maximum likelihood estimator.
; Rst = list	imposes fixed value and equality constraints.
; Calibrate	fixes parameters at previously estimated values.
; ASC	initially fit model with just ASCs.

Criterion Function for CLOGIT

; GME [= number of support points] generalized maximum entropy. Used by **MLOGIT** and **CLOGIT**.

; Sequential	sequential two step estimator for nested logit. (Generally not used.)
; Conditional	conditional estimator for two step nested logit. (Generally not used.)

Simulation Based Estimation

; Pts = number	sets number of replications for simulation estimator. Used by ECM and MNP. (Also used by LCM to specify number of latent classes.)
; Shuffled	uses shuffled uniform draws to compute draws for simulations.
; Halton	uses Halton sequences for simulation based estimators.

Simulation Processor (BINARY CHOICE Command for PROBIT and BLOGIT)

; Simulation [= list of choices] simulates effect of changes in attributes on aggregate outcomes.

; Scenarios	specifies changes in attributes for simulations.
; Arc	computes arc elasticities during simulations.
; Merge	merges revealed and stated preference data during simulations.

Specific NLOGIT Model Commands

; LCM [= list of variables] specifies latent class model. Optionally, specifies variables that enter the class probabilities. (Command is also **LCLOGIT**.) Also used by **PROBIT** and **BLOGIT**.
; ECM = list of specifications specifies error components logit model. (Command is also **ECLOGIT**.)
; HEV specifies heteroscedastic extreme value model. (Command is also **HCLOGIT**.)

Heteroscedastic Models

; Het specifies a heteroscedastic model. Used by RPL, ECL and HEV.
; Hfr = names specifies heteroscedastic function in RPL, HEV and covariance heterogeneity form of nested logit model.
; Hfe = names specifies heteroscedasticity for ECM.

Nested Logit Model

; Tree = spec specifies tree structure in nested logit model.
; GNL specifies generalized nested logit model. (Command is also **GNLOGIT**.)
; RU1 specifies parameterization of second and third levels of the tree.
; RU2 specifies parameterization of second and third levels of the tree.
; RU3 specifies parameterization of second and third levels of the tree.
; IVSET: spec imposes constraints on inclusive value parameters.
; IVB = name keeps branch level inclusive values as a variable.
; IVL = name keeps limb level inclusive values as a variable.
; IVT = name keeps trunk level inclusive values as a variable.
; Prb = name keeps branch level probabilities as a variable.
; Cprob = name keeps conditional probabilities for alternatives.

Random Parameters Logit Model

; RPL [= list of variables] requests mixed logit model. Optionally specifies variables to enter means of random parameters.
; AR1 AR(1) structure for random terms in random parameters.
; Fcn: defines names and types of random parameters.
; Correlation specifies that random parameters are correlated.
; Hfr = names defines variables in heteroscedasticity. Also used by HEV and covariance heterogeneity.

Multinomial Probit

; MNP specifies multinomial probit model. (Command is also **MNPROBIT**.)
; EQC = list specifies a set of choices whose pairwise correlations are all equal.
; RCR = list specifies configurations for correlations for multinomial probit model. Also used by RPL.
; SDV = list specifies diagonal elements of covariance matrix. Also used by RPL and HEV.
; REM specifies random effects form of the model.

N4: Data for Binary and Ordered Choice Models

N4.1 Introduction

The data arrangement needed for discrete choice modeling depends on the model you are estimating. For the models described in Chapters N4-N15, you are fitting either cross section or panel models, and the observations are arranged accordingly. This is needed because in this part of the environment, you are fitting models for a single choice, and you need only a single observation to record that choice. For the models in Chapters N16 and N17 and N23-N33, the basic format of your data set will resemble a panel, even though it will usually be a cross section. This is because you are fitting models for choice sets with multiple alternatives, with one 'observation' (data record) for each alternative. For 'panel' models in the discrete choice environment, your data will consist of sets of groups of observations. This is developed in detail in Chapter N20.

N4.2 Grouped and Individual Data for Discrete 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.

With only two exceptions, all of the discrete response models estimated by *LIMDEP* and *NLOGIT* can be estimated with either individual or grouped data. The two exceptions are

- the multivariate probit model described in Chapter N12 (and E33)
- the multinomial probit model described in Chapter N27

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 in estimation 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. We do note, zero and unit 'proportions' data are sometimes indicative of a flawed data set, and can distort your results.

N4.3 Data Used in Estimation of Binary Choice Models

The following lists the specific features of the data needed to enable estimation of binary choice models. Certain features of the data that are inconsequential or irrelevant in linear regression modeling can impede estimation of a discrete choice model.

N4.3.1 The Dependent Variable

Data on the dependent variable for binary choice models may be individual or grouped. The estimation program will check internally, and adjust accordingly where necessary. The log likelihood function computed takes the same form 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.

N4.3.2 Problems with the Independent Variables

There is a special consideration for the independent variables in a 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. Notice, first, it spun for 100 iterations, which is almost certainly problematic. A probit model should take less than 10 iterations. Second, note that the log likelihood function is essentially zero, indicative of a perfect fit. Finally, note that the coefficients are 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. Finally, note the prediction table shows that the model predicts the dependent variable perfectly.

```
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 $
```

```
Maximum of   100 iterations. Exit iterations with status=1.
-----------------------------------------------------------------------------
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  =      6.0 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
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------+
| 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%)|
+------+----------------+----------------+----------------+
-----------------------------------------------------------------------
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                     97.872%
Specificity = actual 0s correctly predicted                     98.113%
Positive predictive value = predicted 1s that were actual 1s   100.000%
Negative predictive value = predicted 0s that were actual 0s    98.113%
Correct prediction = actual 1s and 0s correctly predicted       98.000%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s              .000%
False neg. for true pos. = actual 1s predicted as 0s              .000%
False pos. for predicted pos. = predicted 1s actual 0s            .000%
False neg. for predicted neg. = predicted 0s actual 1s            .000%
False predictions = actual 1s and 0s incorrectly predicted        .000%
-----------------------------------------------------------------------
```

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(z,y) $

```
Error   462: 0/1 choice model is inestimable. Bad variable = Z
Error   463: Its values predict 1[Y       = 1] perfectly.
Error   116: CALC - Unable to compute result. Check earlier message.
```

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. The software does not accept the job of determining the appropriate set of variables to include in the equation. This is up to the analyst.)

N4.3.3 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 equal to one in one of the two cells of the dependent variable, or vice versa. 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 $
```

In this subset of data, *d* is always one when *y* equals zero.

```
REJECT      ; y = 0 & d = 0 $
PROBIT      ; Lhs = y
            ; Rhs = one,x,d $
```

```
Nonlinear Estimation of Model Parameters
Method=NEWTON; Maximum iterations=100
1st derivs.     .35811D+02  -.19962D+02   .12369D+01
Itr  1 F=  .6981D+02 gtHg=  .6608D+01 chg.F=  .6981D+02 max|db|=  .9613D+01
1st derivs.     .49044D+01  -.74989D+01  -.29693D+00
Itr  2 F=  .4521D+02 gtHg=  .2003D+01 chg.F=  .2460D+02 max|db|=  .5302D+00
...
Itr  5 F=  .4282D+02 gtHg=  .1305D-03 chg.F=  .4625D-03 max|db|=  .2534D-04
1st derivs.    -.10201D-08  -.76739D-08  -.32583D-08
Itr  6 F=  .4282D+02 gtHg=  .2445D-08 chg.F=  .8516D-08 max|db|=  .4705D-09
                        * Converged
Normal exit from iterations. Exit status=0.
Function=  .69808104286D+02, at entry,  .42822158396D+02 at exit

+---------------------------------------------+
| Binomial Probit Model                       |
| Dependent variable                  Y       |
| Number of observations            100       |
| Iterations completed                6       |
| Log likelihood function     -42.82216       |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Index function for probability
 Constant|    -.93917517       .23373657    -4.018   .0001
 X       |    1.17177061       .24254318     4.831   .0000    .10291147
 D       |    1.53191876       .35304007     4.339   .0000    .45000000
```

The second model required 24 iterations to converge, and produced these results: The apparent convergence is deceptive, as evidenced by the standard errors.

```
Nonlinear Estimation of Model Parameters
Method=NEWTON; Maximum iterations=100
Itr 21 F=  .1660D+02 gtHg=  .3006D-04 chg.F=  .1614D-08 max|db|=  .2668D-01
1st derivs.    -.19854D-08   .10979D-08  -.28588D-14
Parameters:     .70037D+01   .14126D+01  -.63569D+01
Itr 22 F=  .1660D+02 gtHg=  .1787D-04 chg.F=  .5692D-09 max|db|=  .2530D-01
1st derivs.    -.72119D-09   .39979D-09   .11824D-13
Parameters:     .71645D+01   .14126D+01  -.65178D+01
Itr 23 F=  .1660D+02 gtHg=  .1064D-04 chg.F=  .2012D-09 max|db|=  .2406D-01
1st derivs.    -.26221D-09   .14554D-09  -.35527D-14
Parameters:     .73213D+01   .14126D+01  -.66746D+01
Itr 24 F=  .1660D+02 gtHg=  .6336D-05 chg.F=  .7126D-10 max|db|=  .2294D-01
                        * Converged
Normal exit:  24 iterations. Status=0, F=    16.60262
Function=  .26413087151D+02, at entry,  .16602624379D+02 at exit
```

```
-----------------------------------------------------------------------------
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  =     39.2 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|     7.32134       24162.78      .00  .9998 ***********  47365.49187
       X|     1.41264***      .39338     3.59  .0003       .64163   2.18365
       D|    -6.67459       24162.78      .00  .9998 ***********  47351.49594
--------+--------------------------------------------------------------------
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|                                    |
+------------------------------------------------------------------+
```

N4.3.4 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) or 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. You will generally be able to see their presence in the model results. The sample means for variables which contain missing values will usually look peculiar. In the small example below, *x*2 is a dummy variable. Both coefficients are one, which should be apparent in a sample of 1,000. The results, which otherwise look quite normal, suggest that missing values are being used as data in the estimation. With **SKIP**, the results, based on the complete data, look much more reasonable.

```
CALC        ; Ran(12345) $
SAMPLE      ; 1-1000 $
CREATE      ; x1 = Rnn(0,1)
            ; x2 = (Rnu(0,1) > .5) $
CREATE      ; y = (-.5 + x1 +x2+rnn(0,1)) > 0 $
CREATE      ; If(_obsno > 900)x2 = -999 $
PROBIT      ; Lhs = y
            ; Rhs = one,x1,x2 $
SKIP $
PROBIT      ; Lhs = y
            ; Rhs = one,x1,x2 $
```

```
Normal exit:   5 iterations. Status=0, F=    549.5785

-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    Y
Log likelihood function      -549.57851
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.08623*        .04601    -1.87  .0609     -.17640     .00394
      X1|     .81668***      .05541    14.74  .0000      .70807     .92529
      X2|     .00029*        .00015     1.95  .0517      .00000     .00058
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                     Y
Log likelihood function      -441.38989
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
       Y|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -.57123***      .07004    -8.16  .0000     -.70850   -.43396
      X1|     .97268***      .06611    14.71  .0000      .84310   1.10225
      X2|     .98082***      .10134     9.68  .0000      .78219   1.17945
--------+--------------------------------------------------------------------
```

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.)

N4.4 Bivariate Binary Choice

The bivariate probit model can be fit with either grouped data (you provide four proportions variables) or individual data (you provide two binary variables). In either case, the data must contain observations in both off diagonal cells. If your binary data are such that either the (y_1=0,y_2=1) or the (y_1=1,y_2=0) have no observations, then the correlation coefficient cannot be estimated, and the estimator will iterate endlessly, eventually 'converging' to a value of -1 or +1 for ρ. Note that this does not apply to the bivariate probit with selection, but that is a different model. For the grouped data case, if one of the proportions variables is always zero, the same problem will arise.

N4.5 Ordered Choice Model Structure and Data

Data for the ordered choice models must obey essentially the same rules as those for binary choice models. 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 that appears in the sample. In the grouped data case, J is one less than the number of Lhs variables you provide. There are two additional considerations for ordered choice modeling.

N4.5.1 Empty Cells

If you are using individual data, the Lhs variable must be coded 0,1,...,J. All the values must be present in the data. *NLOGIT* will look for empty cells. If there are any, the estimation is halted. (If the value 'j' is not represented in the data, then the threshold parameter, μ_j cannot be estimated. In this case, you will receive a diagnostic such as

```
ORDE, Panel, BIVA PROBIT: A cell has (almost) no observations.
Empty cell: Y never takes the value 2.
```

This diagnostic means exactly what it says. The ordered probability model cannot be estimated unless all cells are represented in the data

N4.5.2 Coding the Dependent Variable

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 *NLOGIT* 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.

N4.6 Constant Terms

In general, discrete choice models should contain constant terms. Omitting the constant term is analogous to leaving the constant term out of a linear regression. This imposes a restriction that rarely makes sense.

The ordered probit 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.) Other programs sometimes use different normalizations of the model. For example, if the constant term is forced to equal zero, then one will instead, have a nonzero threshold parameter, μ_0, which equals zero in the presence of a nonzero constant term.

In the more general multinomial choice models, when choices are unlabeled, there may be no case for including alternative specific constants (ASCs) in the model, since they are not actually associated with a particular choice. On the other hand, ASCs in a model with unlabeled choices might simply imply that after controlling for the effects of the attributes, the indicated alternative is chosen more or less frequently than the base alternative. It is possible that this might occur because the alternative is close to the reference alternative or that culturally, those undertaking the experiment might tend to read left to right. Failure to include ASCs in the model would in this case correlate the alternative order effect into the other estimated parameters, possibly distorting the model results.

N5: Models for Binary Choice

N5.1 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 *NLOGIT*'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 and ordered choice models. *NLOGIT* 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 detail in Chapter E27. We will focus on the first two of these here.

N5.2 Modeling Binary Choices

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 *NLOGIT* estimates, but the exact interpretation is a function of the modeling framework.

N5.2.1 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, $\mathbf{x}$. The choice of alternative 1 reveals that $U_1 > U_0$, or that

$$\varepsilon_0 - \varepsilon_1 < \boldsymbol{\beta}_0'\mathbf{x} - \boldsymbol{\beta}_1'\mathbf{x}.$$

Let $\varepsilon = \varepsilon_0 - \varepsilon_1$ and let $\boldsymbol{\beta}'\mathbf{x}$ represent the difference on the right hand side of the inequality – $\mathbf{x}$ 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 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.

N5.2.2 Modeling Approaches

NLOGIT provides estimators for three approaches to formulating the binary choice models described above:

Parametric Models – Probit, Logit, Extreme etc.

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.

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 *NLOGIT*, 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 *NLOGIT* 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.

N5.2.3 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 *NLOGIT*, but will not be pursued here.

N5.3 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 *NLOGIT* 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.

N5.4 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 $\boldsymbol{\beta}$ 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.)

N5.5 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.)

N6: Probit and Logit Models: Estimation

N6.1 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 two of *NLOGIT*'s qualitative dependent variable model estimators, the probit and logit models. More extensive treatment and technical background are given in Chapters E27-29. 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.

N6.2 Probit and Logit Models for Binary Choice

These parametric model formulations are provided as internal procedures in *NLOGIT* 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

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

N6.3 Commands

The basic model commands for the two binary choice models of interest here are:

```
PROBIT      ; Lhs = dependent variable
or BLOGIT   ; Rhs = regressors $
```

Data on the dependent variable may be either individual or proportions for both cases. When the dependent variable is binary, 0 or 1, the model command may be **LOGIT** – the program will inspect the data and make the appropriate adjustments for estimation of the model.

N6.4 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.

N6.4.1 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

$$\log L_0 = n[P\log P + (1-P)\log(1-P)]$$

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.
```

N6.4.2 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 next set of results examines the success of the prediction rule

Predict $y_i = 1$ if $P_i > P^*$ 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|         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%)|
+------+----------------+----------------+----------------+
+---------------------------------------------------------+
|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  |    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%
----------------------------------------------------------------------
```

N6.4.3 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

If the matrix is not larger than 5×5, it will be displayed in full. If it is larger, the covariance matrix will be placed in the matrix area in your project window with the name COV.[B^]. By double clicking the name, you can display the matrix in a window. An example appears in Figure N6.1 below.

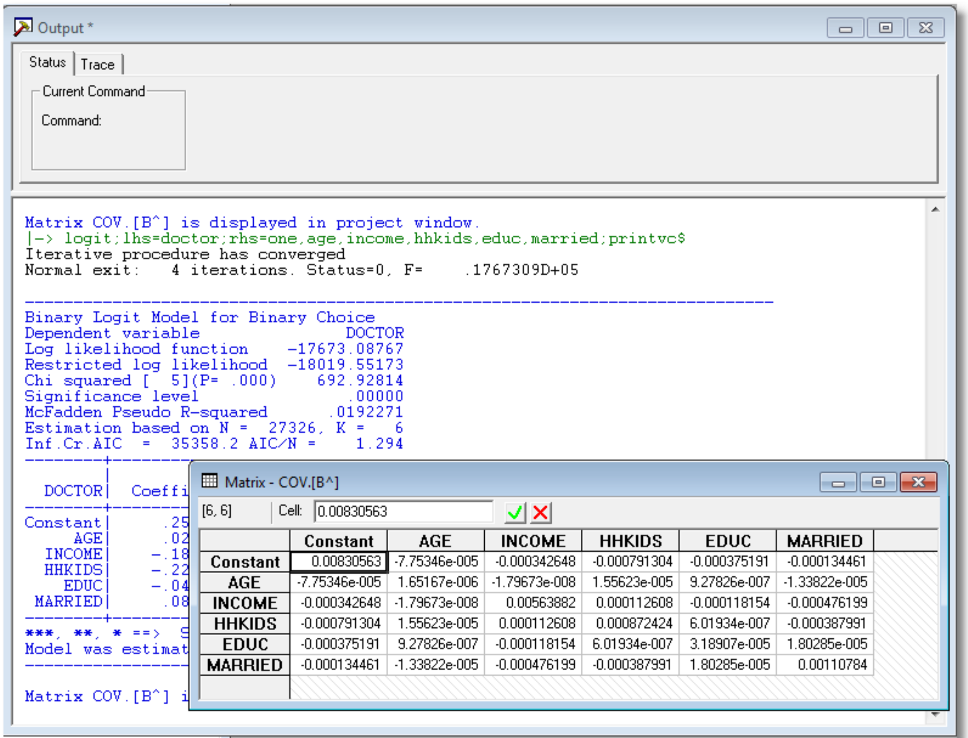


Figure N6.1 Covariance Matrix

N6.4.4 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

Last Model: *b_variables*

Last Function: Prob(y = 1 | **x**) = F(**b′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) $
```

N6.5 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.)

N6.5.1 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 $n/(n-1)$.)

N6.5.2 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:

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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.        |
+---------------------------------------------------------------+
```

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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:

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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.

N6.5.3 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 $(N\text{-}1)/(N\text{-}K)$ where N is the total number of observations in the entire sample and K is the number of parameters (rows in **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.

N6.6 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 **x**. 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 **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

$$\begin{aligned}\Delta F_z &= \text{Prob}[y = 1\,|\,z = 1] - \text{Prob}[y = 1\,|\,z = 0]\\ &= F(\boldsymbol{\beta}'\mathbf{x} + \alpha z\,|\,z = 1) - F(\boldsymbol{\beta}'\mathbf{x} + \alpha z\,|\,z = 0)\\ &= F(\boldsymbol{\beta}'\mathbf{x} + \alpha) - F(\boldsymbol{\beta}'\mathbf{x}).\end{aligned}$$

NLOGIT examines the variables in the model and makes this adjustment automatically.

There are two programs in *NLOGIT* 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** 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.

N6.6.1 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 kth 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 source 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** (or just **PARTIALS**) 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 $

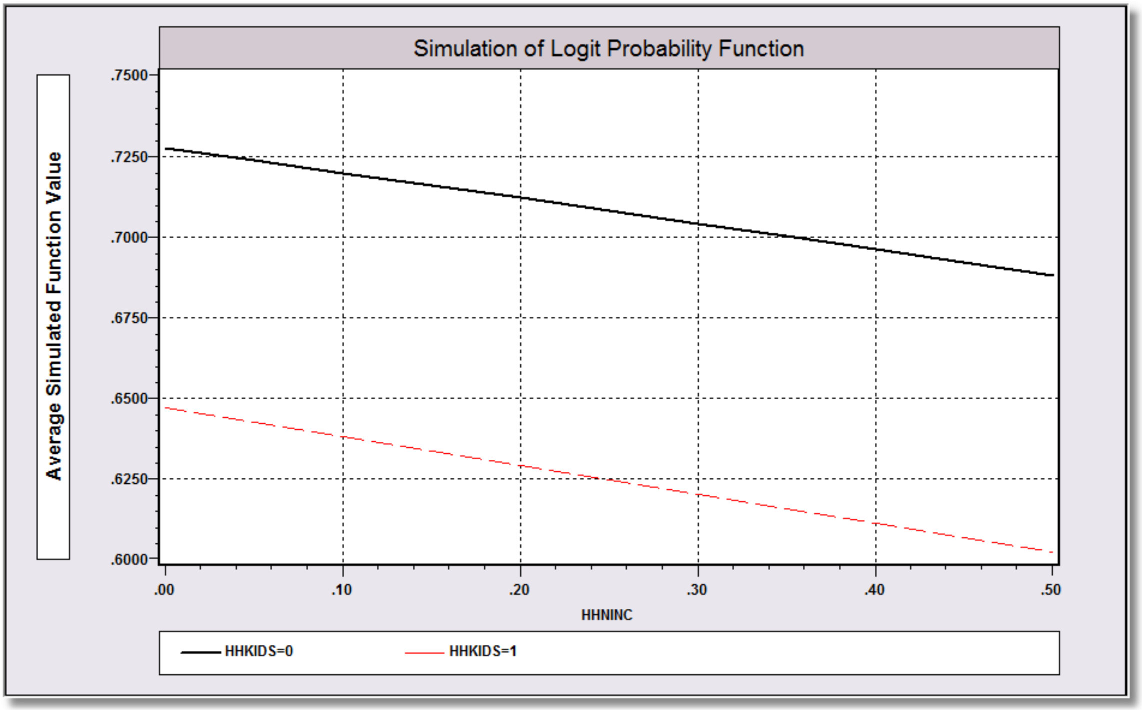


Figure N6.2 Probabilities Varying with Income

N6.7 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. 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.

N6.8 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 *NLOGIT* 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.

N6.9 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. Two general options that are likely to be useful are

; Keep = name to retain predicted values
; 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 **β** and **γ** with **; Rst = ...**, (i.e., force an element in **β** to equal one in **γ**), 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(γ_1*female* + γ_2*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%)|
+------+----------------+----------------+----------------+
```

N7: Tests and Restrictions in Models for Binary Choice

N7.1 Introduction

We define models in which the response variable being described is inherently discrete as qualitative response (QR) models. Chapter N6 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.

N7.2 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.

N7.2.1 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 *NLOGIT* allows.) The Wald statistic is computed by the formula

$$W \;=\; \mathbf{c}\left(\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)\left\{Est.Asy.Var\left(\hat{\boldsymbol{\beta}}\right)\right\}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)'\right]^{-1}\mathbf{c}\left(\hat{\boldsymbol{\beta}}\right)$$

where
$$\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right) \;=\; \frac{\partial \mathbf{c}\left(\hat{\boldsymbol{\beta}}\right)}{\partial \hat{\boldsymbol{\beta}}'}$$

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.

N7.2.2 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   =    549.2873320
[CALC] DF      =     12.0000000
[CALC] PVALUE  =       .0000000
Calculator: Computed   4 scalar results
```

The homogeneity test shown above can be automated in the probit command. For the preceding, we would use

```
PROBIT      ; For [(test) female] ; Lhs = doctor ; Rhs = x ; Quiet $
```

This will produce the following:

```
----------------------------------------------------
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
FEMALE   = ****    27326
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
Subsample analyzed for this command is FEMALE   =  1
Full pooled sample is used for this iteration.
-----------------------------------------------------------------------
Homogeneity Test for Estimated Model
-----------------------------------------------------------------------
The model was estimated for  2 subsamples and the full sample
The likelihood ratio statistic is 2[Sum(g=1...G)logL(g) - logL(pooled)]
Chi squared =    549.2873   Estimated degrees of freedom =   6
Estimated P value for this test is  .0000
-----------------------------------------------------------------------
```

N7.2.3 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;

$$LM = \mathbf{g}\left(\hat{\beta}_R\right)'\left[Est.Asy.Var\left\{\mathbf{g}\left(\hat{\beta}_R\right)\right\}\right]^{-1}\mathbf{g}\left(\hat{\beta}_R\right)$$

where $\hat{\beta}_R$ = MLE of the parameters of the model, with restrictions imposed

$\mathbf{g}\left(\hat{\beta}_R\right)$ = derivatives of log likelihood of full model, evaluated at $\hat{\beta}_R$

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 *NLOGIT* 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

β_age + β_educ = 0
β_married = 0
β_hhninc + β_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
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.

N7.3 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.

N7.3.1 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*.

N7.3.2 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,

$$LM = \left(\sum_{i=1}^{N} d_i\mathbf{z}_i\right)' \left(\sum_{i=1}^{N} c_i\mathbf{z}_i\mathbf{z}_i'\right)^{-1} \left(\sum_{i=1}^{N} d_i\mathbf{z}_i\right)$$

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
```

N7.4 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 $
```

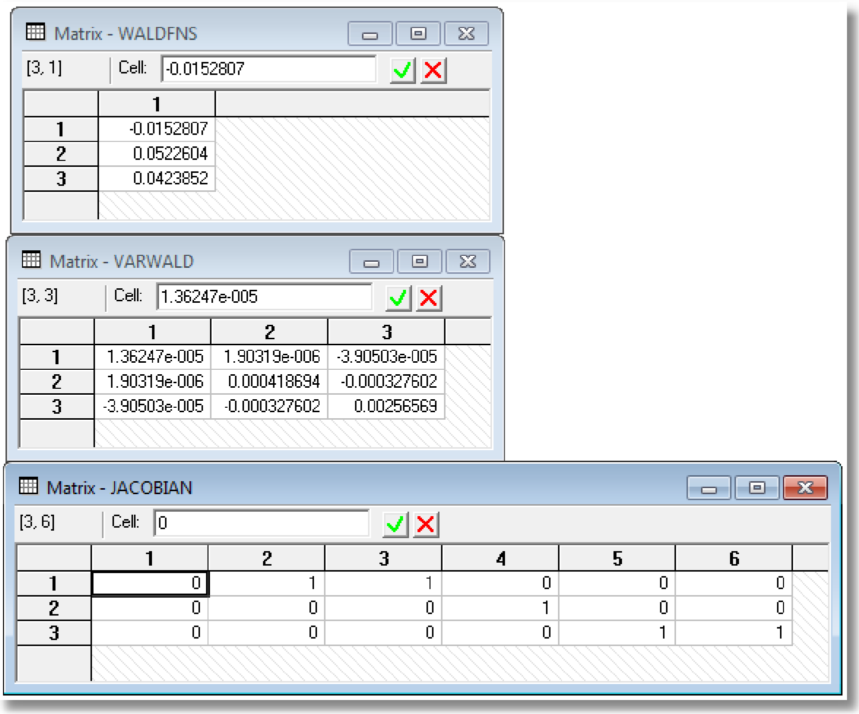
(The latter restriction doesn't make much sense, but we can test it anyway.) The results of this pair of commands are shown below. (The **PROBIT** command was shown earlier.)

```
-----------------------------------------------------------------------------
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



Matrix - WALDFNS

[3, 1] Cell: -0.0152807

	1
1	-0.0152807
2	0.0522604
3	0.0423852

Matrix - VARWALD

[3, 3] Cell: 1.36247e-005

	1	2	3
1	1.36247e-005	1.90319e-006	-3.90503e-005
2	1.90319e-006	0.000418694	-0.000327602
3	-3.90503e-005	-0.000327602	0.00256569

Matrix - JACOBIAN

[3, 6] Cell: 0

	1	2	3	4	5	6
1	0	1	1	0	0	0
2	0	0	0	1	0	0
3	0	0	0	0	1	1

Figure N7.1 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
```

N7.5 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 R13.6.3.) 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

N8: Extended Binary Choice Models

N8.1 Introduction

NLOGIT 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.

N8.2 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 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

$$\text{Prob}(y_i = 1\,|\,\mathbf{x}_i,\varepsilon_i) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i + \sigma\varepsilon_i)}.$$

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 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.

N8.3 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 in Section E29.4.

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: $\Phi(\mathbf{b'x} + \mathrm{a}y_2) = \mathrm{Prob}(y_1 = 1 \mid \mathbf{x},y_2)$.

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.
----------------------------------------------------------------------
```

N9: Fixed and Random Effects Models for Binary Choice

N9.1 Introduction

The parametric models discussed in Chapters N5-N6 are extended to panel data formats. Four specific parametric model formulations are provided as internal procedures in *NLOGIT* 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:** $\text{Prob}[y_{it} = 1|\text{class j}] = F(\boldsymbol{\beta}_j'\mathbf{x}_{it})$,
  $\text{Prob}[\text{class} = \text{j}] = F_j(\boldsymbol{\theta})$.

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 N10.

The applications in this chapter are based on the German health care data used throughout the documentation. 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.*

N9.2 Commands

The essential model command for the models described in this chapter are

```
PROBIT      ; Lhs = dependent variable
or LOGIT    ; 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 $
```

The **; Pds = variable** is optional in the **SETPANEL** command. The default name for the created variable is *ti*. You may change this with **; Pds**. 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. *NLOGIT* 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*.

N9.3 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 N9.1. The Model Estimates dialog box is selected at the bottom of the Output page, then the clustering is specified in the next dialog box.

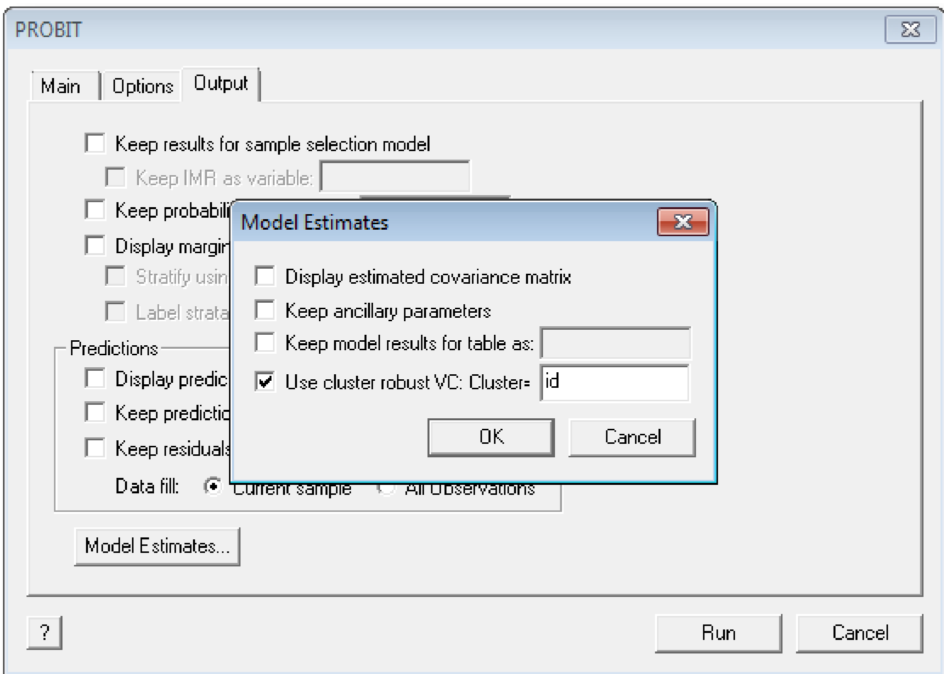


Figure N9.1 Command Builder for a Probit Model

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.

The commands are:

```
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.
----------------------------------------------------------------------
```

N9.4 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 ; Lhs = dependent variable
or LOGIT ; 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.

Further documentation and technical details on fixed effects models for binary choice appear in Chapter E30.

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).

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 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. 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.

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.

The commands are:

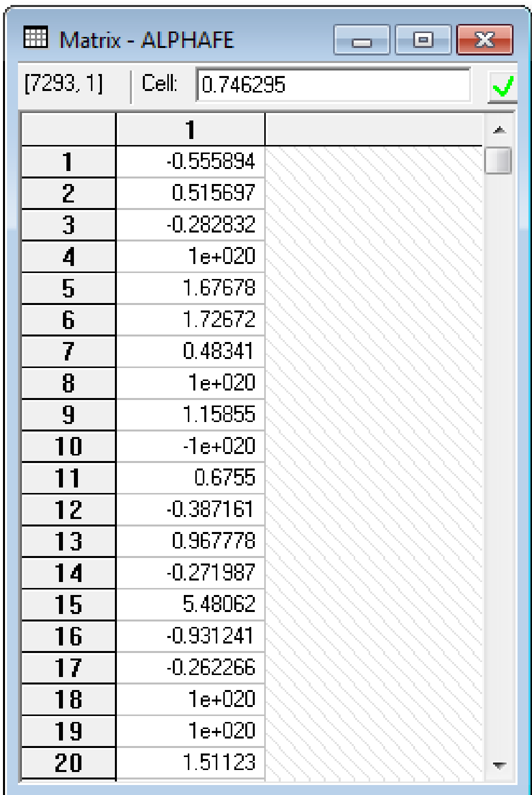
```
SAMPLE      ; All $
SETPANEL    ; Group = id ; Pds = ti $
NAMELIST    ; x = age,educ,hhninc,newhsat $
PROBIT      ; Lhs = doctor ; Rhs = x,one
            ; Partial Effects $
PROBIT      ; Lhs = doctor ; Rhs = x
            ; FEM
            ; Panel
            ; Parameters
            ; Partial 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.
-----------------------------------------------------------------------------
```



	1
1	-0.555894
2	0.515697
3	-0.282832
4	1e+020
5	1.67678
6	1.72672
7	0.48341
8	1e+020
9	1.15855
10	-1e+020
11	0.6755
12	-0.387161
13	0.967778
14	-0.271987
15	5.48062
16	-0.931241
17	-0.262266
18	1e+020
19	1e+020
20	1.51123

Figure N9.2 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
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.
-----------------------------------------------------------------------------
```

N9.5 Conditional MLE of the Fixed Effects Logit Model

Two nonlinear models, the binomial logit and Poisson regression can be estimated by conditional maximum likelihood. This is a specialized approach that was devised to deal with the problem of large numbers of incidental parameters discussed in the preceding section. (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, β, 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

$$A(T,s) = 0 \text{ if } T < s \text{ and } A(T,0) = 1.$$

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.

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.

N9.5.1 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

N9.5.2 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%. Partial effects are shown as well.

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 100,000 groups, you must use the conditional estimator. If, on the other hand, *T* is larger than, say, 10, and *N* is less than 100,000, then the unconditional estimator might be preferred. The additional consideration discussed in the next section might also weigh in favor of the unconditional estimator.

N9.5.3 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.

N9.5.4 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.

N9.6 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,

$$u_i \sim \text{N}[0,\sigma_u^{\,2}],$$

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 these structural parameters appears at the end of this section.

The model command for this form of the model is

PROBIT ; Lhs = dependent variable
or LOGIT ; Rhs = independent variables - not including one
; Panel
; Random Effects $

NOTE: For this model, your Rhs list should include a constant term, *one*.

Partial 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 Lagrange multiplier (LM) statistic developed by Greene and McKenzie (2015) is reported with the other results, as shown in the example below.

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
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+ρ. 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     -15614.49128
Restricted log likelihood   -16639.23907
Chi squared [  1](P= .000)    2049.49558
Significance level                .00000
(Cannot compute pseudo R2.  Use RHS=one
to obtain the required restricted logL)
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  =  31241.0 AIC/N =    1.143
Unbalanced panel has   7293 individuals
- ChiSqd[1] tests for random effects  -
LM   ChiSqd  807.866   P value    .00000
LR   ChiSqd 2049.496   P value    .00000
Wald ChiSqd 1432.269   P value    .00000
--------+--------------------------------------------------------------------
        |                 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
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
--------+--------------------------------------------------------------------
```

N10: Random Parameter Models for Binary Choice

N10.1 Introduction

The probit and logit models 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

$$F = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i), \qquad f = \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)[1 - \Lambda(\boldsymbol{\beta}'\mathbf{x}_i)]$$

N10.2 Probit and Logit Models with Random Parameters

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 (2001). (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

$$\text{Prob}[y_{it} = 1|\, \mathbf{x}_{it}, \boldsymbol{\beta}_i] = F(\boldsymbol{\beta}_i'\mathbf{x}_{it}),\ i = 1,...,N,\ t = 1,...,T_i.$$

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)

$$\text{E}[\boldsymbol{\beta}_i|\, \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

(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

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i \text{ where } \mathbf{v}_i \sim \text{N}[\mathbf{0},\mathbf{I}].$$

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.

N10.2.1 Command for the Random Parameters Models

The basic model command for this form of the model is

```
PROBIT      ; Lhs = dependent variable
or LOGIT    ; 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,

$$v_{kit} = \rho_k v_{ki,t-1} + w_{kit}.$$

N10.2.2 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].)

The commands are:

```
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
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:

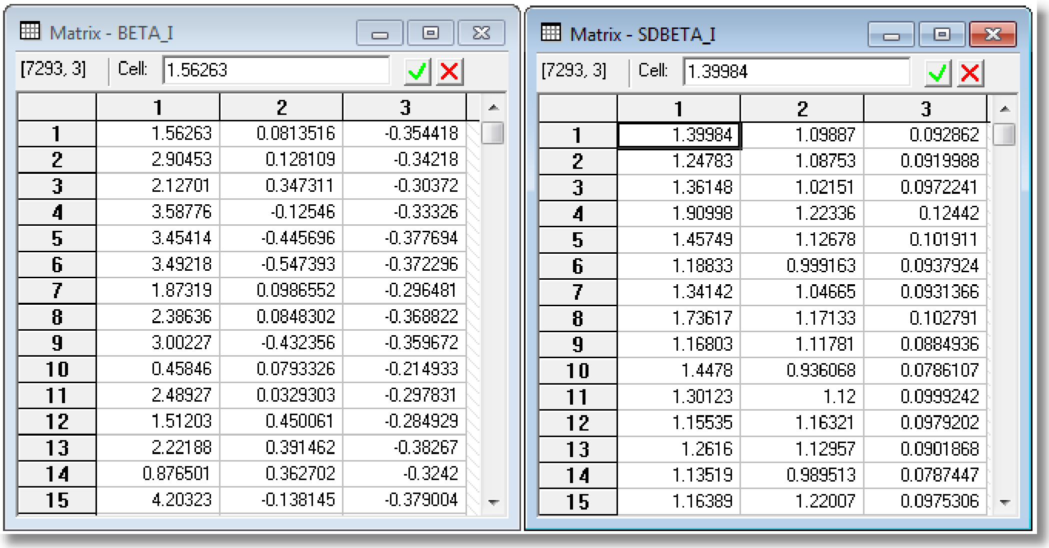
Matrices:	*b*	=	estimate of θ
	varb	=	asymptotic covariance matrix for estimate of θ.
	gammaprm	=	the estimate of Γ
	beta_i	=	individual specific parameters, if **; Par** is requested
	sdbeta_i	=	individual specific parameter standard deviations if **; Par** is requested
Scalars:	*kreg*	=	number of variables 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 N10.1.



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 N10.1 Estimated Conditional Parameter Means

N10.2.3 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. 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 *NLOGIT* 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 *NLOGIT*.

N10.2.4 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 **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:

; 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**
*x*1	α_1
*x*3	α_2
one	$\beta_1 + \sigma_1 v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
*x*2	$\beta_2 + \sigma_2 v_{i2} + \gamma_{21} v_{i1} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$
*x*4	$\beta_3 + \sigma_3 v_{i3} + \gamma_{31} v_{i1} + \gamma_{32} v_{i2} + \delta_{11} z_{i1} + \delta_{12} z_{i2}$

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.

N10.2.5 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 *NLOGIT*. 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.

N10.3 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      ; Lhs = dependent variable
or LOGIT    ; Rhs = independent variables
            ; Panel or Pds = fixed periods or count variable
            ; LCM $
```

The default number of support points is five. You may set J from two to 30 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.

Estimates retained by this model include

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 $J\times(K+1)$ 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

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

N10.3.1 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. 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 N10.2 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.

DOCTOR	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence 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

N11: Semiparametric and Nonparametric Models for Binary Choice

N11.1 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 *NLOGIT*, 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.

N11.2 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 N11.4.

N11.2.1 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 *NLOGIT* 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.

N11.2.2 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 1with

; 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 **b** be the maximum score estimate, *R* be the number of bootstrap replications, and $\mathbf{d}_i$ be the *i*th 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 *k*th column of an identity matrix. The sample score function is evaluated at this vector, and the *k*th 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.

N11.2.3 General Options for MSCORE

The following general options used with the nonlinear estimators in *NLOGIT* 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.

N11.2.4 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 *NLOGIT'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.

N11.3 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

$$\log L \;=\; \frac{1}{n}\sum\nolimits_{i=1}^{n} \left[y_i \log P_n(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)\log(1-P_n(\boldsymbol{\beta}'\mathbf{x}_i))\right]$$

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,

$$P_n(\boldsymbol{\beta}'\mathbf{x}_i) \;=\; \frac{\sum\nolimits_{j=1}^{n} \frac{y_j}{h} K\left(\frac{\boldsymbol{\beta}'(\mathbf{x}_i - \mathbf{x}_j)}{h}\right)}{\sum\nolimits_{j=1}^{n} \frac{1}{h} K\left(\frac{\boldsymbol{\beta}'(\mathbf{x}_i - \mathbf{x}_j)}{h}\right)}.$$

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.

N11.3.1 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 predicted values
; 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 *NLOGIT*'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

N11.3.2 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

(In previous versions, the command was **; Marginal Effects**. This form is still supported.) Partial 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

N11.3.3 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.
-----------------------------------------------------------------------------
```

N11.4 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 |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| < 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(2πz) if |z| < .5, 0 else
7. Triangle:	$K[z]$	= 1 - |z|, if |z| <= 1, 0 else
8. Parzen:	$K[z]$	= 4/3 - 8z^2 + 8|z|^3 if |z| <= .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. 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.

N11.4.1 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, $CVMSPE(h) = (1/n)\,\Sigma_i\,[y_i - F_i^*(x_i)]^2.$

N11.4.2 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.
```

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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 Value	Predicted Value 0	Predicted Value 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 Value	Predicted Probability Prob(y=0)	Predicted Probability 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

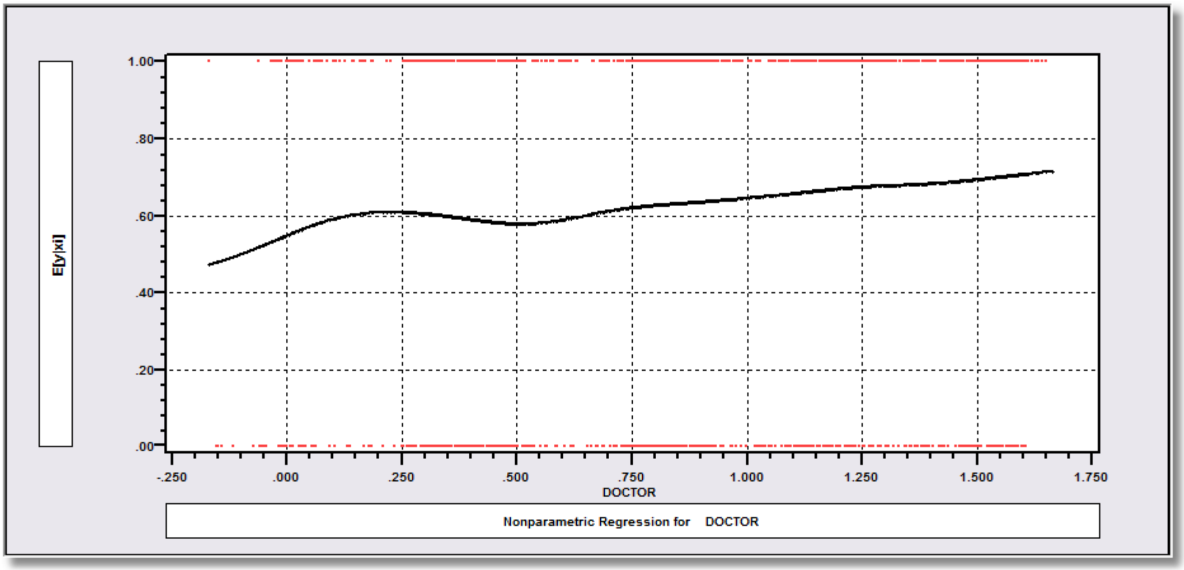


Figure N11.1 Nonparametric Regression

N12: Bivariate and Multivariate Probit and Partial Observability Models

N12.1 Introduction

The basic formulation of the models in this chapter is the bivariate probit model:

$$z_{i1} = \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1},\ \ y_{i1} = 1 \text{ if } z_{i1} > 0,\ y_{i1} = 0 \text{ otherwise,}$$

$$z_{i2} = \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \varepsilon_{i2},\ \ y_{i2} = 1 \text{ if } z_{i2} > 0,\ y_{i2} = 0 \text{ otherwise,}$$

$$[\varepsilon_{i1},\varepsilon_{i2}] \sim \text{bivariate normal (BVN) } [0,0,1,1,\rho],\ -1 < \rho < 1,$$

individual observations on y_1 and y_2 are available for all i.

(This model is also available for grouped (proportions) data. See Section N12.2.2.) 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.

N12.2 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 *NLOGIT* 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 $
```

N12.2.1 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 $
...
ENDPROCEDURE $
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. 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);

$$w_{ij} = \text{population proportion / sample proportion}, \; i,j = 0,1.$$

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* and *NLOGIT*.
; 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.

N12.2.2 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.

N12.2.3 Heteroscedasticity

All bivariate probit specifications, including the basic two equation model, the sample selection model (Section N12.4), and the Meng and Schmidt partial observability model (Section N12.7), may be fit with a multiplicative heteroscedasticity specification. The model is the same as the univariate probit model;

$$\varepsilon_i \sim N\{0, [\exp(\boldsymbol{\gamma}_i'\mathbf{z}_i)]^2\}, i = 1 \text{ and/or } 2.$$

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.)

N12.2.4 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 $\rho = 0$, 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 $\rho = 0$. 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.

N12.2.5 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*.

N12.2.6 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 below. 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.

NLOGIT analyzes the conditional mean function

$$E[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2] = \text{Prob}[y_1 = 1, y_2 = 1 \mid \mathbf{x}_1, \mathbf{x}_2, \rho] \,/\, \text{Prob}[y_2 = 1|\mathbf{x}_1].$$

This is the function analyzed in the bivariate probit marginal effects processor. The bivariate probit estimator in *NLOGIT* 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 N12.4 below.) The conditional mean computed for these two models would be identical,

$$E[y_1|y_2 = 1] = \Phi_2 [w_1, w_2 , \rho] / \Phi(w_2)$$

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 (2012).

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 *NLOGIT* 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 / Y2	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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	
Variable	Direct Efct x1	Indirect Efct x2	Direct Efct h1	Indirect 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 / Y2	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence 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
```

N12.3 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:

$$y_1^* = \mu + \varepsilon_1, \quad y_1 = 1(y_1^* > 0)$$

$$y_2^* = \mu + \varepsilon_2, \quad y_2 = 1(y_2^* > 0)$$

$$(\varepsilon_1,\varepsilon_2) \sim N_2[(0,0),(1,1,\rho)]$$

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. *NLOGIT* 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, $\mathbf{R}$, 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

$$y_1^* = \mu + \varepsilon_1, \quad y_1 = 1(y_1^* > 0)$$

$$y_2^* = \mu + \varepsilon_2, \quad y_2 = 1(y_2^* > 0)$$

$$...$$

$$y_M^* = \mu + \varepsilon_M, \quad y_M = 1(y_M^* > 0)$$

$$(\varepsilon_1,\varepsilon_2,...,\varepsilon_M) \sim N_M[\mathbf{0},\mathbf{R}]$$

The appropriate estimator would be *NLOGIT*'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 $
ENDPROCEDURE $
EXECUTE     ; Quiet $
```

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.

N12.4 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.

N12.5 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.) *NLOGIT* 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.

N12.6 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:

$$\frac{\partial E[y_1 \mid y2 = 1, \mathbf{x}_1, \mathbf{x}_2]}{\partial \mathbf{x}_1} = \frac{g_1}{\Phi(\boldsymbol{\beta}'\mathbf{x}_2)}\boldsymbol{\beta}_1 \ = \text{ direct effects}$$

$$\frac{\partial E[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2]}{\partial \mathbf{x}_2} = \left[\frac{g_2}{\Phi(\boldsymbol{\beta}'\mathbf{x}_2)} - \frac{\Phi_2(\boldsymbol{\beta}'\mathbf{x}_1, \boldsymbol{\beta}'\mathbf{x}_2, \rho)\phi(z_2)}{[\Phi(\boldsymbol{\beta}'\mathbf{x}_2)]^2}\right]\boldsymbol{\beta}_2 \ = \text{indirect effects}$$

The unconditional mean function is

$$\text{E}[y_1 \mid \mathbf{x}_1, \mathbf{x}_2] = \Phi(\boldsymbol{\beta}_2'\mathbf{x}_2)\,\text{E}[y_1 \mid y_2 = 1, \mathbf{x}_1, \mathbf{x}_2] + [1-\Phi(\boldsymbol{\beta}_2'\mathbf{x}_2)]\,\text{E}[y_1 \mid y_2 = 0, \mathbf{x}_1, \mathbf{x}_2]$$

$$= \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1 + \gamma_1, \boldsymbol{\beta}_2'\mathbf{x}_2, \rho) + \Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_1, -\boldsymbol{\beta}_2'\mathbf{x}_2, -\rho)$$

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. 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.’

N12.7 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.

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
```

N12.8 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]

	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]

	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 N12.1 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
```

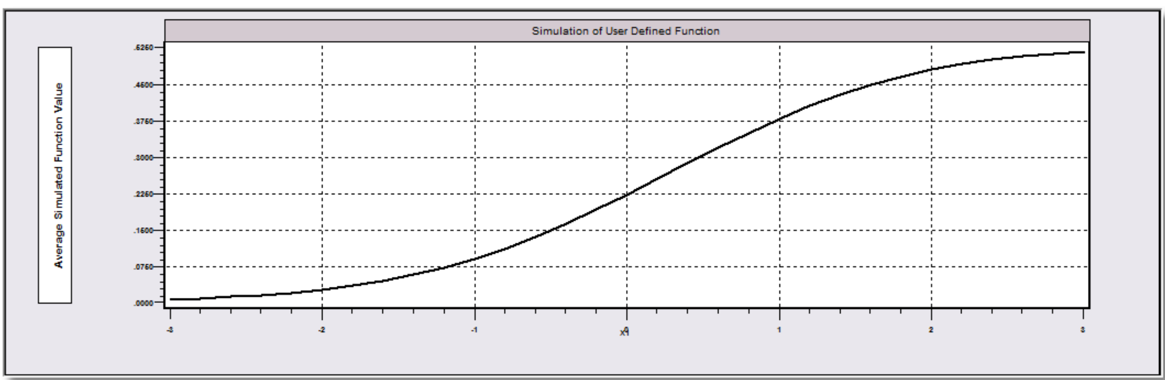


Figure N12.2 Simulation of Estimated Model

N12.9 Multivariate Probit Model

The multivariate probit model is the extension to M equations of the bivariate probit model

$$y_{im}^* = \boldsymbol{\beta}_m'\mathbf{x}_{im} + \varepsilon_{im},\ m = 1,\ldots,M$$

$$y_{im} = 1 \text{ if } y_{im}^* > 0, \text{ and } 0 \text{ otherwise.}$$

$$\varepsilon_{im},\ m = 1,\ldots,M \sim \text{MVN}\,[\mathbf{0},\mathbf{R}]$$

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.

N12.9.1 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

N12.9.2 Partial 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

$$E[y_1|y_2=1,\ldots,y_M=1] = \text{Prob}(y_1=1,\ldots,y_M=1)/\text{Prob}(y_2=1,\ldots,y_M=1) = P_{1\ldots M} / P_{2\ldots M} = E_1.$$

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 $\mathbf{z}_m = \mathbf{x}'\boldsymbol{\gamma}_\mathbf{m} = \boldsymbol{\beta}_m'\mathbf{x}_m$. ($\gamma_m$ will usually have some zeros in it unless all regressors appear in all equations.) Then,

$$\frac{\partial E_1}{\partial \mathbf{x}} = \sum_{m=1}^{M} \left(\frac{1}{P_{2\ldots M}} \frac{\partial P_{1\ldots M}}{\partial \mathbf{z}_m} \right) \boldsymbol{\gamma}_m - E_1 \sum_{m=2}^{M} \left(\frac{1}{P_{2\ldots M}} \frac{\partial P_{2\ldots M}}{\partial \mathbf{z}_m} \right) \boldsymbol{\gamma}_m$$

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.

N12.9.3 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 N12.4. The model structure is

$$\begin{aligned}
y_{i1}^* &= \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1}\\
y_{i2}^* &= \boldsymbol{\beta}_2'\mathbf{x}_{i2} + \varepsilon_{i2}\\
&\ldots\\
y_{i,M-1}^* &= \boldsymbol{\beta}_{M-1}'\mathbf{x}_{i,M-1} + \varepsilon_{I,M-1}\\
y_{iM}^* &= \boldsymbol{\beta}_M'\mathbf{x}_{iM} + \varepsilon_{iM}\\
y_{im} &= 1 \text{ if } y_{im}^* > 0, \text{ and } 0 \text{ otherwise}\\
\varepsilon_{im}, m &= 1,...,M \ \sim \text{MVN}\,[\mathbf{0},\mathbf{R}]\\
&y_{i,1},y_{i,2},\ldots,y_{i,M\text{-}1} \text{ only observed when } y_{iM} = 1
\end{aligned}$$

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.

N13: Ordered Choice Models

N13.1 Introduction

The basic ordered choice model is based on the latent regression,

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \ \varepsilon_i \sim F(\varepsilon_i\,|\boldsymbol{\theta}),\ \text{E}[\varepsilon_i|\mathbf{x}_i] = 0,\ \text{Var}[\varepsilon_i|\mathbf{x}_i] = 1,$$

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. 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 N14.2.) 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

$$\text{Prob}[y_i = j] = \text{Prob}[y_i^* \text{ is in the } j\text{th range}].$$

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.) (See, as well, Chapters E34-E36.) The functional forms of the two models considered here are

Probit

$$F(z) = \int_{-\infty}^{z} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt = \Phi(z), \qquad f(z) = \phi(z),$$

Logit

$$F(z) = \frac{\exp(z)}{1+\exp(z)} = \Lambda(z), \qquad f(z) = \Lambda(z)[1 - \Lambda(z)].$$

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. 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 N14. Panel data models for ordered choice are discussed in Chapter N15.

N13.2 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 model, add

; Model = Logit

to the command or change the verb to **OLOGIT**. The standardized logistic distribution (mean zero, standard deviation approximately 1.81) is used as the basis of the model instead of the standard normal.

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.

N13.3 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. *NLOGIT* 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 *NLOGIT* 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.

N13.4 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.

N13.4.1 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.

N13.4.2 Saved Results

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 $E[y_i] = \sum_{i=0}^{J} i \times \text{Prob}[Y_i = i]$.

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 EFFECTS ; Effects: hhninc ; Outcome = 3 $

and likewise for **SIMULATE**. A full examination of all outcomes is obtained by using

; Outcome = *

N13.5 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,

$$\text{Prob[cell j]} = F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i).$$

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:

$$\partial\text{Prob[cell } j]/\partial\mathbf{x}_i = [f(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i) - f(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i)] \times \boldsymbol{\beta},$$

where $f(.)$ is the appropriate density for the standard normal, $\phi(\bullet)$, logistic density, $\Lambda(\bullet)(1-\Lambda(\bullet))$, 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
```

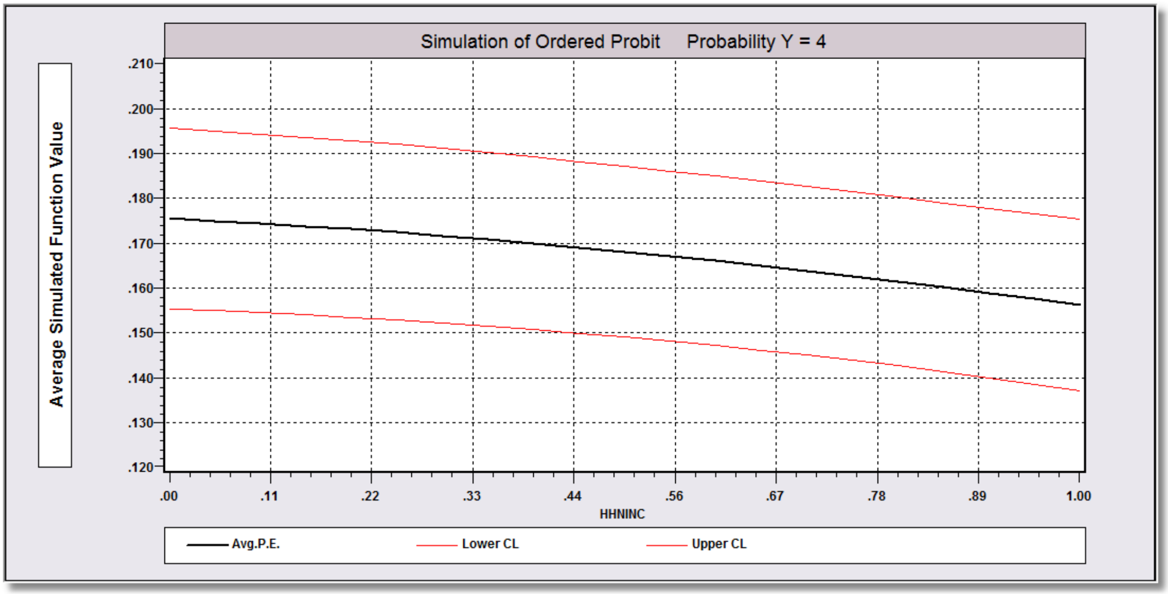


Figure N13.1 Ordered Probit Simulation

N14: Extended Ordered Choice Models

N14.1 Introduction

The basic ordered choice model is based on the 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. 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 N13 (and Chapter E34). A variety of additional specifications and extensions are supported.

N14.2 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

$$\log L = \sum_{i=1}^{n} w_i \log \text{Prob}(\textit{observed outcome}_i)$$

where $\text{Prob[cell } j] = F(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i) - F(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i).$

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

$$\text{Prob[cell } j] = F[(\mu_j - \boldsymbol{\beta}'\mathbf{x}_i)/w_i] - F[(\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i)/w_i]$$

and

$$\log L = \sum_{i=1}^{n} \log \text{Prob}(\textit{observed outcome}_i)$$

N14.3 Multiplicative Heteroscedasticity

The model with multiplicative heteroscedasticity,

$$\text{Var}[\varepsilon_i] = [\exp(\boldsymbol{\gamma}'\mathbf{z}_i)]^2,$$

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

$$\text{Prob}(y = 1 \mid \mathbf{x}, \mathbf{z}) = F\left(\frac{\mu_j - \boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{z})}\right) - F\left(\frac{\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{z})}\right)$$

N14.3.1 Testing for Heteroscedasticity

The model with homoscedastic disturbances is nested in this model (**γ** = **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 γ 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
```

N14.3.2 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

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
```

N14.4 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 N12.4.

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 N12.6.

N14.4.1 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 *NLOGIT*'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

$$\boldsymbol{\Theta} = [\beta_1,...,\beta_K,\alpha_1,...,\alpha_L,\mu_1,...,\mu_{J-1},\rho].$$

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.

N14.4.2 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.

N14.4.3 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
--------+--------------------------------------------------------------------
  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.
-----------------------------------------------------------------------------
```

N14.5 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 *NLOGIT* 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.

$$
\begin{aligned}
\text{Form 1:}\ \mu_j &= \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z}) \\
\text{Form 2:}\ \mu_j &= \exp(\theta_j + \boldsymbol{\delta}_j'\mathbf{z})
\end{aligned}
$$

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 N14.3.*

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
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.
-----------------------------------------------------------------------------
```

N14.6 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

$$d^* = \boldsymbol{\alpha}'\mathbf{w} + u, \quad d = 1\,(d^* > 0)$$

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \quad y = 0 \text{ if } y^* \le 0 \text{ or } d = 0 \leftarrow$$

$$1 \text{ if } 0 < y^* \le \mu_1 \text{ and } d = 1,$$

$$2 \text{ if } \mu_1 < y^* \le \mu_2 \text{ and } d = 1,$$

and so on.

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

$$\text{Form 1: } \mu_j = \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z})$$

$$\text{Form 2: } \mu_j = \exp(\theta_j + \boldsymbol{\delta}_j'\mathbf{z})$$

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.
-----------------------------------------------------------------------------
```

N14.7 Bivariate Ordered Probit and Polychoric Correlation

The bivariate ordered probit model is analogous to the SUR model for the ordered probit case:

$$y_{ji}^* = \boldsymbol{\beta}_j'\mathbf{x}_{ji} + \varepsilon_{ji}$$

$$y_{ji} = 0 \text{ if } y_{ji}^* \leq 0,$$

$$1 \text{ if } 0 < y_{ji}^* < \mu_1,$$

$$2, \ldots \text{ and so on}, j = 1,2,$$

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. *NLOGIT*'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.

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 *NLOGIT*'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.
-----------------------------------------------------------------------------
```

N15: Panel Data Models for Ordered Choice

N15.1 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:

$$
\begin{aligned}
y_i &= 0 \text{ if } y_i \le \mu_0,\\
&= 1 \text{ if } \mu_0 < y_i \le \mu_1,\\
&= 2 \text{ if } \mu_1 < y_i \le \mu_2,\\
&\dots\\
&= J \text{ if } y_i > \mu_{J-1}.
\end{aligned}
$$

The latent 'preference' variable, y_i^* is not observed. The observed counterpart to y_i^* is y_i. 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

$$\text{Prob}[y_i = j] = \text{Prob}[y_i^* \text{ is in the } j\text{th range}].$$

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 *NLOGIT*, fixed effects, random effects, random parameters, and latent class.

N15.2 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 = Logit] $
ORDERED ; Lhs = dependent variable
; Rhs = independent variables
; Pds = fixed number of periods or count variable
; Fixed Effects
[; Model = Logit] $

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 *NLOGIT* 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 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.

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.

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
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.
-----------------------------------------------------------------------------
```

N15.3 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 $\sigma_u = \sqrt{\rho/(1-\rho)}$. 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.

N15.3.1 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.)

N15.3.2 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 $\boldsymbol{\beta}$
varb = asymptotic covariance matrix for estimate of $\boldsymbol{\beta}$.

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 $\boldsymbol{\mu}$ and σ_u be included in *b* and the additional rows and columns be included in *varb*. The *Last Model* is [*b_variable*,*ru*]. 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 $\rho = 0$, 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.

N15.3.3 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
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.
-----------------------------------------------------------------------------
```

N15.4 Random Parameters and Random Thresholds Ordered Choice Models

The structure of the random parameters model is based on the conditional probability

$$\text{Prob}[y_{it} = j \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i] = F(j, \boldsymbol{\mu}, \boldsymbol{\beta}_i'\mathbf{x}_{it} + \alpha_i),\ i = 1,...,N,\ t = 1,...,T_i.$$

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

$$\text{E}[\boldsymbol{\beta}_i \mid \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i,$$

(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. 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

$$\mu_{ij} = \mu_{i,j\text{-}1} + \exp(\alpha_j + \boldsymbol{\delta}'\mathbf{w}_i + \theta u_{ij}),\ \mu_0 = 0,\ u_{ij} \sim N[0,1]$$

$$\varepsilon_{it} \sim N[0,\sigma_i^2],\ \sigma_i = \exp(\boldsymbol{\gamma}'\mathbf{f}_i + \tau h_i),\ h_i \sim N[0,1]$$

N15.4.1 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 = Logit ] $
ORDERED     ; Lhs  = dependent variable
            ; Rhs  = independent variables
            ; Pds  = fixed periods or count variable
            ; RPM
            ; Fcn  = random parameters specification
            [ ; Model = Logit ] $
```

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, $v_{kit} = \rho_k v_{ki,t-1} + w_{kit}$.

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. Some 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 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 *NLOGIT* 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 *NLOGIT*.

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. Distributions can be specified with

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)
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, op. cit., 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 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.

N15.4.2 Results

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:	Prob($y_{it} = J|\mathbf{x}_{it}$) = Probability of the highest cell. May be changed with **; Outcome = j** or **; Outcome = ***.	

N15.4.3 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[β_{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
```

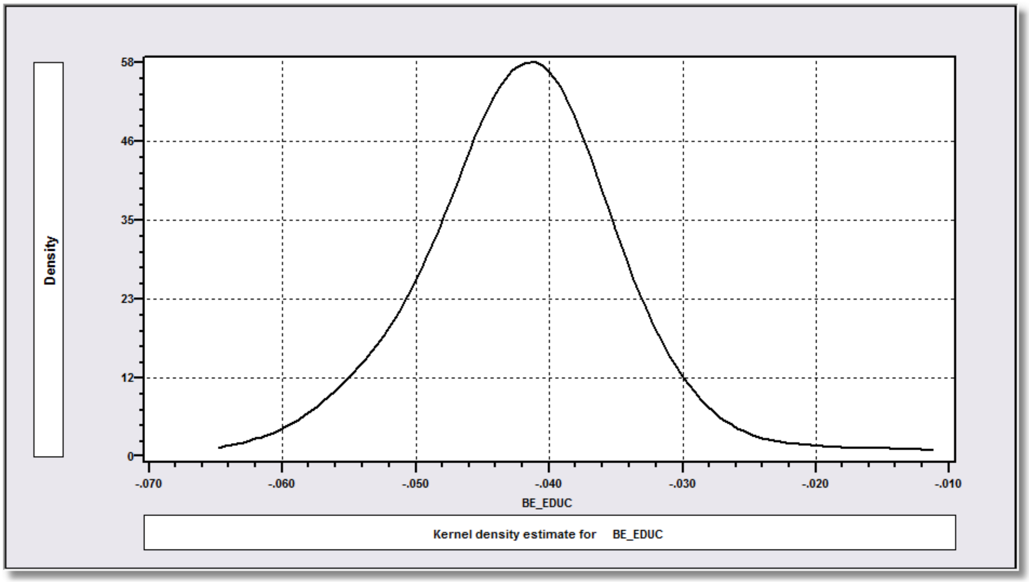


Figure N15.1 Estimators of E[β(educ)|y,x]

```
(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.
-----------------------------------------------------------------------------
```

N15.4.4 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

$$\varepsilon_{it} \sim N[0,\sigma_i^2],\quad \sigma_i = \exp[\boldsymbol{\gamma}'\mathbf{z}_i + \tau v_i] \text{ where } v_i \sim N[0,1].$$

Finally, the thresholds are formed as shown for the cross section variant of this model.

$\mu_{ij} = \mu_{i,j-1} + \exp(\alpha_j + \boldsymbol{\delta}'\mathbf{w}_i + \theta_j u_{ij}]$, where $u_{ij} \sim N[0,1]$

$\mu_0 = 0$ and $\mathbf{x}_{it}$ contains a constant term.

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 **Γ**, 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	

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
```

N15.5 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

$$\theta_{ij} \;=\; \boldsymbol{\theta}_j'\mathbf{z}_i,\ \boldsymbol{\theta}_J \;=\; \mathbf{0}.$$

N15.5.1 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.

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 $
```

N15.5.2 Results

Results saved by this estimator are

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 $J\times(K$+#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 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

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     ; Quiet ; 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     ; Quiet ; 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     ; Quiet ; 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
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 $
```

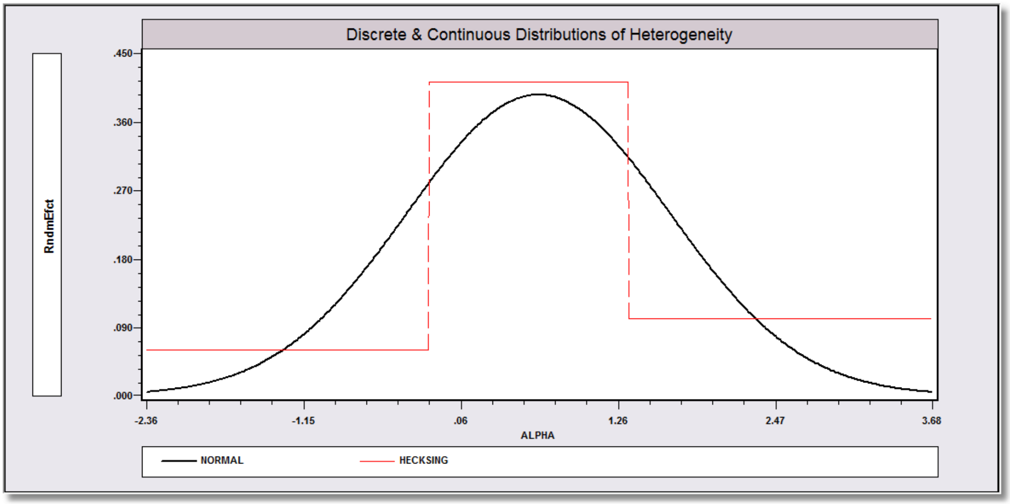


Figure E15.2 Discrete and Continuous Distributions of Heterogeneity

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
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. 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
```

N16: The Multinomial Logit Model

N16.1 Introduction

This chapter and Chapter N17 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.

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 *NLOGIT* and *LIMDEP*. *LIMDEP* contains two basic forms of the model. The *NLOGIT* program provides the major extensions that are documented in this and the remaining chapters 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. Essential features of the model and commands are documented here. This form of the multinomial logit model is supported in *LIMDEP* as well as *NLOGIT*. Further details appear in Chapter E37.

Chapter N17 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. (We emphasize, these naming distinctions are meaningless in the modeling framework – we just use them here only to organize the applicable parts of *LIMDEP* and *NLOGIT*. In practice, all of the models considered in this chapter and Chapter N17 are multinomial logit models. The basic CLOGIT model is also supported by *LIMDEP* and discussed in Chapter E38.

N16.2 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 N17. 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 β_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 500 choices by using CLOGIT, as discussed in Chapter N17.

N16.3 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.

N16.3.1 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) $
MLOGIT        ; 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.

N16.3.2 Starting Values

The parameter vector for this model is a $J \times K$ column vector,

$$\boldsymbol{\Theta} \;=\; [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_J']' .$$

You may provide starting values with **; Start = list**.

N16.4 Robust Covariance Matrix

You can 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

; Robust.

The estimator of the asymptotic covariance matrix produced with this request is the standard 'sandwich' estimator,

$$\mathbf{V} \;=\; [-\mathbf{H}]^{-1} \, (\mathbf{G}'\mathbf{G}) \, [-\mathbf{H}]^{-1}$$

where **H** is the estimated second derivatives matrix of the log likelihood and **G** is the matrix with rows equal to the first derivatives, usually labeled the OPG or 'outer product of gradients' estimator.

N16.5 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 *c*th 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*-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.

N16.6 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 ; ... $
```

A 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

$$\partial^2 \log L/\partial \boldsymbol{\beta}_l \partial \boldsymbol{\beta}_m' \quad = \quad \Sigma_t\, w_t\{-[\mathbf{1}(l=m)P_l - P_l P_m\,]\}\mathbf{X}'\mathbf{X}.$$

$$\partial \log L/\partial \boldsymbol{\beta}_{\mathrm{j}} \quad = \quad \Sigma_t\, w_i(\mathrm{d}_{tj} - t_{ij})\mathbf{x}_t \text{ where } d_{tj} = 1 \text{ if person } t \text{ makes choice } j;$$

$$\mathbf{BHHH}\text{(in blocks)} = \Sigma_t\, w_i(\mathrm{d}_{tl} - P_{tl})(\mathrm{d}_{tm} - P_{tm})\mathbf{x}_t\mathbf{x}_t'$$

and w_t = population frequency for choice made by individual t divided by sample proportion for choice made by individual t.

N16.7 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 *NLOGIT* 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 $
MLOGIT    ; 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,

$$\log L_0 \quad = \quad \Sigma_j\, n_j \log P_j$$

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.

N16.8 Partial Effects

The partial effects in this model are

$$\boldsymbol{\delta}_j = \partial P_j/\partial \mathbf{x}, \quad j = 0,1,...,J.$$

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

$$\mathbf{1}(j = l) = 1 \text{ if } j = l, \text{ and } 0 \text{ otherwise.}$$

N16.8.1 Computation of Partial Effects with the Model

This full set of results is produced automatically when your **LOGIT** command includes

; Partial Effects.

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 and elasticities 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 N16.1 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.

N16.8.2 Partial Effects Using the PARTIALS EFFECTS Command

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*. The **PARTIALS** command will also report elasticities with respect to continuous variables such as *hhnnc* by enclosing the name in brackets, such as

PARTIALS ; Effects: <hhninc> $

N16.9 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, *NLOGIT* 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.

N16.10 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 *NLOGIT*'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

V =	$-1/\sqrt{N}$, $+1/\sqrt{N}$	for $H = 2$
=	$-1/\sqrt{N}$, 0, $+1/\sqrt{N}$	for $H = 3$
=	$-1/\sqrt{N}$, $-.5/\sqrt{N}$, [0], $+.5/\sqrt{N}$, $+1/\sqrt{N}$	for $H = 4$ or 5
=	... [0], $+.33/\sqrt{N}$, $+.67/\sqrt{N}$, $+1/\sqrt{N}$	for $H = 6$ or 7
=	... [0], $+.25/\sqrt{N}$, $+.50/\sqrt{N}$, $+.75/\sqrt{N}$, $+1/\sqrt{N}$	for $H = 8$ or 9

(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.

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)
```

HSAT	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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.                 |
+----------------------------------------------------------------------+
```

N16.11 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

$$\partial \log L/\partial \boldsymbol{\beta}_{\mathrm{j}} = \Sigma_t (\mathrm{d}_{tj} - P_{tj})\mathbf{x}_t.$$

$$\partial^2 \log L/\partial \boldsymbol{\beta}_l \partial \boldsymbol{\beta}_m' = \Sigma_t -[\mathbf{1}(l=m)P_{tl} - P_{tl}P_{tm}]\mathbf{x}_t\mathbf{x}_t'.$$

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 *NLOGIT*, 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.

N16.12 Sequential Logit Model

The sequential logit model treats the outcome of a multinomial choice as a sequence of transitions. The outcomes are ordered by some construction (which might suggest an ordered choice model as a natural formulation.) Consider the outcome, level of schooling: less than high school (y = 0), high school (y =1), college (y = 2) and advanced degree (y = 3). The sequential logit model treats the level as a sequence of transitions starting with transition from level 0, which everyone enters so $P_0 = 1$. The probability of stopping at level 0 is the probability of entering and not passing level 1, which is $P_0(1 - P_1)$. The probability of stopping at level 1 (high school) is the probability of leaving level 1 and not leaving level 2, $P_0P_1(1 - P_2)$. For level 2, the probability is, $P_0P_1P_2(1 - P_3)$ and finally for level 3, $P_0P_1P_2P_3$ This is treated as a sequence of conditional binary logit models;

Prob(observed level = j) = Probability(exit levels m = 1,…,j) × Probability(no exit at level j+1).

The outcome variable must be coded 0,1,… as for other forms of the multinomial logit model. By this formulation, for the outcomes listed above,

Prob(observed level = j) = $P_{0,1,\ldots,j}(1 - P_{J+1})$.

The command for the sequential logit model is

SEQLOGIT ; Lhs = outcome variable
; Rhs = exogenous influences $

The standard options for nonlinear models, including **; Cluster = specification**, are available for this model. The outcomes are labeled ‘name = 0’, ‘name = 1,’ and so on in the estimation results. You may provide a set of labels with

; Choices = a list of labels for the outcomes.

For purposes of partial effects and simulations of the outcome variable, the default function is an expected value

$$E[y] = 0\times\text{Prob}(y = 0) + 1\times\text{Prob}(y = 1) + \ldots$$

In a particular application, the outcomes might represent quantifiable levels, such as years of education. For this case, you may supply a set of levels to be used instead of (0,1,…,J) with

; Levels = list of values.

The following example is based on the health care data used in the previous example.

CREATE ; edlevel = (educ > 10) + (educ > 12) + (educ > 16) $
SEQLOGIT ; Lhs = edlevel ; Rhs = one,income,married
; Choices = lths,hs,college,graduate
; Levels = 10,12,16,18 $
SIMULATE $

(We have used a rather arbitrary coding of the years of education variable for purposes of this numerical example.)

```
-----------------------------------------------------------------------------
Sequential Multinomial Logit Model
Dependent variable                EDLEVEL
Log likelihood function    -27123.76693
Restricted log likelihood  -28275.56899
Chi squared [  6](P= .000)   2303.60411
Significance level               .00000
McFadden Pseudo R-squared      .0407349
Estimation based on N =  27326, K =   9
Inf.Cr.AIC  =  54265.5 AIC/N =    1.986
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
 EDLEVEL|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Transition from LTHS     to HS       etc...........................
Constant|     .48944***      .04215    11.61  .0000      .40684    .57204
  INCOME|    3.11877***      .11521    27.07  .0000     2.89296   3.34458
 MARRIED|    -.29636***      .03638    -8.15  .0000     -.36766   -.22505
        |Transition from HS       to COLLEGE  etc...........................
Constant|   -1.65743***      .04400   -37.67  .0000    -1.74367  -1.57119
  INCOME|    2.62319***      .09537    27.51  .0000     2.43627   2.81011
 MARRIED|    -.94624***      .03830   -24.71  .0000    -1.02131   -.87118
        |Transition from COLLEGE  to GRADUATE etc...........................
Constant|   -1.31118***      .08099   -16.19  .0000    -1.46992  -1.15244
  INCOME|    2.23477***      .17356    12.88  .0000     1.89460   2.57494
 MARRIED|     .03204         .06843      .47  .6397     -.10209    .16616
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
---------------------------------------------------------------------
Model Simulation Analysis for E[Outcome] in Sequential Logit Model
---------------------------------------------------------------------
Simulations are computed by average over sample observations
---------------------------------------------------------------------
User Function      Function   Standard
(Delta method)      Value      Error     |t|  95% Confidence Interval
---------------------------------------------------------------------
Avrg. Function    15.03525     .02874  523.12    14.97892    15.09158
```

N16.13 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.

N16.13.1 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. 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 *NLOGIT*'s other random parameter models, which is how we do the estimation, by maximum simulated likelihood. There are two additional versions of the essential structure:

1. Independent effects: $\mathbf{\Gamma}$ = A diagonal matrix.

2. True random effects: $\mathbf{\Gamma}$ = 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;

```
MLOGIT    ; Lhs  = dependent variable
          ; Rhs  = list of variables including one $
MLOGIT    ; Lhs  = dependent variable
          ; Rhs  = list of variables including one
          ; RPM ; Fcn = remnl
          [; Halton]
          [; Pts  = ...]
          ; Pds  = panel specification  $
```

(In earlier versions of NLOGIT, the **; Fcn = remnl** would have been **; Fcn = one(n)** instead. You may still use this syntax.) 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.

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 $
SETPANEL   ; Group = it ; Pds = ti $
MLOGIT     ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids $
MLOGIT     ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids
           ; RPM ; Fcn = remnl ; Common ; Halton ; Pts = 50 ; Panel $
MLOGIT     ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids ; Quiet $
MLOGIT     ; Lhs  = hsat; Rhs = one,age,hhninc,hhkids
           ; RPM ; Fcn = remnl ; Halton ; Pts = 50 ; Panel $
MLOGIT     ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids ; Quiet $
MLOGIT     ; Lhs = hsat ; Rhs = one,age,hhninc,hhkids
           ; RPM ; Fcn = remnl ; Correlated ; Halton ; Pts = 50 ; Panel $
```

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     |
| 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
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*          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
```

N16.13.2 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

$$P_{ijt} \mid \alpha_{i1},...,\alpha_{iJ} = \frac{\exp(\boldsymbol{\beta}'_j\mathbf{x}_{it} + \boldsymbol{\gamma}'_j\mathbf{z}_{it} + \alpha_{ij})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}'_j\mathbf{x}_{it} + \boldsymbol{\gamma}'_j\mathbf{z}_{it} + \alpha_{ij})} \quad t = 1,...,T_i,\ j = 0,1,...,J,\ i = 1,...,N$$

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. The core of the treatment is to treat the first period as an equilibrium, with no lagged effects,

$$P_{ij0} \mid \theta_{i1},...,\theta_{iJ} = \frac{\exp(\boldsymbol{\delta}'_j\mathbf{x}_{i0} + \theta_{ij})}{\sum_{j=1}^{J}\exp(\boldsymbol{\delta}'_j\mathbf{x}_{i0} + \theta_{ij})},\ t = 0,\ j = 0,1,...,J,\ i = 1,...,N$$

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, $(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_J,\boldsymbol{\gamma}_1,...,\boldsymbol{\gamma}_J,\boldsymbol{\delta}_1,...,\boldsymbol{\delta}_J,\boldsymbol{\Gamma},\Phi)$ 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 $1,2,...,T_i$ 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)
```

N17: Conditional Logit Model

N17.1 Introduction

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.

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 the basic model specification and estimation. Other features of the model, including those extensions contained in *LIMDEP* and *NLOGIT* are described in Chapters N18-N22.

N17.2 The Conditional Logit Model – CLOGIT

In the multinomial logit model described in Chapter N16, 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:

$$U(\text{choice } j \text{ for individual } i) \;=\; U_{ij} \;=\; \boldsymbol{\beta}'\mathbf{x}_{ij} \;+\; \boldsymbol{\gamma}'\mathbf{z}_i \;+\; \varepsilon_{ij},\; j = 1,\ldots,J.$$

The random, individual specific terms, $(\varepsilon_{i1},\varepsilon_{i2},\ldots,\varepsilon_{iJ})$ 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 N16.2.

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* and *NLOGIT*, 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.

N17.3 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* and *NLOGIT* These data provide a compact illustration of how data should be arranged for the CLOGIT model. 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, 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**	

N17.3.1 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 **IMPORT** or **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 – that is, after the data are read into the program.

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* and *NLOGIT* 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.

See Chapters N18 and N20 for further details on choice sets and data types also fixed and variable number of choices and restricting the choice set during estimation.

N17.4 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 commands **DISCRETE CHOICE** and **NLOGIT** in this form 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.

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.

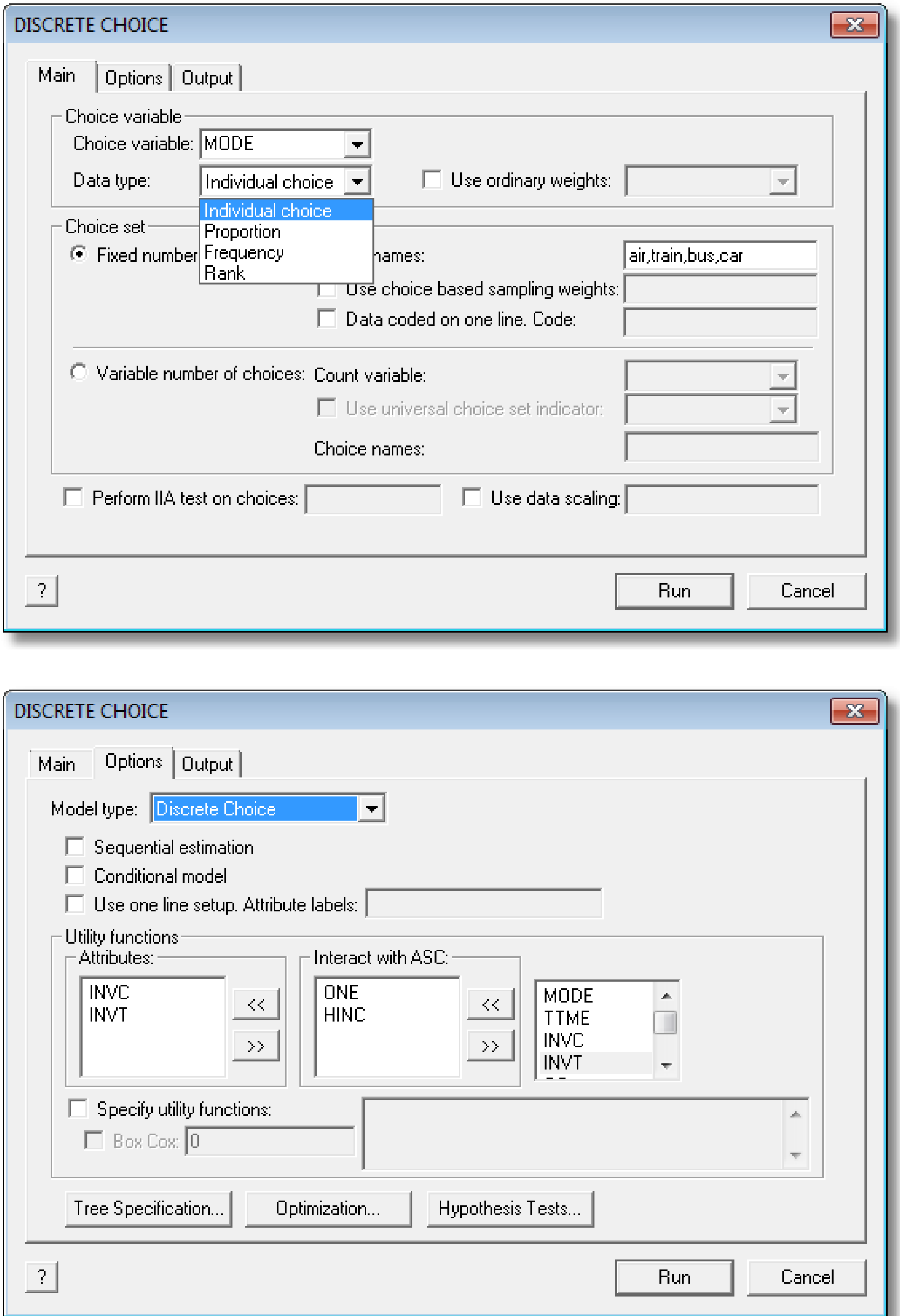


Figure N17.1 Command Builder for the Conditional Logit Model

N17.5 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 N23.6. 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 $1 - K(\log L\,/\text{Log}L_0)$

$$Adjusted\ R^2 \;=\; 1 \;-\; 1.02632\,(-62.54818/-81.0939) \;=\; 0.20794.$$

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 Chapters N21 and N22.

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.
-----------------------------------------------------------------------------
```

N17.5.1 Robust Standard Errors

The 'cluster' estimator 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* and *NLOGIT* – the same construction is used here.

N17.5.2 Descriptive Statistics

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
```

N17.6 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

$$P_{ji} = \text{Prob}[y_i = j] = \text{Prob}[d_{ji} = 1] = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ji})}{\sum_{m=1}^{J_i}\exp(\boldsymbol{\beta}'\mathbf{x}_{mi})}$$

$$\text{Log } L = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ji}\log P_{ji}$$

$$\bar{\mathbf{x}}_i = \sum_{j=1}^{J_i} P_{ji}\mathbf{x}_{ji},$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ji}(\mathbf{x}_{ji} - \bar{\mathbf{x}}_i),$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} P_{ji}(\mathbf{x}_{ji} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ji} - \bar{\mathbf{x}}_i)',$$

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.
-----------------------------------------------------------------------------
```

N17.7 Generalized Maximum Entropy Estimator

The CLOGIT multinomial logit model may be estimated using the generalized maximum entropy estimator described in Section N16.10 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
--------+--------------------------------------------------------------------
```

N17.8 MLOGIT and CLOGIT

When there are no choice varying attributes, CLOGIT is the same model as MLOGIT. From Chapter N16, the functional form for MLOGIT is

$$\text{Prob}(y_i = j|\mathbf{x}_i) \quad = \quad \frac{\exp(\boldsymbol{\beta}'_j \mathbf{x}_i)}{\sum_{m=1}^{J} \exp(\boldsymbol{\beta}'_m \mathbf{x}_i)}, \; j = 0,...,J,$$

From the introduction in this chapter,

$$\text{Prob}(\text{choice} = j \mid \mathbf{X}_i, \mathbf{z}_i) \quad = \quad \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{ji} + \boldsymbol{\gamma}'_j \mathbf{z}_i)}{\sum_{m=0}^{J} \exp(\boldsymbol{\beta}' \mathbf{x}_{mi} + \boldsymbol{\gamma}'_m \mathbf{z}_i)}, \; j = 1,...,J.$$

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 |
--------+---------+---------+---------+---------+
```

N18: Data Setup for *NLOGIT*

N18.1 Introduction

In general, the data for the models described in Chapters N23-N33 will be arranged in a format that is set up to work well with the specific *NLOGIT* estimators. In almost all cases, the data used for all models that you fit with *NLOGIT* will be set up as if they were a panel. That is, each individual choice situation will have a set of observations, with one 'line' of data for each choice in the choice set. Thus, in the analogy to a panel, the 'group' is a person and the group size would be the number of choices. You will use this arrangement in nearly all cases. This chapter will explain the various aspects of setting up the data for the *NLOGIT* models. We note one specific feature of the data set that is unusual is the 'ignored value code,' -888, described in Section N18.9. This special code is used to signal values that are deliberately omitted from the data set by the observed individual – they are 'missing values,' with a specific understanding for why they are missing.

N18.2 Basic Data Setup for *NLOGIT*

In the base case, the 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:

		YQW		
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 the arrow marks the row of the respondent's actual choice.

When you read these data, the data set is not treated any differently from any other panel. *Nobs* would be the total number of rows in the data set, in the hypothetical case, 75, not 25. The separation of the data set into the above groupings would be done at the time your particular model is estimated.

NOTE: Missing values are handled automatically by estimation programs in *NLOGIT*. You should not reset the sample or use **SKIP** with the *NLOGIT* models. Observations that have missing values are bypassed as a group.

Thus far, it is assumed that the observed outcome is an indicator of which choice was made among a fixed set of up to 100 choices. Numerous variations on this are possible:

- Data on the observed outcome may be in the form of frequencies, market shares, or ranks. These possibilities are discussed further in Section N18.3.

- The number of choices may differ across observations. This is discussed further in Section N20.2.

The preceding described the base case model for a fixed number of choices using individual level data. There are several alternative formulations that might apply to the data set you are using.

N18.3 Types of Data on the Choice Variable

We allow several types of data on the choice variable, *y*. If you have *grouped data*, the values of *y* will be *proportions* or *frequencies*, instead of *individual choices*. In the first case, within each observation (*J* data points), the values of *y* will sum to one when summed *down* the *J* rows. (This will be the only difference in the grouped data treatment.) In the second case, *y* 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. Or, the data might be counts of responses to particular questions in a survey in which groups of people in different locations or at different times were surveyed. Finally, *y* might be a set of *ranks*, in which case, instead of zeros and ones, *y* would take values 1,2,...,*J* (not necessarily in that order) within, and reading down, each block.

More specifically, data on the dependent (Lhs) variable may come in these four forms:

- **Individual Data:** The Lhs variable consists of zeros and a single one which indicates the choice that the individual made. When data are individual, the observations on the Lhs variable will sum exactly to 1.0 for every person in the sample. A sum of 0.0 or some other value will only arise if a data error has occurred. Individual choice data may also be simulated. See Section N18.3.1 below.

- **Proportions Data:** The Lhs variable consists of a set of sample proportions. Values range from zero to one, and again, 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 individuals.

- **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 Section N21.4.1 is not available.
    - The number of choices may be fixed or variable, as described above.
    - You may keep probabilities or inclusive values as described in Chapter N21.
    - Ranks data may only be used with the conditional logit model (CLOGIT) and the mixed logit (random parameters) model (RPLOGIT).

- **Best/Worst Data:** This would be a variant of ranks data. When data are in the form of best and worst, there will be three values of the outcome variable. The choice variable for the best (most preferred) outcome is coded 1 as usual. The least favored outcome is coded with any value larger than 1, such as 2, 9, or any other value. Outcomes between these that are not chosen either best or worst will be coded 0 as usual.

The first three data types are detected automatically by *NLOGIT*. You 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. (See below for one exception.) The ranks data are an exception for which you would use

NLOGIT ; ... as usual ...; Ranks $

If you are using frequency or proportions data, and your data contain zeros or ones, 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 **NLOGIT** 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.)

Best/worst data can come in three forms. The simple case is that in which the chooser simultaneously identifies the most and least favored alternatives. For this case, use

; Best worst.

If the chooser identifies the best alternative and then chooses the worst among those that remain, then this choice is sequential. Indicate this with

; Best worst ; Seq = bw.

Lastly, if the worst alternative is indicated first and the best is chosen from among the remainder, use

; Best worst ; Seq = wb.

The actual data will look the same for the three cases. The difference that is implied relates to the way the likelihood is formulated for estimation. We note, in practical terms, the three are likely to produce similar, if not indistinguishable results. This makes sense. Most of the information about preferences is provided by the most favored alternative. This will be the same in all cases. The change from simultaneous choice to sequential is likely only to lead to marginal changes in the model results.

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 at any ranks other than the lowest.

N18.3.1 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*).

N18.3.2 Simulated Choice Data

For some kinds of experiments and simulations, you might want to draw a random sample of choices given known utility functions. *NLOGIT* allows simulation of the Lhs variable in a choice model using

$$Y = j^* \text{ from } \mathrm{Max}(U_{ij}),$$

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 **NLOGIT** or **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. For example, suppose you wanted to compare the results of two different nesting specifications using the simulated data. The utilities are in variable *utility*.

The command set might appear as follows:

```
CALC      ; Ran(56791) $
NLOGIT    ; Lhs = utility ; Choices = air,train,bus,car
          ; Tree = (air,train,bus),(car)
          ; ... $
CALC      ; Ran(56791) $
NLOGIT    ; Lhs = utility ; Choices = air,train,bus,car
          ; Tree = (train,bus),(air,car)
          ; ... $
```

N18.3.3 Checking Data Validity

NLOGIT 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
```

N18.4 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.)

N18.5 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), and it is the only estimator in *NLOGIT* 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 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
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         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.

```
--------+--------------------------------------------------------------------
        |                  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
--------+--------------------------------------------------------------------
```

N18.6 Entering Data on a Single Line

Data for *NLOGIT* 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

$$\begin{array}{r} \\ \text{Air} \\ \text{Train} \\ \text{Bus} \\ \text{Car} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Choice} & \text{Cost} & \text{Time} & \text{Income} \end{array} \\ \left[\begin{array}{cccc} y_{air} & cost_{air} & time_{air} & income \\ y_{train} & cost_{train} & time_{train} & income \\ y_{bus} & cost_{bus} & time_{bus} & income \\ y_{car} & cost_{car} & time_{car} & income \end{array}\right] \end{array}$$

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:

$$\begin{array}{c} \text{choice} \quad \text{cost time} \quad \text{constants} \qquad\qquad \text{income} \end{array}$$

$$\mathbf{X}_i = \begin{bmatrix} y_{air} & c_a & t_a & 1 & 0 & 0 & income & 0 & 0 \\ y_{train} & c_t & t_t & 0 & 1 & 0 & 0 & income & 0 \\ y_{bus} & c_b & t_b & 0 & 0 & 1 & 0 & 0 & income \\ y_{car} & c_c & t_c & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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

The following options are unavailable when data are arranged on a single line:

- Data scaling: See Section N18.10.
- Ranks data: See Section N18.3.
- Keeping predictions, probabilities, inclusive values, etc. See the relevant parts of Chapter N21.
- **Model: U(...) = spec...:** You must use **; Rhs** and/or **; Rh2**. See Chapter N19.

N18.7 Converting One Line Data Sets for *NLOGIT*

Data for the several discrete choice models in *NLOGIT* are assumed to be arranged in a 'stack' for each observation. For example, suppose you are studying mode choice for transportation (of course), and your observation consists of the following (as in the preceding example):

- The choice variable, *choice* = 1, 2, or 3 for *car, train, bus,*
- For each mode, *time, cost* – note that this differs by choice,
- For the individual, *age, income* – note that this does not differ by choice.

NLOGIT would usually expect each observation in the sample to consist of three rows, such as the following

$$\begin{array}{l} \\ \text{car} \\ \text{train} \\ \text{bus} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{choice} & \text{time} & \text{cost} & \text{age} & \text{income} \end{array} \\ \begin{bmatrix} 0 & 44 & 125 & 37 & 56.5 \\ 1 & 29 & 40 & 37 & 56.5 \\ 0 & 56 & 25 & 37 & 56.6 \end{bmatrix} \end{array}.$$

Suppose that your data were arranged not in this fashion, but in a single observation, as in

$$\begin{bmatrix} choicei & ctime & ttime & btime & ccost & tcost & bcost & agei & incomei \\ 2 & 44 & 29 & 56 & 125 & 40 & 25 & 37 & 56.6 \end{bmatrix}.$$

The estimator in *NLOGIT* can handle either arrangement, but for several purposes it will usually be more convenient to use the first. You can convert this one line observation to the three record format in order to use *NLOGIT*'s estimation programs. There are two ways to do so. *NLOGIT* provides a command that does the full conversion of the data set internally for you – essentially it creates a new data set for you. The second way to convert the data set is to write a new data file (using *NLOGIT*'s commands) containing the necessary variables, and read in the newly created data set. You could use this operation to create a data set for export as well. We note, there are relatively few commercial packages available that do the kinds of modeling that you will do with *NLOGIT* – for several of the models, *NLOGIT* is unique. As far as we are aware, other software generally use the more cumbersome single line format. You will find the operation useful when you import data from other programs into *NLOGIT*.

N18.7.1 Converting the Data Set to Multiple Line 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

```
NLCONVERT  ; 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

```
NLCONVERT  ; 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 **NLCONVERT** 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. **NLCONVERT** 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.

Note that the count of Rhs variables is an exact multiple of the number of choices in the **; Choices** list.

When **NLCONVERT** is executed, the sample is reset to the number of observations in the new sample. There is an additional option with **NLCONVERT**. 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:

$$\begin{bmatrix} choicei1 & choicei2 & ctime & ttime & btime & ccost & tcost & bcost & agei & incomei \\ 2 & 3 & 44 & 29 & 56 & 125 & 40 & 25 & 37 & 56.6 \\ 1 & 1 & 19 & 44 & 20 & 160 & 18 & 50 & 42 & 98.6 \\ 3 & 2 & 28 & 55 & 15 & 85 & 50 & 9 & 10 & 22.0 \end{bmatrix}.$$

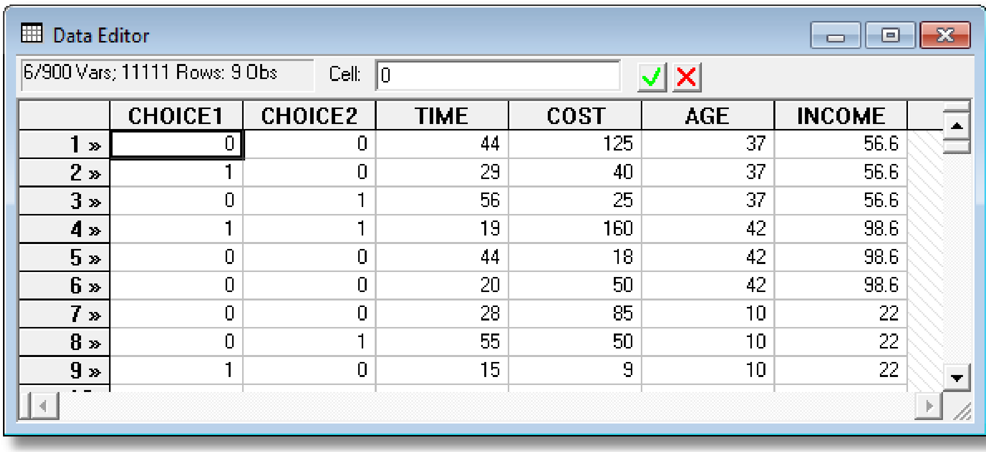
We wish to convert this data set to *NLOGIT*'s multiple line format. There are three choices in the choice set, so there will be three rows of data for each observation.

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 $
NLCONVERT ; 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 $



	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 N18.1 Converted Data Set

```
===========================================================================
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
=================================================================
```

N18.7.2 Writing a Multiple Line Data File for *NLOGIT*

If you need to create a data file in the multiple line format, you can, of course, use **NLCONVERT**, then just use **WRITE** to create the file. The following shows a way that you can bypass **NLCONVERT** if you wish. The first command creates the three choice variables (one will appear in each row of the new data set).

```
CREATE      ; car = (choice=1) ; train = (choice=2) ; bus = (choice=3) $
```

The next command writes out the 15 variables, but only allows five items to appear on each line, which is what you need to recreate the data file.

```
WRITE       ; car,   ctime, ccost, age,  income,
              train, ttime, tcost, age,  income,
              bus,   btime, bcost, age,  income
            ; File = whatever you choose
            ; Format = (exactly 5 format codes, not 15)  $  ←
```

For example, **; Format = (5F10.3)**. See Chapter R3 for discussion of using formats for reading and writing data files.

The **WRITE** command takes advantage of a very useful feature of this type of formatting. The **WRITE** command instructs *NLOGIT* to write 15 values, but it provides only five format codes. What happens is that the program will write the first five values according to the format given, *then start over in the same format, on a new line*. That is exactly what we want. This **WRITE** command writes three lines per observation. When it is done, the data can be read back into *NLOGIT* with no further processing necessary, in the format required for *NLOGIT*.

N18.8 Merging Invariant Variables into a Panel

Some panel data sets contain variables that do not vary across the observations in a group. A common example is the data shown in the preceding two sections. Some variables in the data set will be attributes of the choices, and, as such, will be different for each choice. Others may be characteristics of the individual, and will, therefore, be repeated on each record in the panel. *NLOGIT* allows you to keep separate data files for the variable and invariant data. This may result in a large amount of space saving. The data may be merged when they are read into *NLOGIT*, rather than in the data set. For example, consider a panel with three individuals, and a variable number of observations per individual, two, then three, then two. The two data sets might look like

```
     File=var.dat
     Variable data
xyniz
ind=1   1.1   4   2
        1.2   2   2
ind=2   3.7   8   3
        4.9   3   3
        5.0   1   3
ind=3   0.1   2   2
        1.2   5   2
```

```
File=invar.dat
Invariant data

ind=1   100.7
ind=2    93.6
ind=3    88.2
```

Note the usual count variable for handling panels. To merge these files, use this setup

READ ; File = var.dat ; Nobs = 7 ; Nvar = 3 ; Names = x,y,ni $

This reads the original panel data set. Now, to expand the invariant data, the syntax is

READ ; File = invar.dat ; Nobs = 3 ; Nvar = 1 ; Names = z ; Group = ni $

The new feature is the **; Group = ...** specification. **; Group** specifies either a count variable, as above, or a fixed group size, as usual for *NLOGIT*'s handling of panel data sets. The resulting data will be

	x	*y*	*ni*	*z*
ind=1	1.1	4	2	100.7
	1.2	2	2	100.7
ind=2	3.7	8	3	93.6
	4.9	3	3	93.6
	5.0	1	3	93.6
ind=3	0.1	2	2	88.2
	1.2	5	2	88.2

Note the following checks and errors:

- *Nobs* must be given on the second **READ** command.
- *Nobs* must match exactly the number of groups in the existing data set.
- The existing panel must be properly blocked out by the **; Groups** variable or by a constant group size.
- This form may not be used with spreadsheet files.
- This form may not be used to read data **; By Variables**.
- This form may not be used with the **APPEND** command.
- The first data set could be read with a simple **IMPORT ; File = var.dat $** command, however, the second requires a fully specified **READ** command because of the merging feature.

N18.9 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 to remove 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 an extreme value on the choice process. Hensher, Rose, and Greene (2005) 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. *NLOGIT* allows you to automate precisely this formulation in all discrete choice models with a special data coding.

> *For respondents which 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 autodetects this feature and adjusts the model accordingly. You do not have to add any other codes to any *NLOGIT* commands to signal this aspect of the data. The model output will contain a diagnostic box noting when this option is being used when *NLOGIT* finds these values in the data. Some aspects of this convention are:

- At least some respondents must actually consider the attribute. It cannot be omitted from the model for everyone.

- In the multinomial, multiperiod probit model, if an attribute is ever ignored, it must be ignored in all periods. This is not the case for LCM or RPL which use repeated choice situation data. A respondent may ignore attributes in some choice situations (say the later ones in an experiment) and not in others (say the early ones).

- In nested logit models, this feature can only be used at the lowest, twig level of the tree. It will not be picked up if it used at branch or higher levels. For example, in nested logit models, one often puts the demographic data in the model at the branch level. This feature will not be picked up in branch level variables.

- In computing elasticities, 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.

- In computing descriptive statistics with the **; Describe** option, 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.

- In models that can produce person specific parameters (mixed logit, latent class), the saved parameters for the individual will contain the requested zeros if the indicated attribute is noted as not used.

N18.10 Scaling the Data

In some applications involving stated preference data, it is useful to estimate the model with different scales of the same data. That is, if all of the data on all attributes are collected in a matrix, **X**, then we estimate the discrete choice model with the data set

$$\mathbf{X}^* = \theta\mathbf{X},$$

for different values (near 1.0) of the scalar θ. There are two ways to do this. Suppose the attributes in **X** are named *x*1, *x*2, ..., *xk*. To set up the procedure, we create a placeholder for **X***:

```
CREATE      ; x1s = x1 ; x2s = x2 ; ... $
```

Now, define the matrices:

```
NAMELIST    ; x = x1, x2,..., xk
            ; xs = x1s, x2s, ... , xks $
```

Finally, define a procedure which sets up the **NLOGIT** estimation in terms of the variables in *xs* instead of *x*, along with a **MATRIX** command that does the scaling:

```
PROCEDURE $
  CREATE    ; xs = x $
  MATRIX    ; xs = Xmlt(theta) $
  NLOGIT    ; ... $
ENDPROCEDURE $
```

Now, the model can be fit with any desired scaling of the data with the command

```
EXECUTE     ; theta = the desired value $
```

NLOGIT also provides a more fully automated procedure for scaling when you wish to change only some of the variables in a model. You can specify as part of the command

; Scale (list of variables) = θ_{low} , θ_{high} , number of points.

This requests *NLOGIT* to examine '*number of points*' equally spaced values ranging from θ_{low} to θ_{high}. The value associated with the highest value of the log likelihood is then used to reestimate the model. (No output is produced during the search.) You may also specify a second round, finer search with

; Scale (list of variables) = θ_{low} , θ_{high} , number of points , nfine.

If you specify the second round search (nfine), evenly spaced points ranging from the adjacent values below and above the value found in the first search are examined to try to improve the value of the log likelihood. For example, if you specify the grid .5,1.5,11,11, the first search will examine the values .5, .6, ..., 1.5. If the best value were found at, say, 1.2, then the finer search would examine 1.10, 1.12, .., 1.30.

N18.11 Data for the Applications

The documentation of the *NLOGIT* program in the chapters to follow includes numerous applications based on the data set clogit.dat, that is distributed with *NLOGIT*. These data are 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. Data for *NLOGIT* 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**	**0**	**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**	

N18.12 Merging Revealed Preference (RP) and Stated Preference (SP) Data Sets

For applications in which you wish to merge RP and SP data sets, we assume that a data set is built up for each individual in the sample from an RP observation and one or more SP observations, for the same person. To construct the data for the simulation, you will require two variables:

1. a numeric *identification* (*id*) that is the same for the RP and SP observations,

2. a treatment or choice set *type* index, coded 0 for the RP observation and 1,...,*T* (may vary by person) for the SP data.

It is assumed that there is exactly one RP observation and any number up to *T* SP observations. The type code need not obey any particular convention; you may code it any way you wish. What is essential is that this type code equal zero for the RP observation and some positive value for the SP observation(s). The SP observations may have the same or different values for this coding. From this information *NLOGIT* can deduce the form of the choice set.

NOTE: This feature of the simulator cannot be used if the data are already arranged as RP,SP1,RP,SP2,RP,SP3,RP… That is, the RP observation must not be repeated.

The **; Choices = list** specification in the model command must include the full universal choice list for both RP and SP. In most applications of this sort, the RP observations will use one subset and the SP observations will use the remainder and there will be no overlap. For example, the universal choice set might include a set of, say, five RP choices and 15 SP choices in which each RP choice setting involves some smaller number, say four, of the latter. However, this partitioning is not necessary. For example, you might have survey data in which variants on an existing choice set are presented to individuals, for example, as in ‘would you choose option A,B,C... if price were changed by ...?’. The additional specification for **NLOGIT** will be

; MergeSPRP (id = name of unique identifier,
type = the name of the treatment indicator variable)

where *id* is the unique identifying variable that links the SP and RP observations (or any observations associated with the same *id* from two data sets).

The effect of the preceding specification is to expand each observation into *T* combined sets of data, in the form shown above. (*NLOGIT* wants to do the expansion itself.) This does not actually modify your data set. The observations are created temporarily during the computations.

N19: *NLOGIT* Commands and Results

N19.1 Introduction

This and the next three chapters will describe the common features of the *NLOGIT* models and commands used to fit them. Section N19.2 presents the generic command structure for *NLOGIT*. The specification of models for *NLOGIT* follows the general pattern for model commands in *LIMDEP*. Section N19.3 describes optional command specifications. The different models, such as nested logit, mixed logit and multivariate probit, are requested by modifying the basic command. Section N19.4 describes output features, such as estimation results and elasticities, that are common to all the models. The subsequent chapters, N20-N22 will provide greater details on the model specifications, including choice sets and utility functions in Chapter N20 and partial effects, hypothesis tests and model simulations in Chapter N21.

NLOGIT is built around estimation of the parameters of the random utility model for discrete choice,

$$U(\text{choice } j \text{ for individual } i) = U_{ij} = \boldsymbol{\beta}_{ij}'\mathbf{x}_{ij} + \varepsilon_{ij},\ j = 1,...,J_i,$$

in which individual i makes choice j if U_{ij} is the largest among the J_i utilities in the choice set. The parameters in the model are the weights in the utility functions and the deeper parameters of the distribution of the random terms. In some cases, the 'taste' parameters in the utility functions might vary across individuals and in most cases, they will vary across choices. The latter is simple to accommodate just by merging all parameters into one grand $\boldsymbol{\beta}$ and redefining $\mathbf{x}$ with some zeros in the appropriate places. But, for the former case, we will be interested in a lower level parameterization that involves what are sometimes labeled the 'hyperparameters.' Thus, it might be the extreme case (as in the random parameters logit model) that $\boldsymbol{\beta}_{ij} = \mathbf{f}(\mathbf{z}_i, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\beta}, \mathbf{v}_i)$ where $\boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\beta}$ are lower level parameters, $\mathbf{z}_i$ is observed data, and $\mathbf{v}_i$ is a set of latent unobserved variables. The parameters of the random terms will generally be few in number, usually consisting of a small number of scaling parameters as in the heteroscedastic logit model, but they might be quite numerous, again in the random parameters model. In all cases, the main function of the routines is estimation of the structural parameters, then use of the estimated model for analysis of individual and aggregate behavior.

N19.2 *NLOGIT* Commands

The essential command for the set of discrete choice models in *NLOGIT* is the same for all, with the exception of the model name:

Model ; Lhs = variable which indicates the choice made
; Choices = a set of J names for the set of choices
(utility functions)
; Rhs = choice varying attributes in the utility functions
; Rh2 = choice invariant variables, including *one* for ASCs $
(or)
; Model: utility specifications… $

The various models are as follows, where either of the two forms given may be used:

Model	Command	Alternative Command Form
Conditional Logit	**CLOGIT**	**NLOGIT**
Random Regret Logit	**RRLOGIT**	**NLOGIT ; RRM**
Scaled Multinomial Logit	**SMNLOGIT**	**NLOGIT ; SMNL**
Error Components Logit	**ECLOGIT**	**NLOGIT ; ECM = ...**
Heteroscedastic Extreme Value	**HLOGIT**	**NLOGIT ; HET**
Nested Logit	**NLOGIT**	**NLOGIT ; Tree = ...**
Generalized Nested Logit	**GNLOGIT**	**NLOGIT ; GNL**
Random Parameters Logit	**RPLOGIT**	**NLOGIT ; RPL**
Generalized Mixed Logit	**GMCLOGIT**	**NLOGIT ; GMX**
Nonlinear Random Parameters	**NLRPLOGIT**	**NLOGIT ; NLRP =**
Latent Class Logit	**LCLOGIT**	**NLOGIT ; LCM**
Latent Class Random Parameters	**LCRPLOGIT**	**NLOGIT ; RPL ; LCM**
Multinomial Probit	**MNPROBIT**	**NLOGIT ; MNP**

The description to follow in the rest of this chapter applies equally to all models. For convenience, we will use the generic **NLOGIT** command in most of the discussion, while you can use the specific model names in your estimation commands.

The command builders for these models can be found in Model:Discrete Choice. There are several model options as shown in Figure N19.1

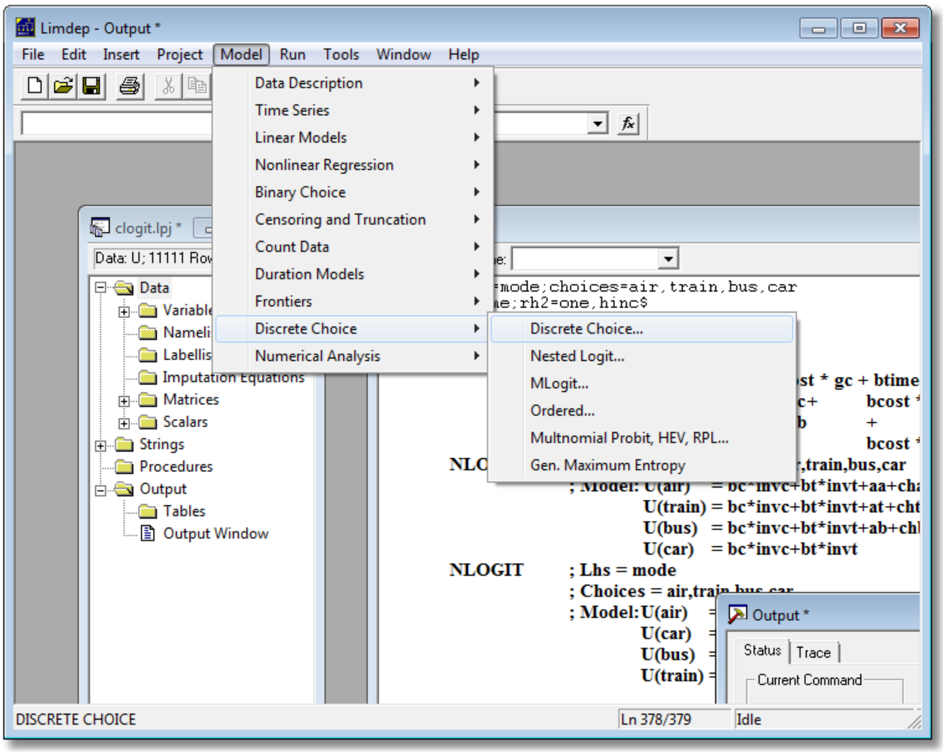


Figure N19.1 Command Builders for *NLOGIT* Models

The Main and Options pages of the command builder for the conditional logit model are shown in Figures N19.2, N19.3 and N19.4. (Some features of the models, and the ECM model, are not provided by the command builders. Most of the features of these models are much easier to specify in the editor using the command mode of entry.) 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 variables of the model are defined in the left window and the Rh2 variables are specified in the right window.

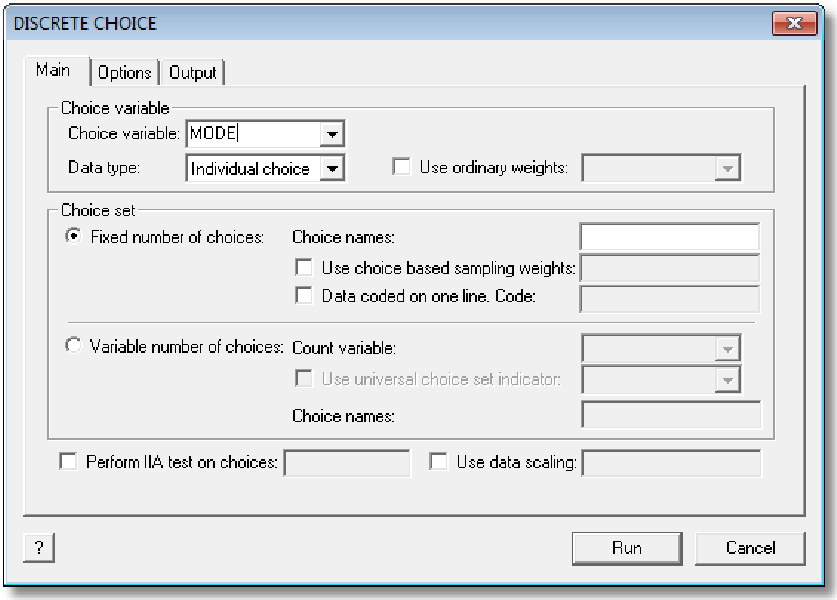


Figure N19.2 Discrete Choice Command Builder Main Page

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 100.

There are *K* attributes (Rhs variables) measured for the choices. The sections below 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 100.

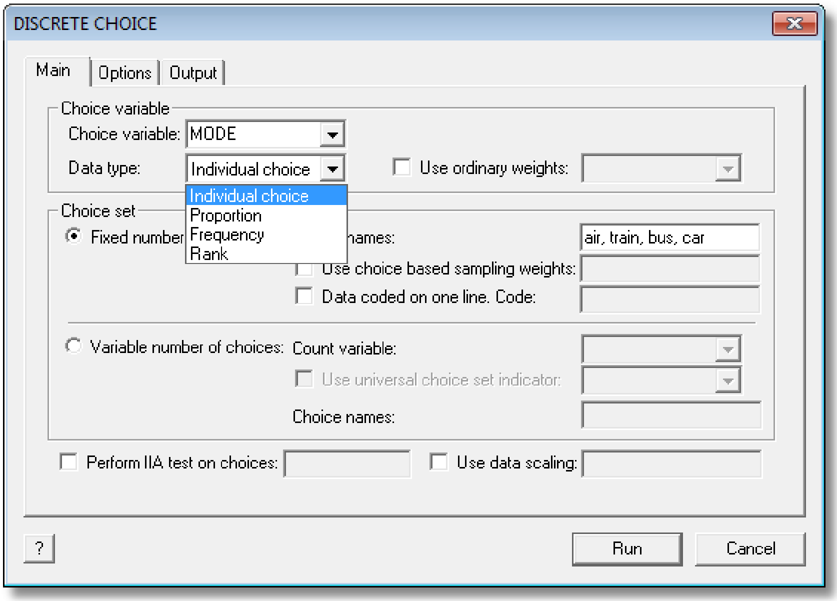


Figure N19.3 Specifying Choices on Command Builder Main Page

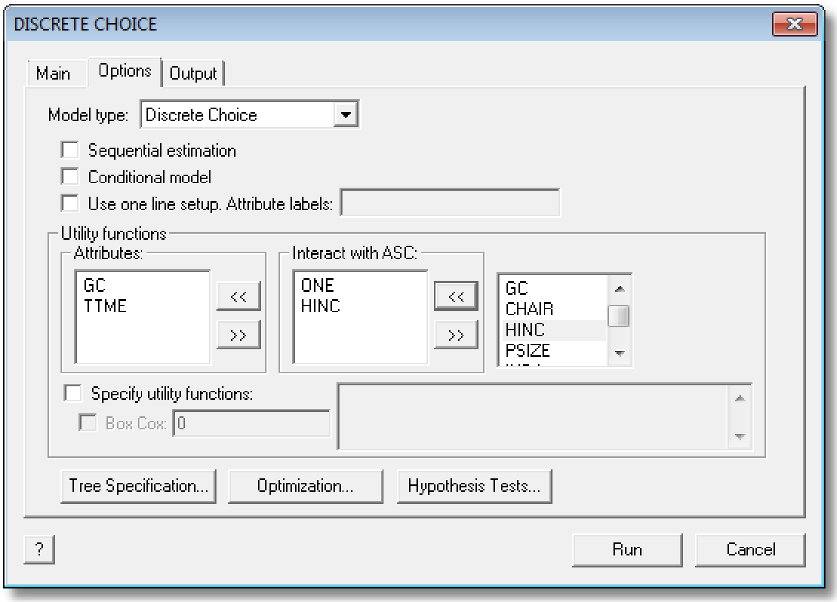


Figure N19.4 Options Page of Command Builder for Conditional Logit Model

The random utility model specified by this setup is precisely of the form

$$U_{i,j} = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_{K1} x_{i,K1} + \gamma_{1,j} z_{i,1} + \ldots + \gamma_{K2,j} z_{i,K2} + \varepsilon_{i,j},$$

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 several points below.

N19.3 Other Optional Specifications on *NLOGIT* Commands

The **NLOGIT** command operates like other *LIMDEP* model commands. The following lists command features and options that are used with this command. There are numerous additional command specifications that are used with the specific models fit with *NLOGIT*, such as **; RPM** to specify a random parameters model, and **; Umax** which is a technical specification if it is necessary to control the accumulation of rounding error in estimating certain models.

Controlling Output from Model Commands

; Par	saves person specific parameter vectors, used with the random parameters logit model and heteroscedastic extreme value model.
; Effects: spec	displays partial effects and elasticities of probabilities.
; Table = name	adds model results to stored tables.

Robust Asymptotic Covariance Matrices

; Covariance Matrix	displays estimated asymptotic covariance matrix (normally not shown).
; Cluster = spec	computes robust cluster corrected asymptotic covariance matrix.
; Robust	computes robust sandwich estimator for asymptotic covariance matrix.

Optimization Controls for Nonlinear Optimization

; Start = list	provides 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.
; Alg = name	specifies optimization method. 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.
; Umax = value	specifies maximum allowed value for utility in MNL probability. (Generally not needed save for extreme data. Default = 300.)

Predictions and Residuals

; List	lists predicted probabilities and predicted outcomes with model results.
; Keep = name	keeps fitted values as a new (or replacement) variable in data set. (Several other similar specifications are used with **NLOGIT**.)
; Prob = name	keeps probabilities as a new (or replacement) variable.

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	imposes fixed value and equality constraints.

; Maxit = 0 ; Start = the restricted values specifies Lagrange multiplier test.

N19.4 Estimation Results

This section will detail the common results produced by the different models in *NLOGIT*.

N19.4.1 Descriptive Headers for *NLOGIT* Models

The output for the *NLOGIT* estimators may contain a description of the model before the statistical results. The description consists of a table that shows the sample proportions and the tree structure if you fit a nested logit model, and a table that lists the components of the utility functions. You can request these listings by adding

; Show Model

to your **NLOGIT** command. (We used this device in several earlier examples.) 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

```
NLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = invc,invt,gc
          ; Rh2 = one,hinc
          ; Show Model $
```

```
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               |
+----------------------------------------------------------------+
```

The initial header includes a display of the tree structure when you fit a nested logit model. For example, the command

NLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = invc,invt,gc
; Rh2 = one,hinc
; Tree = Public[(air),(train,bus)],Private[(car)]
; Show Model $

produces the header:

```
Tree Structure Specified for the Nested Logit Model
Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.
----------------+----------------+----------------+----------------+------+---
Trunk    (prop.)|Limb     (prop.)|Branch   (prop.)|Choice   (prop.)|Weight|IIA
----------------+----------------+----------------+----------------+------+---
Trunk{1} 1.00000|PUBLIC    .71905|B(1|1,1)  .27619|AIR       .27619| 1.000|
                |                |B(2|1,1)  .44286|TRAIN     .30000| 1.000|
                |                |                |BUS       .14286| 1.000|
                |PRIVATE   .28095|B(1|2,1)  .28095|CAR       .28095| 1.000|
----------------+----------------+----------------+----------------+------+---
```

(Note, this particular model is not identified – we specified it only for purpose of illustrating the display of its tree structure.)

N19.4.2 Standard Model Results

Estimation results for the model commands consist of the initial display of diagnostic followed by notes about the model, then the estimated coefficients. The preceding command, without the tree structure or the initial echo of the model specification,

NLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = invc,invt,gc
; Rh2 = one,hinc $

produces the following results:

```
Normal exit from iterations. Exit status=0.
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -246.10979
Estimation based on N =    210, K =   9
Inf.Cr.AIC  =    510.2 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

```
NLOGIT     ; Lhs = mode ; Show
           ; Choices = air,(train),(bus),car
           ; Rhs = gc,ttme ; Rh2 = one $
```

```
+--------------------------------------------------------+
|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.
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:   6 iterations. Status=0, F=    62.58418

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -62.58418
Estimation based on N =    117, K =   3
Inf.Cr.AIC  =    131.2 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).....
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

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 is -62.58418. 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) = 0.22829 (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 - $K(\log L / \text{Log} L_0)$;

$$Adjusted\ R^2 = 1 - 1.02632\ (-62.54818/-81.0939) = 0.20794.$$

N19.4.3 Retained Results

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
lastiter = number of iterations completed for estimation
opt_exit = exit code. 0.0 indicates successful estimation
modelaic = AIC for estimated model
modelbic = BIC for estimate model

Last Model: *b_variable* = the labels kept for the **WALD** command

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:

```
NLOGIT     ; Lhs = mode ; Show
           ; Choices = air,train,bus,car
           ; Rhs = gc,ttme ; Rh2 = one $
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              =     78.54713
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)|    12.9101***      1.45668     8.86  .0000     10.0550    15.7651
--------+--------------------------------------------------------------------
```

N19.4.4 Descriptive Statistics for Alternatives

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:

```
NLOGIT	; Lhs = mode ; Choices = air,train,bus,car
	; Rhs = invc,invt,gc ; Rh2 = one,hinc
	; Show Model
	; 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 the two sets of results below. Note the first cross tabulation is based on the fitted probabilities while the second is based on the observed choices.

```
+-------------------------------------------------------+
| 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
```

N19.5 Calibrating a Model

When the data consists of two subsets, for example an RP data set and a counterpart SP data set, it is sometimes useful to fit the model with one of the data sets, then refit the second one while retaining the original coefficients, and just adjusting the constants. Consider the application below:

```
SAMPLE    ; 1-420 $
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Model: U(air)   = aa + gc * gc + ttme * ttme + invt * invt /
                   U(train) = at + gc * gc + ttme * ttme /
                   U(bus)   = ab + gc * gc + ttme * ttme /
                   U(car)   =      + gc * gc + ttme * ttme $
SAMPLE    ; 421-840 $
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Model: U(air)   = aa + gc[ ] * gc + ttme[ ] * ttme + invt[ ] * invt /
                   U(train) = at + gc * gc + ttme * ttme /
                   U(bus)   = ab + gc * gc + ttme * ttme /
                   U(car)   =      + gc * gc + ttme * ttme $
```

The model is first fit with the first half of the data set (observations 1 - 105). Then, for the second estimation, we want to refit the model, but only recompute the constant terms but keep the previously estimates slope parameters. The device to use for the second model is the **‘[ ]’** specification, which indicates that you wish to use the previously estimated parameters. The commands above will, in principle, produce the desired result, with one consideration. Newton’s method is very sensitive to the starting values for this model, and with the constraints imposed in the second model, will generally fail to converge. (See the example below.) The practical solution is to change the algorithm to BFGS, which will then produce the desired result. You can do this just by adding

; Alg = BFGS

to the second command. An additional detail is that the second model will now replace the first as the ‘previous’ model. So, if you want to do a second calibration, you have to refit the first model. To preempt this, you can use

; Calibrate

in the second command. This specification changes the algorithm and also instructs *NLOGIT* not to replace the previous estimates with the current ones. Three notes about this procedure:

- You may use this device with any discrete choice model that you fit with *NLOGIT*.
- The second sample must have the same configuration as the first.
- The device can only be used to fix the utility function parameters.

The third point implies that if you do this with a random parameters model, the random parameters will become fixed – have the variances fixed at zero.

The commands above (with the addition of **; Calibrate** to the second **CLOGIT** command) produce the following results: (Some parts of the results are omitted.) The note before the second set of results has been produced because the estimator converges very quickly – this will usually happen when the model contains only the alternative specific constants.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -93.51621
Estimation based on N =    105, K =   6
Inf.Cr.AIC  =    199.0 AIC/N =    1.896
Number of obs.=   105, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      AA|    7.94929***     1.44243     5.51  .0000     5.12217  10.77641
      GC|    -.01705***      .00626    -2.72  .0064     -.02931   -.00478
    TTME|    -.08983***      .01452    -6.19  .0000     -.11829   -.06136
    INVT|    -.01974**       .00775    -2.55  .0109     -.03494   -.00455
      AT|    4.31669***      .64859     6.66  .0000     3.04549   5.58790
      AB|    2.60715***      .72991     3.57  .0004     1.17656   4.03774
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      AA|-.22520D+33***     1.00000 ********  .0000 -.22520D+33  -.22520D+33
      GC|    -.01705    .....(Fixed Parameter).....
    TTME|    -.08983    .....(Fixed Parameter).....
    INVT|    -.01974    .....(Fixed Parameter).....
      AT| .24951D+34    .....(Fixed Parameter).....
      AB| .68897D+33    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -97.65109
Number of obs.=   105, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      AA|    8.06593***      .29707    27.15  .0000     7.48368   8.64817
      GC|    -.01705    .....(Fixed Parameter).....
    TTME|    -.08983    .....(Fixed Parameter).....
    INVT|    -.01974    .....(Fixed Parameter).....
      AT|    2.94882***      .34838     8.46  .0000     2.26600   3.63164
      AB|    3.09656***      .31503     9.83  .0000     2.47910   3.71402
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

N20: Choice Sets and Utility Functions

N20.1 Introduction

Chapter N17 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.

N20.2 Choice Sets

Every multinomial model fit by *NLOGIT* must include a specification for the choice variable and a definition of the choice set. The basic formulation would appear as

; Lhs = the dependent, or choice variable
; Choices = the names of the choices in the model

Several variations on this formula appear in Sections N20.3 and N20.4. In general, your dependent variable is the name of a variable which indicates by a one or zero whether a particular alternative is selected, or it gives the proportion or frequency of individuals sampled that selected a particular alternative. When they are enumerated, the **; Choices** list gives names and possibly sampling weights for the set of alternatives.

All command builders begin with these two specifications. The discrete choice and nested logit models allow the full set of variants discussed in this section while the other command builders expect the simple form with a fixed choice set. The Main page of the conditional logit command builder shown in Figure N20.1 illustrates. (A similar Main page is used for the nested logit command builder.) The command builder allows you to specify the choice variable and type of choice set in the three sections of this dialog box.

NOTE: The command builder for the multinomial probit, HEV and RPL models requires you to provide a fixed sized choice set. This is a limitation of the command builder window, not the estimator. With the exception of the multinomial probit model, this is not a requirement of the models themselves. Only the multinomial probit model requires the number of choices to be fixed. For the HEV and RPL models, if you build your command in the text editor, rather than with the command builder, you may specify a variable choice set, as described in Section N20.2.1.

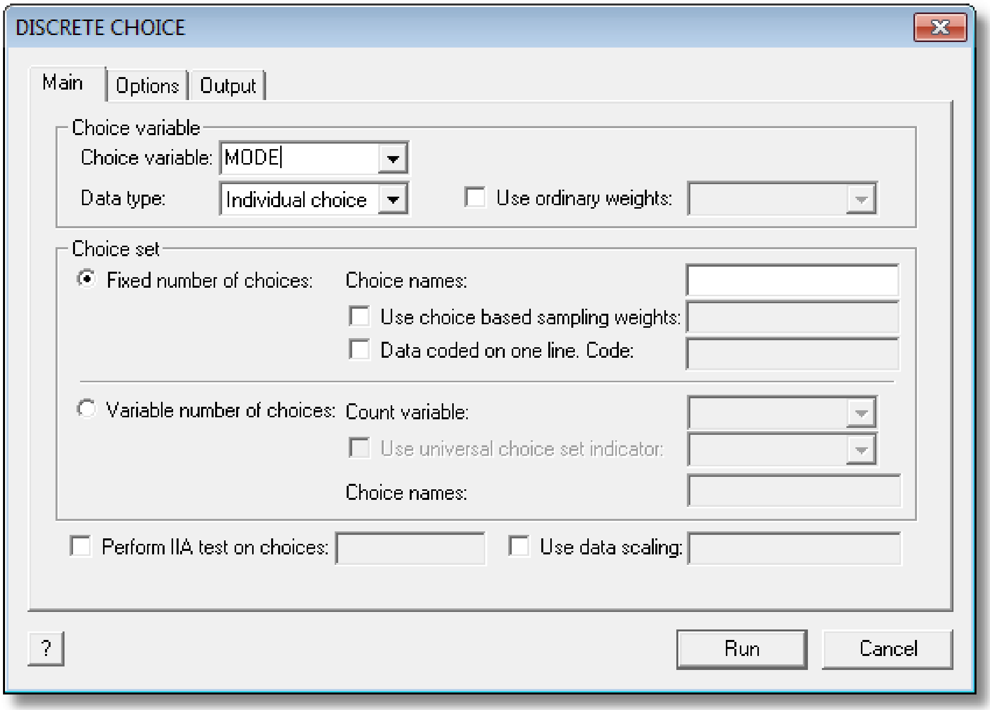


Figure N20.1 Main Page of Command Builder for Conditional Logit Model

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.

The following modifications apply to all multinomial models that are fit with *NLOGIT*. We use **NLOGIT** as the generic verb for this description. Any of the others described in the next chapter will be treated the same. Note, as well, the **NLOGIT** commands, which do not contain any additional model specifications, will be equivalent to and act like **CLOGIT** commands. That is, the command, **NLOGIT**, with no additional model specifications is equivalent to **CLOGIT**. (It is also the same as **DISCRETE CHOICE**, which although no longer used by *NLOGIT*, remains acceptable as the basic model verb.)

N20.2.1 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 N17, and will consist of *n* sets of *J* 'observations.' You indicate this case with a command such as:

NLOGIT ; Lhs = the choice variable
; Choices = ... a list of J names for the choices
; ... the rest of the command $

For example,

NLOGIT ; Lhs = mode
; Choices = air,train,bus,car
; etc. $

A fixed choice set can be specified in the command builder as shown in Figure N20.2.

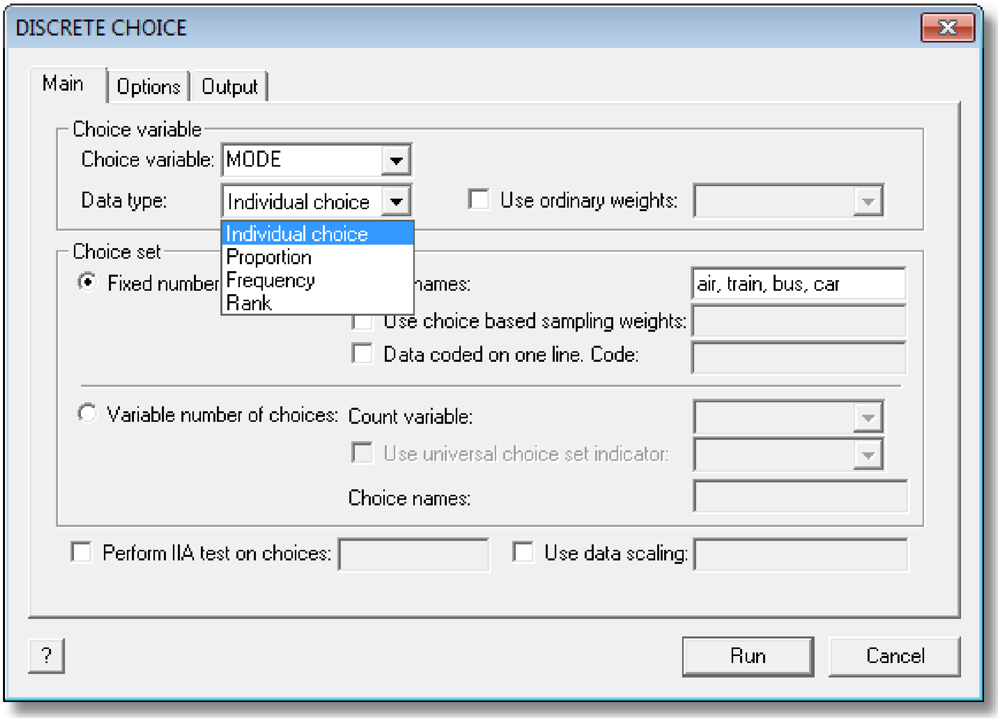


Figure N20.2 Fixed Choice Set Specified in Command Builder

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,…

Any of these cases can be accommodated with *NLOGIT*. 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

```
NLOGIT    ; 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. *NLOGIT* 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 *NLOGIT*. The model command might be

NLOGIT ; 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,

NLOGIT ; 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, *NLOGIT* 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.

N20.2.2 Restricting the Choice Set

The IIA test described later in Section N21.4.1 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 train and bus 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|   |
+----------------+------+---+
```

N20.2.3 A Shorthand for Choice Sets

You may use

; Choices = number_name

To define a set of choice labels of the form name1, name2, … For example,

; Choices = 5_brand

Creates choice labels *brand1*, *brand2*, *brand3*, *brand4*, *brand5*. This sort of construction is likely to be useful for unlabeled choice experiments.

N20.2.4 Large Choice Sets – A Panel Data Equivalence

The conditional logit estimator can fit a model with up to 500 choices, which is quite large. Chamberlain's fixed effects model for the binary logit model described in Section N9.5 can also be used to fit a discrete choice model. The log likelihood function for this model is

$$L_c = \frac{\prod_{t=1}^{T_i} \exp[y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}]}{\sum_{\text{all arrangements of } T_i \text{ outcomes with the same sum}} \prod_{s=1}^{T_i} \exp[y_{is}\boldsymbol{\beta}'\mathbf{x}_{is}]}$$

$$= \frac{\exp\left[\sum_{t=1}^{T_i} y_{it}\boldsymbol{\beta}'\mathbf{x}_{it}\right]}{\sum_{\text{all arrangements of } T_i \text{ outcomes with the same sum}} \exp\left[\sum_{s=1}^{T_i} d_{is}\boldsymbol{\beta}'\mathbf{x}_{is}\right]}.$$

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 Chapter N17. 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. A 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
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
--------+--------------------------------------------------------------------
```

N20.2.5 An Alternative Data Arrangement

NLOGIT uses two variables keep track of different choice set sizes and groups of choice tasks in choice experiments. These are shown in Figure N20.3. The data shown in the figure are for two individuals in a choice experiment. The first makes five choices, the second makes three. The choice set sizes are (3,4,3,4,2) for individual 1 and (2,3,3) for individual 2. The variables *count* and *ntask* would be used by *NLOGIT* to account for these parameters. Your command would appear

… ; Lhs = choice, count ; Pds = ntask

The data for your analysis might be arranged in a different format from what *NLOGIT* expects. The first two columns in Figure N20.3 show an alternative arrangement that we have seen in some applications. This is not the standard format used by the program. However, it is possible to deduce '*count*' and '*ntask*' from '*person*' and '*task*.' A processor is provided for you to automate the conversion. This is a onetime setting that you must make before you use these data for estimating choice models. The command is

SETCHOICE ; Group = the name of the 'person' variable in your data
; Choice = the name of the 'task' variable in your data
; Cset = the name of the 'count' variable that will be created
; Nset = the name of the 'ntask' variable that will be created

(**; Cset** and **; Nset** can provide any names you desire.) Once this command is executed, the configuration of the data will be maintained internally, and you will use the default program settings from then on. For example, assuming that *count* and *ntask* did not already exist in your data set, you would use

SETCHOICE ; Group = person
; Choice = task
; Cset = count
; Nset = ntask $

and thereafter,

; Lhs = choice,count ; Pds = ntask.

	A	B	C	D	E
1	Person	Task	Count	Ntask	Choice
2	1	1	3	5	0
3	1	1	3	5	1
4	1	1	3	5	0
5	1	2	4	5	1
6	1	2	4	5	0
7	1	2	4	5	0
8	1	2	4	5	0
9	1	3	3	5	0
10	1	3	3	5	0
11	1	3	3	5	1
12	1	4	4	5	0
13	1	4	4	5	1
14	1	4	4	5	0
15	1	4	4	5	0
16	1	5	2	5	1
17	1	5	2	5	0
18	2	6	2	3	0
19	2	6	2	3	1
20	2	7	3	3	0
21	2	7	3	3	0
22	2	7	3	3	1
23	2	8	3	3	1
24	2	8	3	3	0
25	2	8	3	3	0

Figure N20.3 Alternative Data Arrangement

N20.3 Specifying the Utility Functions with Rhs and Rh2

There are several ways to specify the utility functions in your **NLOGIT** command, in the text editor and in the command builder. In order to provide a simple explanation that covers the cases, we will develop the application that will be used in the chapters to follow to illustrate the models. The application is based on the data summarized in Section N18.11. We will model travel mode choice for trips between Sydney and Melbourne with utility functions for the four choices as follows:

		gc	***ttme***	***one***	***hinc***	***one***	***hinc***	***one***	***hinc***	***one***	***hinc***
U(*air*)	=	GC	TTME	A_AIR	AIR_HIN1	0	0	0	0	0	0
U(*train*)	=	GC	TTME	0	0	A_TRAIN	TRA_HIN2	0	0	0	0
U(*bus*)	=	GC	TTME	0	0	0	0	A_BUS	BUS_HIN3	0	0
U(*car*)	=	GC	TTME	0	0	0	0	0	0	0	0

The columns are headed by the names of variables, generalized cost (*gc*), terminal time (*ttme*) and household income (*hinc*). The entries in the body of the table are the names given to coefficients that will multiply the variables. Note that the generic coefficients in the first two columns are given the names of the variables they multiply while the interactions with the constants are given compound names. It is important to note the last two columns. The last one in a set of choice specific constants or variables that are interacted with them must be dropped to avoid a problem of collinearity in the model. In what follows, for brevity, we will omit these two columns. Before proceeding, we note the format of a set of parameter estimates for a model set up in exactly this fashion:

```
--------+----------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*          Interval
--------+----------------------------------------------------------------
      GC|    -.01093**       .00459    -2.38  .0172     -.01992   -.00194
    TTME|    -.09546***      .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|    5.87481***      .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|    -.00537         .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|    5.54986***      .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|    -.05656***      .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|    4.13028***      .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|    -.02858*        .01544    -1.85  .0642     -.05885    .00169
--------+----------------------------------------------------------------
```

Note the construction of the compound names includes what might seem to be a redundant number at the end. This is necessary to avoid constructing identical names for different variables.

N20.3.1 Utility Functions

A basic four choice model which contains *cost*, *time*, *one* and *income* will have utility functions

$$\begin{aligned}
U_{i,air} &= \beta_{cost}\, cost_{i,air} + \beta_{time}\, time_{i,air} + \alpha_{air} + \gamma_{air}\, income_i + \varepsilon_{i,air},\\
U_{i,train} &= \beta_{cost}\, cost_{i,train} + \beta_{time}\, time_{i,train} + \alpha_{train} + \gamma_{train}\, income_i + \varepsilon_{i,train},\\
U_{i,bus} &= \beta_{cost}\, cost_{i,bus} + \beta_{time}\, time_{i,bus} + \alpha_{bus} + \gamma_{bus}\, income_i + \varepsilon_{i,bus},\\
U_{i,car} &= \beta_{cost}\, cost_{i,car} + \beta_{time}\, time_{i,car} + \varepsilon_{i,bus}.
\end{aligned}$$

The device you will use to construct utility functions in this fashion is

; 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 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 in Section N18.11 actually does contain the superfluous set of four choice specific constants, *aasc*, *tasc*, *basc* and *casc*.

NOTE: If you include *one* in your Rhs list, 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.)

N20.3.2 Generic Coefficients

The way to specify generic coefficients in a model is to use *NLOGIT*'s standard construction, by specifying a set of Rhs variables. The specification

; Rhs = gc,ttme

produces the utility functions in the first two columns in the table. Rhs variables are assumed to vary across the choices and will receive generic coefficients.

N20.3.3 Alternative Specific Constants and Interactions with Constants

The logit model is homogeneous of degree zero in the attributes. Any attribute which does not vary across the choices, such as age, marital status, income etc., will simply fall out of the probability. Consider an example with a constant, one attribute and one characteristic,

$$\begin{aligned}\text{Prob}(choice\ j) &= \frac{\exp(\alpha + \beta_1 cost_{ij} + \beta_2 income_i)}{\Sigma_{j=1}^{J} \exp(\alpha + \beta_1 cost_{ij} + \beta_2 income_i)} \\ &= \frac{\exp(\alpha + \beta_2 income_i)\exp(\beta_1 cost_{ij})}{\Sigma_{j=1}^{J} \exp(\alpha + \beta_2 income_i)\exp(\beta_1 cost_{ij})} \\ &= \frac{\exp(\alpha + \beta_2 income_i)\exp(\beta_1 cost_{ij})}{\exp(\alpha + \beta_2 income_i)\Sigma_{j=1}^{J} \exp(\beta_1 cost_{ij})} \\ &= \frac{\exp(\beta_1 cost_{ij})}{\Sigma_{j=1}^{J} \exp(\beta_1 cost_{ij})}.\end{aligned}$$

With a generic coefficient, the choice invariant characteristic and the single constant term fall out of the model. A model which contains such a characteristic with a generic coefficient is not estimable. This carries over to all of the more elaborate models such as the HEV, nested logit and MNP models as well. The solution to this complication is to create choice specific constant terms and, if need be, interact the invariant characteristic with the constant term. This is what appears in the last eight columns in the example above. (This is how the MLOGIT model in Chapter N16 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.

$$\text{Prob}(choice = j) = \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\sum_{j=1}^{J} \exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}.$$

There remains an indeterminacy in the model after it is expanded in this fashion. Suppose the same constant, say θ, is added to each γ_j. The resulting model is

$$\begin{aligned}
\text{Prob}(choice = j) &= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + (\gamma_j + \theta)Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + (\gamma_j + \theta)Income_i)} \\
&= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i + \theta Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i + \theta Income_i)} \\
&= \frac{\exp(\theta Income_i)\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\exp(\theta Income_i)\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)} \\
&= \frac{\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}{\sum_{j=1}^{J}\exp(\beta_1 cost_{i,j} + \alpha_j + \gamma_j Income_i)}.
\end{aligned}$$

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. *NLOGIT* sets the last one to zero. The same result applies to the choice specific constant terms that you create with *one*. This produces the data matrix shown earlier, with the last two columns (in the dashed box) normalized to zeros.

Finally, while it is necessary for choice invariant variables to appear in the Rh2 list, 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 *cost* 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 right hand side. 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 $
NLOGIT      ; ...  ; 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.

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, *NLOGIT* 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*, *NLOGIT* will create a full set of *J*-1 choice specific constants. (One of them must be dropped to avoid what amounts to the dummy variable trap.)

HINT: You need not have choice specific dummy variables in your data set. The Rh2 setup described here allows you to produce these variables as part of the model specification.

The remaining columns of the utility functions in the example above are produced with

; Rh2 = one,hinc

You should note, in addition, how the variables are expanded, as a set, in constructing the utility functions.

N20.3.4 Command Builders

You can specify utility functions in this format in any of the command builders, as shown in Figure N20.4. The two windows allow you to select variables from the list at the right and assemble the Rhs list at the left or the Rh2 list in the center.

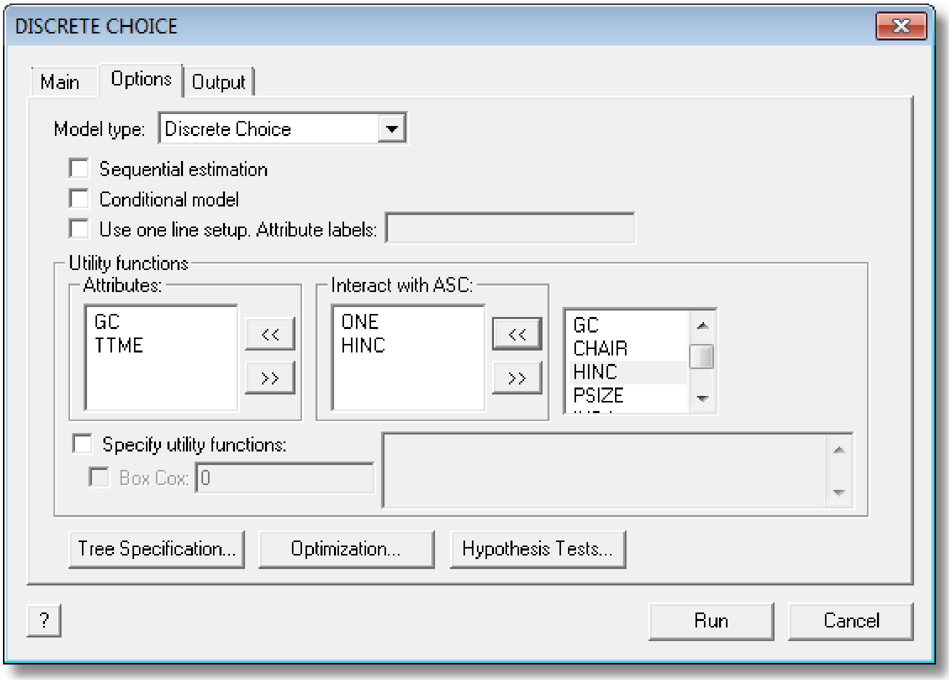


Figure N20.4 Specifying Utility Functions in the Command Builder

N20.4 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.

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

choice cost time constants income

$$\mathbf{Z}_i = \begin{bmatrix} y_{air} & c_a & t_a & 1 & 0 & 0 & income & 0 & 0 \\ y_{train} & c_t & t_t & 0 & 1 & 0 & 0 & income & 0 \\ y_{bus} & c_b & t_b & 0 & 0 & 1 & 0 & 0 & income \\ y_{car} & c_c & t_c & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

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

$$\begin{aligned}
U_{i,air} &= \beta_{cost1}cost_{i,air} \phantom{+ \beta_{cost2}cost_{i,train}} + \beta_{time}time_{i,air} + \alpha_{air} + \gamma_{air}income_i + \varepsilon_{i,air}\\
U_{i,train} &= \beta_{cost2}cost_{i,train} + \beta_{time}time_{i,train} + \alpha_{train} + \gamma_{train}income_i + \varepsilon_{i,train}\\
U_{i,bus} &= \beta_{cost2}cost_{i,bus} + \beta_{time}time_{i,bus} + \alpha_{bus} + \gamma_{bus}income_i + \varepsilon_{i,bus}\\
U_{i,car} &= \beta_{cost1}cost_{i,car} + \beta_{time}time_{i,car} + \alpha_{car} + \gamma_{car}income_i + \varepsilon_{i,car}
\end{aligned}$$

This section will describe how to structure the utility functions individually, rather than generically with Rhs and Rh2 and transformations of variables.

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 : ….** *NLOGIT* 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:

```
NLOGIT    ; 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 also 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

```
NLOGIT    ; 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

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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

N20.4.1 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 would use the variable names instead of the supplied parameter names for the two parameters, but the models will be the same.

N20.4.2 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.

The command builders provide space for you to build the utility functions in this fashion. See Figure N20.5. Since this is done by typing out the functions in the windows – there is no menu construction that would allow this – these will not save much effort.

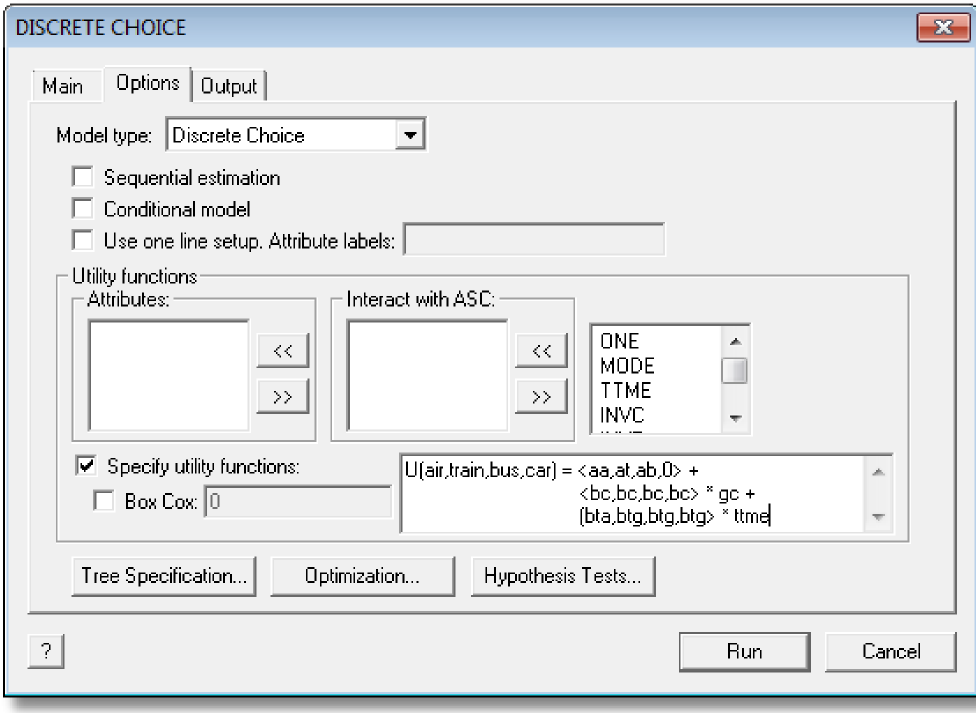


Figure N20.5 Utility Functions Assembled in Command Builder Window

Note that in the window, you must provide the entire specification for the utility functions, including the listing of which alternatives the definitions are to apply to. The model shown in the window in Figure N20.5 produces these results.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable              Choice
Log likelihood function     -199.68246
Estimation based on N =    210, K =   6
Inf.Cr.AIC  =    411.4 AIC/N =    1.959
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2963 .2895
Chi-squared[ 3]          =    168.15262
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
--------+--------------------------------------------------------------------
      AA|   6.41354***    1.10452     5.81  .0000     4.24871    8.57836
      AT|   3.69564***     .52116     7.09  .0000     2.67418    4.71711
      AB|   2.96222***     .54485     5.44  .0000     1.89433    4.03011
      BC|   -.01702***     .00471    -3.61  .0003     -.02626    -.00778
     BTA|   -.10758***     .01792    -6.00  .0000     -.14270    -.07246
     BTG|   -.08940***     .01419    -6.30  .0000     -.11722    -.06158
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

N20.4.3 Logs and the Box Cox Transformation

Variables may be specified in logarithms. This will be useful when you are using aggregate data and you wish to include, e.g., market size in a choice. To indicate that you wish to use logs, use **Log**(*variable name*) instead of just *variable name* in the utility definition. (The syntax **; Rhs = ... Log(x)** as described above is not available. This option may only be used when you are explicitly defining the utility functions.) Thus, the model above might have been

```
NLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Model: U(air)   = ba + bcost * Log(gc)  /
                   U(car)   = bc + bcost * Log(gc)  /
                   U(bus)   = bb + bcost * Log(gc)  /
                   U(train) =      bcost * Log(gc)  $
```

When a variable appears in more than one utility function, you should take logs each time it appears. Although this is not mandatory, if you do not, your model will contain a mix of levels and logs, which is probably not what you want. Also, it will be necessary for you to be aware in your results when you have used this transformation. The model results will not contain any indication that logs have appeared in the equation. The preceding, for example, produces the following estimation results:

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
BA	-.59298***	.21340	-2.78	.0055	-1.01124	-.17473
BCOST	-2.63022***	.45171	-5.82	.0000	-3.51555	-1.74489
BC	-.95454***	.24331	-3.92	.0001	-1.43141	-.47767
BB	-.97857***	.22952	-4.26	.0000	-1.42841	-.52872

You may also use the Box-Cox transformation to transform variables. Indicate this with Bcx(x) where x is the variable (which must be positive). The transformation is

$$\text{Bcx}(x) = (x^{\lambda} - 1) / \lambda,$$

which is Log(x) if λ equals 0 and is x**-**1 (not x) if λ equals 1. The Bcx(**.**) function may appear any number of times in the model specification. In general, if a variable is transformed with this function, it should be transformed every time it appears in the model. Not doing so is analogous to including both levels and logs of a variable, which while not invalid, is usually avoided. The default value of the transformation parameter, λ, is 1.0. The same value is used in all transformations. You may specify a different value by including the specification

; Lambda = value

in your **NLOGIT** command. *Lambda* is treated as a fixed value during estimation, not an estimated parameter. Thus, no standard error is computed for *lambda* (since you provide the fixed value) and the standard errors for the other estimates are not adjusted for the presence of *lambda*. I.e., by this construction, the Box-Cox transformation is treated like the log function – just a transformation. In this case, the model results will contain an indication that the transformation has appeared in the utility functions. For example, the preceding, with λ = 0.5, produces:

```
Normal exit:   4 iterations. Status=0, F=    267.4253

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -267.42533
Estimation based on N =    210, K =   4
Inf.Cr.AIC  =    542.9 AIC/N =    2.585
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .0576 .0515
Chi-squared[ 1]          =     32.66687
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Box-Cox model. LAMBDA used is    .50000
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      BA|    -.64256***      .21843    -2.94  .0033    -1.07068   -.21445
   BCOST|    -.24334***      .04456    -5.46  .0000     -.33068   -.15601
      BC|    -.84570***      .23246    -3.64  .0003    -1.30132   -.39008
      BB|    -.99967***      .22980    -4.35  .0000    -1.45007   -.54927
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Do note, however, that the results can only indicate that a Box-Cox transformation using λ = 0.5 has appeared in the model. It is not possible to report where it appears.

N20.4.4 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  $
```

N20.5 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.

In the generic form of the utility functions, when you use **; Rhs** and **; Rh2**, you may also provide starting values for your parameters with

; Start = the list of values

The values must be provided in the order in which the model constructs them from your lists. Thus, the Rhs variables appear first, followed by the Rh2 variables interacted with the alternative specific constants. For the example earlier,

; Rhs = gc,ttme ; Rh2 = one,income

the coefficients are $\boldsymbol{\beta} = (\beta_{gc},\beta_{ttme},\alpha_{air},\gamma_{air},\alpha_{train},\gamma_{train},\alpha_{bus},\gamma_{bus})$.

There are cases in which some starting values are better than others in terms of the path of the iterations to a solution. However, since the log likelihood function is globally concave, if the solver is going to find the MLE, it will find the same MLE regardless of the starting values. In principle, this makes starting values irrelevant. But, providing starting values does allow you to compute the log likelihood function at a particular set of parameters. You can also use **; Maxit = 0** to instruct the estimator to compute a Lagrange multiplier statistic based on a particular set of values. The LM statistic is discussed in Chapter N21.

N20.5.1 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 immediately 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.

For the generic utility function setup using the Rhs and Rh2 lists, you can also fix coefficients at specific values by using

; Fix = name[value], ...

for as many coefficients as you like. The '*name*' is the name that is given to the coefficient. If the coefficient multiplies a Rhs variable, that is just the variable name. If it is an Rh2 variable, that will be the compound constructed name. These are a bit complex, but a strategy you can use is to fit the model first without the fixed value constraint. The output will show the constructed names that you can then use in your specification.

N20.5.2 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.) 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.

N21: Post Estimation Results for Conditional Logit Models

N21.1 Introduction

This chapter documents the three post estimation calculations:

- Partial effects and elasticities
- Predictions of probabilities, utilities and several other variables,
- Specification testing for the IIA assumption

A fourth post estimation computation is described in Chapter N22:

- Model simulation and examination of the effects of changing scenarios on market shares.

N21.2 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

```
NLOGIT    ; 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, in addition to the summary matrix shown above, we obtain, for each alternative, two tables of results. The first displays the average partial effects and estimates of the sampling standard errors of these estimates. (Computation of standard errors for partial effects is discussed in Section N21.2.2.) The second table displays, in addition to the sample mean of the partial effects, the sample standard deviation, minimum and maximum for each effect for each alternative. The results below show the two tables for the *air* alternative:

```
-----------------------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in AIR
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AIR|     .00604***      .00175     3.46  .0005       .00262    .00946
   TRAIN|    -.00201***      .00068    -2.94  .0033     -.00334   -.00067
     BUS|    -.00124***      .00042    -2.97  .0030     -.00205   -.00042
     CAR|    -.00280***      .00081    -3.46  .0005     -.00438   -.00122
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in =AIR
--------+--------------------------------------------------------------------
        |   Average        Sample Standard      Sample       Sample
  Choice|  Elasticity         Deviation         Minimum      Maximum
--------+--------------------------------------------------------------------
     AIR|     .00604           .00017          .001180       .00908
   TRAIN|    -.00201           .00008         -.005658     -.00042
     BUS|    -.00124           .00006         -.005170     -.00008
     CAR|    -.00280           .00014         -.007631     -.00007
--------+--------------------------------------------------------------------
```

Corresponding results will be shown for the other alternatives (*train*, *bus*, *car*).

N21.2.1 Elasticities

Rather than see the partial effects, you may want to see elasticities,

$$\eta_{im}(k|j) \;=\; \partial \log \text{Prob}[y_i = m]/\partial \log x_i(k|j) \;=\; x_i(k|j)/P_{im}\times\delta_{im}(k|j)$$
$$= [\mathbf{1}(j = m) - P_{ij}]\; x_i(k|j)\beta_k$$

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 shown earlier is produced.

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.

N21.2.2 Standard Errors for Estimated Partial Effects

The sample standard deviations shown in the second table in the results below are not estimates of asymptotic standard errors for the sample statistic that is the estimated elasticity. The elasticities or other effects are computed separately for each individual in the sample, The second table below shows the sample descriptive statistics for this set of computed results. The asymptotic standard error for the estimated elasticity is computed by using the method of Krinsky and Robb. The method generates a sample of P draws from the asymptotic distribution of the model parameters. The full set of elasticities is computed using each draw and the standard error is estimated by the empirical standard deviation of this set of draws. The Krinsky and Robb results are shown in the first table.

```
--------------------------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in AIR
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AIR|     .00604***      .00175     3.46  .0005       .00262     .00946
   TRAIN|    -.00201***      .00068    -2.94  .0033     -.00334    -.00067
     BUS|    -.00124***      .00042    -2.97  .0030     -.00205    -.00042
     CAR|    -.00280***      .00081    -3.46  .0005     -.00438    -.00122
--------+-----------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

--------------------------------------------------------------------------------
Average partial effect  on prob(alt) wrt GC       in AIR
--------+-----------------------------------------------------------------------
        |    Average       Sample Standard       Sample        Sample
  Choice|  Elasticity         Deviation          Minimum       Maximum
--------+-----------------------------------------------------------------------
    =AIR|     .00604           .00017            .001180        .00908
   TRAIN|    -.00201           .00008           -.005658       -.00042
     BUS|    -.00124           .00006           -.005170       -.00008
     CAR|    -.00280           .00014           -.007631       -.00007
--------+-----------------------------------------------------------------------
```

N21.2.3 Influential Observations and Probability Weights

Elasticities and partial effects in *NLOGIT* 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. *NLOGIT* 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

   $$\eta_{im}(k|j) = \partial \text{logProb}[y_i = m]/\partial \text{log} x_i(k|j) = x_i(k|j)/P_{im} \times \delta_{im}(k|j)$$

   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
```

N21.2.4 Saving Elasticities in the Data Set

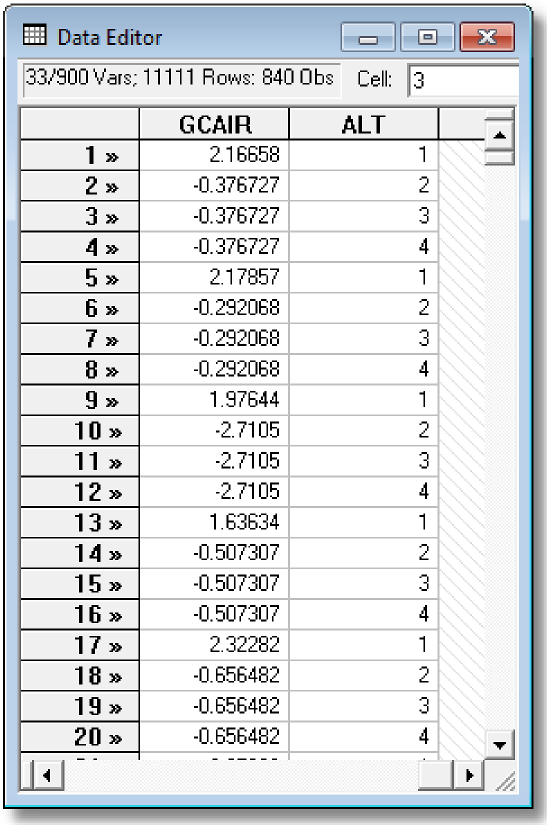
You can save the individual estimates of the own and cross elasticities as a variable in the data set by using

; 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 N21.1 shows the first few observations in the data area. The commands are:

```
NLOGIT    ; 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
-----------------+-------------------------------------------------------
```



	GCAIR	ALT
1 »	2.16658	1
2 »	-0.376727	2
3 »	-0.376727	3
4 »	-0.376727	4
5 »	2.17857	1
6 »	-0.292068	2
7 »	-0.292068	3
8 »	-0.292068	4
9 »	1.97644	1
10 »	-2.7105	2
11 »	-2.7105	3
12 »	-2.7105	4
13 »	1.63634	1
14 »	-0.507307	2
15 »	-0.507307	3
16 »	-0.507307	4
17 »	2.32282	1
18 »	-0.656482	2
19 »	-0.656482	3
20 »	-0.656482	4

Figure N21.1 Estimated Elasticities

N21.2.5 Computing Partial Effects at Data Means

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 first table above, the results would be as follows:

```
+-----------------------------------------------------+
| Derivative (times 100) Computed at sample means.    |
| Attribute is GC       in choice AIR                 |
| Effects on probabilities of all choices in model:  |
| * = Direct Derivative effect of the attribute.      |
|                                   Mean     St.Dev   |
| *     Choice=AIR                  .7263      .0000   |
|       Choice=TRAIN               -.3010      .0000   |
|       Choice=BUS                 -.1434      .0000   |
|       Choice=CAR                 -.2819      .0000   |
+-----------------------------------------------------+
```

Note that the changes are substantial. The literature is divided on this computation. Current practice seems to favor the first approach.

Rather than see the partial effects, you may want to see elasticities,

$$\eta_{im}(k|j) \;=\; \partial \log \text{Prob}[y_i = m]/\partial \log x_i(k|j) = \; x_i(k|j)/P_{im}\times\delta_{im}(k|j)$$
$$= \; [\mathbf{1}(j = m) - P_{ij}]\; x_i(k|j)\beta_k.$$

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 below. This aspect of the model is specific to the basic multinomial logit model. As will emerge in the chapters to follow, the IIA property which produces this result is absent from every other model in *NLOGIT*.

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             Averaged over observations.|
| Attribute is GC       in choice AIR               |
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
|                                 Mean    St.Dev    |
| *     Choice=AIR              2.6002     .8212    |
|       Choice=TRAIN           -1.1293     .9295    |
|       Choice=BUS             -1.1293     .9295    |
|       Choice=CAR             -1.1293     .9295    |
+---------------------------------------------------+
+---------------------------------------------------+
| Elasticity             Averaged over observations.|
| Attribute is GC       in choice TRAIN             |
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
|                                 Mean    St.Dev    |
|       Choice=AIR             -1.2046     .8221    |
| *     Choice=TRAIN            3.5259    2.1605    |
|       Choice=BUS             -1.2046     .8221    |
|       Choice=CAR             -1.2046     .8221    |
+---------------------------------------------------+
| Elasticity             Averaged over observations.|
| Attribute is GC       in choice BUS               |
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
|                                 Mean    St.Dev    |
|       Choice=AIR              -.5695     .2859    |
|       Choice=TRAIN            -.5695     .2859    |
| *     Choice=BUS              3.6181    1.4924    |
|       Choice=CAR              -.5695     .2859    |
+---------------------------------------------------+
```

```
+---------------------------------------------------+
| Elasticity             Averaged over observations.|
| Attribute is GC       in choice CAR               |
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
|                                  Mean    St.Dev   |
|       Choice=AIR               -.8688     .5119   |
|       Choice=TRAIN             -.8688     .5119   |
|       Choice=BUS               -.8688     .5119   |
| *     Choice=CAR               2.5979    1.5604   |
+---------------------------------------------------+
```

The force of the independence from irrelevant alternatives (IIA) assumption of the multinomial logit model can be seen in the identical elasticities in the tables above. The table also shows two 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 we estimated seems to 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.

N21.2.6 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 $
```

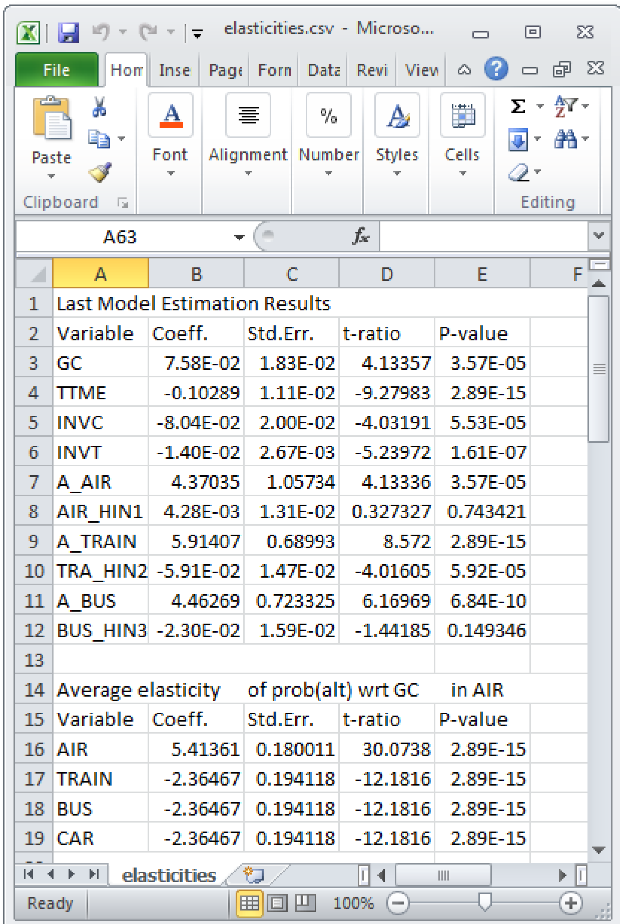
The **; Export output** setting requests that the model estimates also be included in the export file. This is followed by the tables of elasticities. The figure shows the results after the file has been read into *Excel*.

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

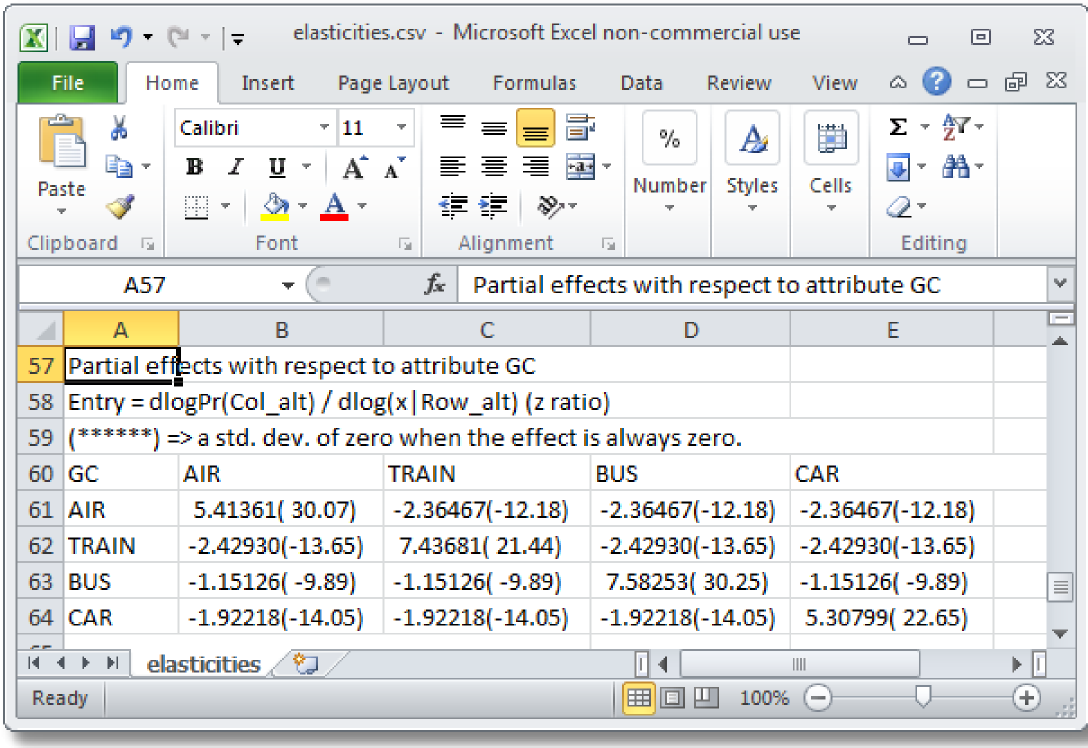
instead. Figure N21.3 shows the effect on the table shown in Figure N21.2.

> **HINT:** The export file is created while the computations are being done. However, there is a delay between when results are computed (by *NLOGIT*) and when they arrive in the file (by *Windows*). You should not try to open the export file (for example in *Excel*) while *NLOGIT* is still creating it. The results will be incomplete. Open the export file after you exit *NLOGIT*. Also, you should not try to write to an export file from *NLOGIT* 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.



	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 N21.2 Exported Model Results and Elasticities



	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 N21.3 Exported Elasticities in Matrix Format

N21.3 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 *NLOGIT*.

N21.3.1 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

```
NLOGIT    ; 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.

N21.3.2 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 cross tabulations 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.
-----------------------------------------------------------------------------
```

To continue our example,

```
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.

N21.3.3 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.

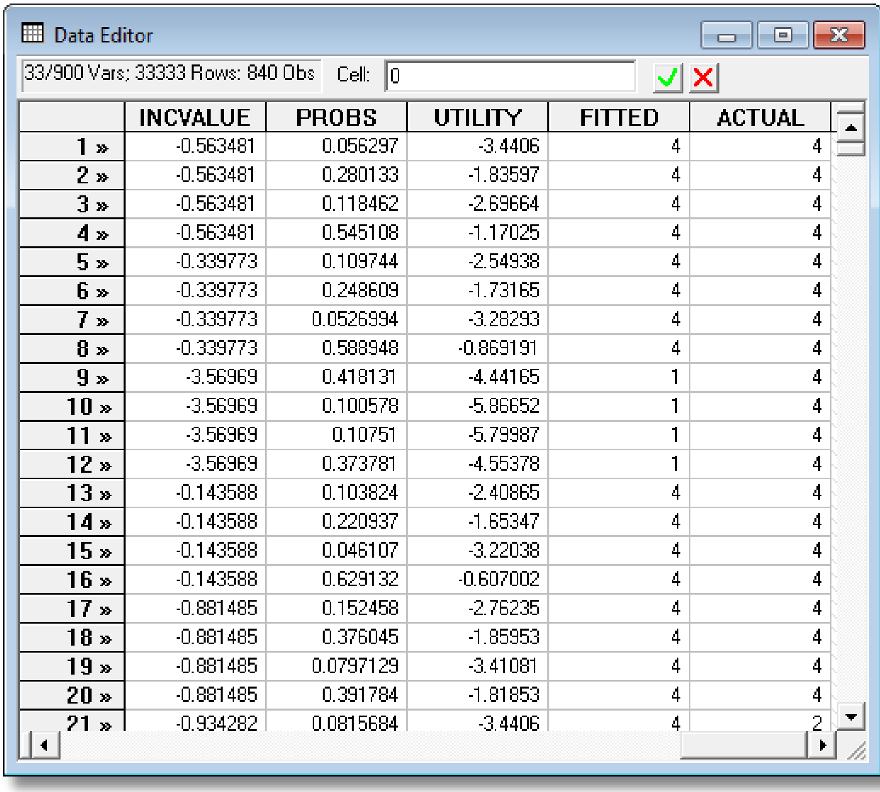
N21.3.4 Fitted Values of the Choice Variable

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 $
```



	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 N21.4 Model Predictions

N21.4 Specification Tests of IIA and Hypothesis

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.

N21.4.1 Hausman-McFadden Test of 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 (2015) and Greene (2012).) 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 *NLOGIT* 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, *NLOGIT* will send up a diagnostic about a singular Hessian (it is). Hausman and McFadden 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 = gc,invc,invt,tasc,basc,aasc,hinca $
MATRIX    ; bu = b(1:5) ; vu = Varb(1:5,1:5) $
```

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 without the air ASC and *hinca*,

```
CLOGIT      ; Lhs = mode
            ; Choices = train,bus,car
            ; Rhs = gc,invc,invt,tasc,basc $
```

Retrieve the restricted results and compute the Hausman statistic.

```
MATRIX      ; br = b(1:5) ; vr = Varb(1:5,1:5)
            ; db = br - bu ; vdb = Nvsm(vr,-vu) $
CALC        ; List
            ; q = Qfr(db,vdb)
            ; 1 - Chi(q,5) $
```

The results are:

```
[CALC] Q       =     40.5144139
[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.

N21.4.2 Small-Hsiao Likelihood Ratio Test of IIA

Small and Hsiao (1985) proposed an alternative procedure for testing IIA in the context of the CLOGIT model. The approach is similar to Hausman and McFadden, in that it is based on comparing two estimates of β that should be similar under IIA but will not be if the assumption is not met. This test is carried out via a packaged command set, rather than in internal procedure. We will lay out this routine around the specific application. Modifications needed for a different problem will be obvious. In the **NLOGIT** estimation commands, **; Quiet** is used to suppress the intermediate results.

The Small-Hsiao test is based on the likelihood function, rather than the Wald distance. The test is carried out in four steps as follows:

Step 1. Split the sample roughly equally into groups 0 and 1.
Using group 0, estimate β and retain as $b_{0.}$

Step 2. Using group 1, refit the model and retain the estimator as b_1.
Compute $b_{01} = (1/\sqrt{2})b_0 + [1-(1/\sqrt{2})]b_1$.

Step 3. Using group 1 again, fit the model using the restricted choice set.
Retain the log likelihood function, $\text{LogL}_{1.}$

Step 4. Still using group 1 and the restricted choice set, recompute the log likelihood function at $b_{01.}$
The log likelihood function is logL_{01}.

The likelihood ratio statistic is $2^*(\text{logL}_1 - \text{logl}_{01})$. By construction, this is positive, since logL_1 is the maximized value of a log likelihood while logL_{01} is the same log likelihood function computed at a value of the parameters that does not maximize it. Under the assumption of IIA, the first three steps produce what should be estimates of the same parameter vector. The logic of the test is based on the difference between b_{01} and the result at Step 3. The log likelihood function is used instead of a Wald statistic to measure the difference.

Small-Hsiao Test of IIA

The model is estimated using the full choice set, {A}= A1,….,AJ, and a restricted set of choices, B1,B2,…,BM which is a subset of {A}. (In the previous example, {A} = (air,train,bus,car) and {B} = (train,bus,car). The model contains *x* in two parts, *xtheta* is variables that are identified in both choice situations [e.g., (*gc*,*invc*,*invt*,*tasc*,*basc*) and *xgamma* is variables that are not identified by the restricted choice set [e.g., (*aasc*,*hinca*)]. The routine is as follows:

```
NAMELIST    ; xgamma = gc,invc,invt,tasc,basc ; xtheta = aasc,hinca
            ; x = xgamma,xtheta $
CALC        ; kgamma = Col(xgamma) ; nperson = 210 ; numalt = 4 $
CREATE      ; y = the choice variable $
CLIST       ; alts = air,train,bus,car $
```

We randomly select blocks of observations to split the sample. The following assumes a fixed choice set size. If not, then there must exist a variable in the data set that gives a sequential identification number to the person, repeated for each alternative within the choice set. (For the first person, if J = 5, this variable would equal 1,1,1,1,1.

```
SAMPLE      ; All $
CREATE      ; i = Trn(numalt,0) $
```

From this point, the program is generic, and need not be changed by the user. We now randomly split the sample into two sets of observations.

```
CALC        ; Ran(123457) $
MATRIX      ; split = Rndm(nperson) $
CREATE      ; ab_split = split(i) > 0 $
```

The following now carries out the test:

```
NLOGIT 1    ; For[ab_split = 0] ; Quiet ; Lhs = y ; Choices = alts ; Rhs = x $
MATRIX      ; gamma0 = b(1:kgamma) $
NLOGIT 2    ; For[ab_split = 1] ; Quiet ; Lhs = y ; Choices = alts ; Rhs = x $
MATRIX      ; gamma1 = b(1:kgamma) $
MATRIX      ; gamma01 = .7071*gamma0 + .2929*gamma1 $
NLOGIT 3    ; For[ab_split = 1] ; Quiet ; Lhs = y ; Choices = alts
            ; IAS = air ; Rhs = xgamma $
CALC        ; logl1 = logl $
NLOGIT 4    ; For[ab_split = 1] ; Quiet ; Lhs = y ; Choices = alts
            ; IAS = air ; Rhs = xgamma  ; Start = gamma01 ; Maxit = 0 $
CALC        ; List ; hs_stat = 2*(logl1 - logl) ; cvalue = Ctb(.95,kgamma) $
```

The results of this test are shown below. The chi squared statistic with five degrees of freedom is 69.921. The critical value is 11.07, so on the basis of this test, the IIA restriction is rejected. Using the Hausman-McFadden procedure in the preceding section produced a chi squared value of 40.514. The hypothesis is once again rejected.

```
-----------------------------------------------------
Setting up an iteration over the values of AB_SPLIT
The model command will be executed for     1 values
of this variable.  In the current sample of     840
observations, the following counts were found:
Subsample   Observations    Subsample  Observations
AB_SPLIT =   0       448
----------------------------------------------------
Actual subsamples may be smaller if missing values
are being bypassed.  Subsamples with 0 observations
will be bypassed.
-----------------------------------------------------
Setting up an iteration over the values of AB_SPLIT
The model command will be executed for     1 values
of this variable.  In the current sample of     840
observations, the following counts were found:
Subsample   Observations    Subsample  Observations
AB_SPLIT =   1       392
----------------------------------------------------
-----------------------------------------------------------------
Subsample analyzed for this command is AB_SPLIT =       1
-----------------------------------------------------------------
--> CALC                 ; List ; hs_stat = 2*(logl1 - logl)
                         ; cvalue = ctb(.95,kgamma) $
[CALC] HS_STAT =     69.9219965
[CALC] CVALUE  =     11.0704978
Calculator: Computed   2 scalar results
```

N21.4.3 Lagrange Multiplier, Wald, and Likelihood Ratio Tests

NLOGIT 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 **NLOGIT** command to set up Wald tests of linear restrictions on the parameters. In general, the names are constructed during estimation, so it may be necessary to estimate the model without restrictions to determine what compound names are being used for the parameters. The example below shows a test of the hypothesis that the income coefficients in the air and train utility functions are the same. The names are constructed by the program, so it is necessary to fit the model first without restriction to determine the names to use in the restriction.

```
NLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = gc,ttme ; Rh2 = one,hinc
          ; Test: air_hin1 - tra_hin2 $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -189.52515
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    395.1 AIC/N =    1.881
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3321 .3235
Chi-squared[ 5]          =    188.46723
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
Wald test of  1 linear restrictions <---
Chi-squared =  12.07, P value =  .00051
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient     Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
      GC|   -.01093**      .00459    -2.38  .0172     -.01992   -.00194
    TTME|   -.09546***     .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|   5.87481***     .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|   -.00537        .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|   5.54986***     .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|   -.05656***     .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|   4.13028***     .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|   -.02858*       .01544    -1.85  .0642     -.05885    .00169
--------+--------------------------------------------------------------------
```

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, *NLOGIT* 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: *NLOGIT* 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 *NLOGIT* 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.

N22: Simulating Probabilities in Discrete Choice Models

N22.1 Introduction

The simulation program described here allows you to fit a model, use it to predict the set of choices for your sample, then examine how those choices would change if the attributes of the choices changed. You can also examine scenarios that involve restricting the choice set from the original one. Finally, you can use your estimated model and this simulator to do these analyses with data sets that were not actually used to fit the model. The calculation proceeds as follows:

Step 1. Set the desired sample for the model estimation. Estimate the model using *NLOGIT*. This processor is supported for the following discrete choice models that are specific to *NLOGIT*:

Model	Command	Alternative Command
Conditional Logit	**CLOGIT**	**NLOGIT**
Best/Worst Multinomial Logit	**BWMNL**	**CLOGIT ; BestWorst**
Best/Worst RP Logit	**BWRPL**	**RPLOGIT ; BestWorst**
Scaled Multinomial Logit	**SMNLOGIT**	**NLOGIT ; SMNL**
Random Regret Logit	**RRLOGIT**	**NLOGIT ; RRM**
RP Random Regret logit	**RPRRLOGIT**	**RPLOGIT ; RRM**
Error Components Logit	**ECLOGIT**	**NLOGIT ; ECM = ...**
Heteroscedastic Extreme Value	**HLOGIT**	**NLOGIT ; HEV**
Nested Logit	**NLOGIT**	**NLOGIT ; Tree = ...**
Generalized Nested Logit	**GNLOGIT**	**NLOGIT ; GNL**
Random Parameters Logit	**RPLOGIT**	**NLOGIT ; RPL**
Generalized Mixed Logit	**GMXLOGIT**	**NLOGIT ; GMXL**
Nonlinear Random Par	**NLRPLOGIT**	none
Latent Class Logit	**LCLOGIT**	**NLOGIT ; LCM**
Latent Class RP Logit	**LCRPLOGIT**	none
Latent Class RR logit	**LCRRLOGIT**	**LCLOGIT ; RRM**
Multinomial Probit	**MNPROBIT**	**NLOGIT ; MNP**

Step 2. The model is viewed as a random utility model in which the utility functions are functions of attributes $x_1,...,x_K$. The model is then fit to describe the choice among J alternatives, $C_1,...,C_J$. This may be a very simple model such as the basic multinomial logit model (MNL) of Chapter N16 or as complicated as a four level nested logit model as described in Chapter N28. In any event, the model is ultimately viewed in terms of these attributes and choices.

Step 3. (If desired) Reset the sample to any desired setup that is consistent with the model. This may be all or a subset of the data used to fit the model, or a set of individuals that were not used in fitting the model, or any mixture of the two.

Step 4. Specify which of the choices (possibly but not necessarily all) are to be used as the choice set for the simulation. The simulation is then produced to predict choice among this possibly reduced set of choices. (Probabilities for the full choice set are reallocated, but not necessarily proportionally. This would only occur in the MNL model which satisfies IIA.)

Step 5. Specify how the attributes that enter the utility functions will change – for example that a particular price is to rise by 25%.

Step 6. Simulate the model by computing the probabilities and predicting the outcomes for the specified sample and summarize the results, comparing them to the original, base case.

Steps 3-6 may be repeated as many times as desired once a model has been estimated. The model is not reestimated; the existing model is used to compute the simulation results. The simulation produces an output table that compares absolute frequencies and shares for each alternative in the full or a restricted choice set to the base case in which the predicted shares are the means of the sample predictions from the model absent the changes specified in the scenario.

In addition, this feature provides a capability for implementing simulation/scenario analysis when one is using mixtures of data (for example stated preference and revealed preference). This option allows you to combine the two types of data in a simulation. An example is shown in the case study below.

N22.2 Essential Subcommands

NLOGIT's models are all built around the specification which indicates the choice set being modeled:

; Choices = the full list of alternatives in the model

This simulation program is used to compute simulated probabilities assuming that the individuals in the sample being simulated are choosing among some or all of these alternatives. The first subcommand for the simulation is

; Simulation = a list of names of alternatives

The list of names must be some or all of the names in the **; Choices** list. If they are to be all of them, then you may use

; Simulation = * (or, just ; Simulation)

NOTE: Simulation on a subset of alternatives in the full choice set is done by analyzing the full set of data while, in process, pretending (simulating) that alternatives not in the simulation list are not available to these individuals even if they are physically in the data set and actually available. (Note, this is just for the purposes of the simulation.) You must not change the sample settings in any way to produce this effect yourself. It is handled completely internally by this program simply by using a set of switches ('on' for included, 'off' for excluded) for the choice set while numerical results are computed.

The second specification you will provide is the name of the attribute that is being set or changed and the names of the alternatives in which this attribute is changing. This is the 'scenario.' The base case, for a single changing attribute is

; Scenario: attribute name (list of alternatives whose attribute levels will change)
= [action] magnitude of action

If you wish to include in the scenario, all the alternatives that are defined in the simulation, simply use the wildcard character, * as the list. Note that this 'all items in list' refers back to your **; Simulation** list, not to the **; Choices** list. The actions in the scenario specification are as follows:

	=	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,
or	=	[/] value to divide the attribute by the specified value,
or	=	[-] value to subtract 'value' from the observed values
or	=	{*} value to change all observed values to this value.

The following example:

; Choices = air,train,bus,car
; Simulation = air,car
; Scenario: gc(car) = [*] 1.5

specifies a simulation over two choices in a four choice model. The scenario is enacted by changing the *gc* attribute for *car* only by multiplying whatever value is found in the original sample by 1.5. Alternatively,

; Scenario : gc(car) = {*} 100

compares the outcome actually observed (the base case) to a scenario in which *gc* for *car* is 100 for all observations.

N22.3 Multiple Attribute Specifications and Scenarios

The simulation may specify that more than one attribute is to change. The multiple settings may provide for changes in different alternatives. The specification is

; Scenario: attribute name 1 (list of alternatives) = [action] magnitude of action /
attribute name 2 (list of alternatives) = [action] magnitude of action /
... repeated up to a maximum of 20 attributes specifications

The different change specifications are separated by slashes. To continue the earlier example, we might specify

; Choices = air,train,bus,car
; Simulation = air,train, car
; Scenario: gc(car) = [*] 1.5 /
ttme (air,train) = [*] 1.25

You may also provide more than one full scenario for the simulation. In this case, each scenario is compared to the base case, then the scenarios are compared to each other. You may compare up to five scenarios in one run with this tool. Use

; Scenario: attribute name 1 (list of alternatives) = [action] magnitude of action ...
&
attribute name 2 (list of alternatives) = [action] magnitude of action ...

Use ampersands (&) to separate the scenarios. Within each scenario, you may have up to 20 attribute specifications separated by slashes.

N22.4 Simulation Commands

The simulation instruction does not produce new model estimates. However all other **NLOGIT** options can be invoked with the command, such as descriptive statistics and computing and retaining predicted probabilities.

N22.4.1 Observations Used for the Simulations

The data set used in the simulation can be the original data set used to estimate the model or a new data set. The base model is fit with an 'estimation' data set. After this operation (Steps 1 and 2 in the introduction), if desired, you may respecify the sample to direct the simulator to do the calculations with a completely different set of observations. This would precede Step 4 above. If you do not change the sample setting, the same data are used for the simulation. (The simulation must follow the estimation. In any case, it will require a second command, which will generally be identical to the first save for the specification of the simulation.)

N22.4.2 Variables Used for the Simulations

If a new data set is used, the attributes must have the exact same names and measurement units and the alternatives must also have the same names as the full or a restricted set of those used in model estimation. A natural application that would obey this convention would be to use one half of a sample to estimate the model, then repeat the simulation using the other half of the same sample.

N22.4.3 Choices Simulated

One can undertake simulation either on the full choice set used in estimation or a restricted set. This latter option is very useful for modelers using mixtures of data (e.g., combined stated and revealed preference data), where some alternatives are only included in estimation but not in application. An extensive example is shown below in the case study.

N22.4.4 Other *NLOGIT* Options

The routine that does simulation also allows you to compute the various elasticities and/or derivatives (**; Effects: ...**) and descriptive statistics (**; Describe** and **; Crosstab**) as described in Chapter N19, and will produce the standard results for these. You might already have done this at the estimation step, but if you change the sample as described in Section N22.4.1, you can use this simulation program to recompute those values.

N22.4.5 Observations Used for the Simulations

This program also allows you to compute, display, and save fitted probabilities, utilities and inclusive values for specific observations, using the standard setup for these as described in the *LIMDEP* documentation. Once again, this is likely to be useful when your estimation and simulation steps are based on different sets of observations.

N22.5 Arc Elasticities

Since the simulated scenarios produce discrete changes in the probabilities from discrete changes in attributes, it is convenient to compute arc elasticities using the results. You can request estimates of arc elasticities in **; Simulation** by adding

; Arc

to the command. Like point elasticities, these be computed either unweighted or probability weighted by adding

; Pwt

to the command. The following results are produced by adding **; Arc** to the application at the beginning of the next section:

```
-----------------------------------------------------------------------------
Estimated Arc Elasticities Based on the Specified Scenario. Rows in the table
report 0.00 if the indicated attribute did not change in the scenario or if
the average probability or average attribute was zero in the sample.
Estimated values are averaged over all individuals used in the simulation.
Rows of the table in which no changes took place are not shown.
-----------------------------------------------------------------------------
Attr Changed in | Change in Probability of Alternative
-----------------------------------------------------------------------------
Choice AIR      | AIR       TRAIN     BUS       CAR
   x = TTME     |   -3.003     2.948     2.948    -9.000
-----------------------------------------------------------------------------
```

N22.6 Plotting Simulated Choice Probabilities

You can plot the set of (up to 5) choice probabilities as they vary over a specified scenario. The syntax is

; attribute (choice) = [Plot] lower limit (interval) upper limit.

For example, to plot the choice probabilities in a simple multinomial logit model, we used

```
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,invc,invt,ttme,one $
CLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,invc,invt,ttme,one
          ; Simulate ; Scenario: invc(car) = [Plot] 20(5)75 $
```

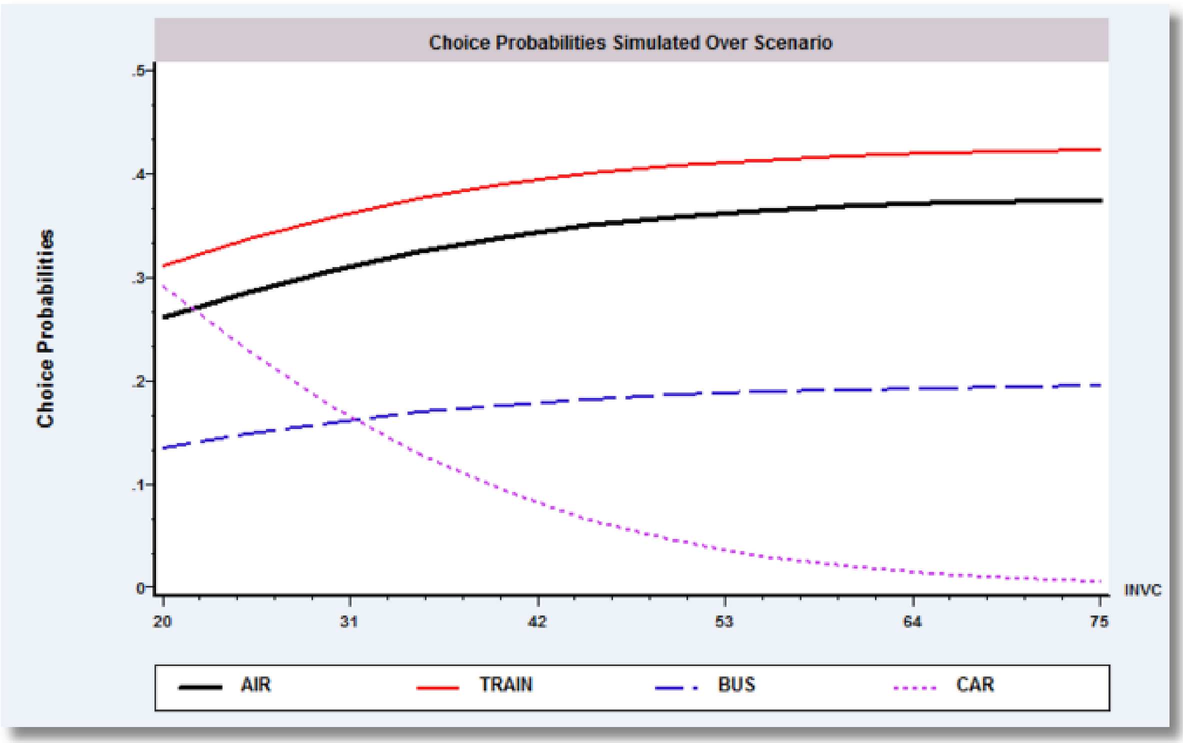


Figure N22.1 Simulated Choice Probabilities

N22.7 Applications

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

$$S(alternative\ j) = N \times \sum_{i=1}^{N} \hat{P}_{ij}$$

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,

NLOGIT ; 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 $
NLOGIT     ; Lhs = mode ; Rhs = one,gc,ttme
           ; Choices = air,train,bus,car $
NLOGIT     ; 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 $
NLOGIT     ; Lhs = mode ; Rhs = one,gc,ttme
           ; Choices = air,train,bus,car $
NLOGIT     ; 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,

```
NLOGIT    ; 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 %Share	Base Number	Scenario %Share	Scenario Number	Scenario - Base ChgShare	Scenario - Base 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 %Share	Base Number	Scenario %Share	Scenario Number	Scenario - Base ChgShare	Scenario - Base 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 %Share	Base Number	Scenario %Share	Scenario Number	Scenario - Base ChgShare	Scenario - Base 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 $
NLOGIT     ; Lhs  = mode ; Rhs = one,gc,ttme
           ; Choices = air,train,bus,car $
NLOGIT     ; 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 |
+----------+--------------+--------------+------------------+
```

N22.8 A Case Study

The data set used to illustrate the application of simulation/scenario analysis is a combined RP-SP data set associated with single-vehicle households choosing among vehicle types. The RP data are a single observation per household and involved choosing among 12 vehicle classes (*mc*,*sm*,*md*,*ua*,*ub*,*lg*,*lx*,*lc*,*fd*,*lt*), all of which are vehicles fueled by conventional fuels (i.e. gasoline and diesel). The SP data are three observations per household, often called treatments or choice sets. These observations are correlated and so it is preferable to run a model such as mixed logit (RPL) to allow for choice set correlation. We have done this in Hensher and Greene (2003), but in the example below we have used the simple multinomial logit form. The SP data set involved households choosing among four conventionally fueled vehicle (*c*1,*c*2,*c*3,*c*4), four electric vehicles (*e*1,*e*2,*e*3,*e*4) and four alternatively fueled vehicles (*a*1,*a*2,*a*3,*a*4). The case study involves running a number of scenarios in which we are interested in only the four electric vehicles, the four alternative fueled vehicles and the 12 conventional fueled vehicles. The reason for excluding *c*1-*c*4 is that they are equivalent to the RP alternatives and are only used to establish more robust parameter estimates in the SP data set that can be used to enrich the RP estimates. See Hensher and Greene for more details.

The initial data setup proceeds as follows, where mnemonics for the variables are suggestive of their content.

```
READ      ; File="C:\projects\ggedata\vehtype\sprp1data\sprp1.txt"
          ; Nvar = 24 ; Nobs = 14120
          ; Names = id,chosen,cset,altz,hweight,price,princ,opcost,rg,ls,
                    lage,acc,ncylind,encap,yr2,yr5,yr10,elec,accevaf,
                    bsize,range,small,altfuel,vexper $
CREATE    ; If(ncylind>0) rpobs=1 ? defining RP vs SP observations by # cyls. > 0
          ; If(rpobs=1 & altz=1)altz=13 ; If(rpobs=1 & altz=2)altz=14
          ; If(rpobs=1 & altz=3)altz=15 ; If(rpobs=1 & altz=4)altz=16
          ; If(rpobs=1 & altz=5)altz=17 ; If(rpobs=1 & altz=6)altz=18
          ; If(rpobs=1 & altz=7)altz=19 ; If(rpobs=1 & altz=8)altz=20
          ; If(rpobs=1 & altz=9)altz=21 ; If(rpobs=1 & altz=10)altz=22
          ; If(altz>12)cset=10 ; If(altz<13)sp=1
          ; pricea = price/1000
          ; hinc = princ/price ; hincn = hinc*1000000
          ; priccalc = princ/hinc ; pricez = -price
          ; opcostz = opcost
          ; If(rpobs=0)pdsz=3 ; If(rpobs=1)pdsz=1 $
DSTAT     ; Rhs = * $
```

Descriptive Statistics
All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
ID	2756.73399	1292.93923	1006.00000	6501.00000	14120
CHOSEN	.089164306	.284990850	.000000000	1.00000000	14120
CSET	11.3002833	.953883844	10.0000000	12.0000000	14120
ALTZ	10.3484419	6.17728995	1.00000000	22.0000000	14120
HWEIGHT	1.22935552	.326829180	.553000000	2.27400000	14120
PRICE	23239.6936	18854.2894	1552.00000	110000.000	14120
PRINC	947.053091	1094.13538	7.08600000	15400.0000	14120
OPCOST	8.52293909	4.28526201	2.00000000	24.0000000	14120
RG	218.160411	98.4963414	50.0000000	550.000000	14120
LS	.321991723	.808399213	.000000000	6.13000000	14120
LAGE	.647481834	1.03715689	.000000000	2.77259000	14120
ACC	5.64133853	8.04051105	.000000000	23.1000000	14120
NCYLIND	1.89123938	2.68257443	.000000000	8.00000000	14120
ENCAP	932.266218	1408.75909	.000000000	4994.00000	14120
YR2	.160764873	.367326945	.000000000	1.00000000	14120
YR5	.161685552	.368175141	.000000000	1.00000000	14120
YR10	.157365439	.364157863	.000000000	1.00000000	14120
ELEC	.216713881	.412020628	.000000000	1.00000000	14120
ACCEVAF	10.9891289	9.19275986	.000000000	29.0000000	14120
BSIZE	.345835694	.286489598	.000000000	.750000000	14120
RANGE	321.076416	248.536479	.000000000	580.000000	14120
SMALL	.219688385	.414050166	.000000000	1.00000000	14120
ALTFUEL	.216713881	.412020628	.000000000	1.00000000	14120
VEXPER	1.30028329	1.15903100	.000000000	3.00000000	14120
RPOBS	.349858357	.476941922	.000000000	1.00000000	14120
SP	.650141643	.476941922	.000000000	1.00000000	14120
PRICEA	23.2396936	18.8542894	1.55200000	110.000000	14120
HINC	.040000427	.024639760	.003000000	.140000000	14120
PRICCALC	23239.6936	18854.2894	1552.00000	110000.000	14120
HINCN	40000.4275	24639.7600	3000.00000	140000.000	14120
PRICEZ	-23239.6936	18854.2894	-110000.000	-1552.00000	14120
OPCOSTZ	-8.52293909	4.28526201	-24.0000000	-2.00000000	14120
PDSZ	2.30028329	.953883844	1.00000000	3.00000000	14120

N22.8.1 Base Model – Multinomial Logit (MNL)

The base model is a MNL model with a fairly complicated set of utility functions.

```
NLOGIT ; Lhs = chosen,cset,altz
       ; Choices = c1,c2,c3,c4,e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
       ; Model: U(c1,c2,c3,c4) = prc*pricez + pic*princ + opc*opcost
                   + y2*yr2 + y5*yr5 + y10*yr10/
              U(e1,e2,e3,e4) = el*elec + prc*pricez + pic*princ + opc*opcost
                   + accev*accevaf + rangevaf*range + smev*small
                   + y2*yr2 + y5*yr5 + y10*yr10/
              U(a1,a2,a3,a4) = af*altfuel + prc*pricez + pic*princ
                   + opc*opcost + rangevaf*range + smaf*small
                   + y2*yr2 + y5*yr 5 + y10*yr10/
              U(mc,sm,md,ua,ub,lg,lx,lc,fd,lt)=<mc,sm,md,ua,ub,lg,lx,lc,fd,0>
                   + prc*pricez + pic*princ + opc*opcost + ag*lage + ac*acc
                   + <ncy4,ncy4,ncy4,0,0,0,0,0,0,0>*ncylinds $
```

```
Normal exit from iterations. Exit status=0.

-------------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Maximum Likelihood Estimates
Dependent variable               Choice
Number of observations             1259
Iterations completed                  6
Log likelihood function       -2636.317
Log-L for Choice   model =   -2636.3166
R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj
No coefficients  -3891.6224  .32257  .32130
Constants only.  Must be computed directly.
                 Use NLOGIT ;...; RHS=ONE $
Response data are given as ind. choice.
Number of obs.=  1259, skipped   0 bad obs.
-------------------------------------------------------------------------------
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
-------------------------------------------------------------------------------
 PRC       .7222769180E-04  .59254181E-05   12.189   .0000
 PIC       .5707622560E-03  .79760574E-04    7.156   .0000
 OPC      -.2789975405E-01  .10864813E-01   -2.568   .0102
 Y2         -.8517427857       .10381185   -8.205   .0000
 Y5         -1.133299963       .11084551  -10.224   .0000
 Y10        -2.019371339       .13924674  -14.502   .0000
 EL          .2578529016       .34663340     .744   .4570
 ACCEV    -.3375590631E-01  .13000471E-01   -2.597   .0094
 RANGEVAF  .1543507538E-02  .58049882E-03    2.659   .0078
 SMEV       -.1436671461       .14858417    -.967   .3336
 AF         -.2122986044       .30794448    -.689   .4906
 SMAF       -.5259584270       .13516611   -3.891   .0001
 MC         -4.715718940       2.6638968   -1.770   .0767
 SM         -4.415489351       2.9062069   -1.519   .1287
 MD         -4.425306017       2.9075390   -1.522   .1280
 UA          .2883292576       .73695696     .391   .6956
 UB          1.116433455       .76581936    1.458   .1449
 LG         -.5133101006       .85317833    -.602   .5474
 LX       -.6684748282E-01     .86242554    -.078   .9382
 LC          1.342303824       .44512975    3.016   .0026
 FD          .8089115596       .46962284    1.722   .0850
 AG         -.2728581631   .80614539E-01   -3.385   .0007
 AC         -.1817342201   .76184184E-01   -2.385   .0171
 NCY4        1.518454640       .71367281    2.128   .0334
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

N22.8.2 Scenarios

We now simulate the model, using several different specifications for different scenarios.

```
NLOGIT    ; Lhs = chosen,cset,altz
          ; Choices = c1,c2,c3,c4,e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
          ; Model: ... exactly as above ...
          ; Simulation = *
          ; MergeSPRP (id = id, type = vexper)
```

These are added to the command above and the command is terminated after the setup:

Scenario 1. Increase prices by 50% for *mc* to *lt*.

```
; Scenario:    pricez(mc,sm,md,ua,ub,lg,lx,lc,fd,lt)   = [*]  1.5 /
               princ(mc,sm,md,ua,ub,lg,lx,lc,fd,lt)    = [*]  1.5
```

Scenario 2. For the second case, we exclude *c*1 - *c*4.

```
; Simulation  = e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
```

Scenario 3. Increase prices by 50% for *e*1, *e*2, *e*3, *e*4.

```
; Simulation  = e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
; Scenario:    pricez(e1,e2,e3,e4)                     = [*]  1.5 /
               princ(e1,e2,e3,e4)                      = [*]  1.5
```

Scenario 4. Reduce prices by 50% for *e*1, *e*2, *e*3, *e*4 and increase price by 50% for *mc* to *lt*.

```
; Simulation  = e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
; Scenario:    pricez(e1,e2,e3,e4)                     = [*]  0.5 /
               princ(e1,e2,e3,e4)                      = [*]  0.5
               &
               pricez(mc,sm,md,ua,ub,lg,lx,lc,fd,lt)   = [*]  1.5  /
               princ(mc,sm,md,ua,ub,lg,lx,lc,fd,lt)    = [*]  0.5
```

Scenario 5. Increase acceleration by 50% for *e*1,*e*2,*e*3,*e*4.

```
; Simulation  = e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
; Scenario:    accevaf(e1,e2,e3,e4)                    = [*]  1.5
```

Scenario 6. Make *yr*2, *yr*5 and *yr*10 take on fixed values for *e*1,*e*2,*e*3,*e*4, *a*1,*a*2,*a*3,a4.

```
; Simulation  = e1,e2,e3,e4,a1,a2,a3,a4,mc,sm,md,ua,ub,lg,lx,lc,fd,lt
; Scenario:    yr2(e1,e2,e3,e4,a1,a2,a3,a4)            =  0.5/
               yr5(e1,e2,e3,e4, a1,a2,a3,a4)           =  0.25/
               yr10(e1,e2,e3,e4, a1,a2,a3,a4)          =  0.25
```

Scenario 1 – All Alternatives

```
+------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+------------------------------------------------------+
```

Specification of scenario 1 is:

Attribute	Alternatives affected				Change type	Value
PRICEZ	MC	SM	MD	more	Scale base by value	1.500
PRINC	MC	SM	MD	more	Scale base by value	1.500

```
+-------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+-------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+-------------------------------------------------------+
```

Simulated Probabilities (shares) for this scenario:

Choice	Base %Share	Base Number	Scenario %Share	Scenario Number	Scenario - Base ChgShare	Scenario - Base ChgNumber
C1	8.973	67	9.439	70	.465%	3
C2	5.083	38	5.349	40	.266%	2
C3	3.817	28	4.010	30	.192%	2
C4	2.730	20	2.870	21	.141%	1
E1	9.456	70	9.931	74	.475%	4
E2	6.772	50	7.103	53	.332%	3
E3	4.800	36	5.029	37	.228%	1
E4	3.549	26	3.718	28	.170%	2
A1	10.189	76	10.708	80	.519%	4
A2	7.928	59	8.332	62	.404%	3
A3	7.189	53	7.551	56	.363%	3
A4	5.564	41	5.840	43	.277%	2
MC	1.826	14	1.645	12	-.181%	-2
SM	6.498	48	5.591	42	-.907%	-6
MD	5.583	42	4.617	34	-.967%	-8
UA	1.603	12	1.305	10	-.298%	-2
UB	5.077	38	4.258	32	-.819%	-6
LG	.838	6	.683	5	-.155%	-1
LX	.392	3	.210	2	-.182%	-1
LC	1.164	9	1.025	8	-.138%	-1
FD	.634	5	.500	4	-.134%	-1
LT	.335	2	.285	2	-.050%	0
Total	100.000	743	100.000	745	.000%	2

Scenario 2 – Excluding Alternatives C1-C4

```
+------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
PRICEZ     MC       SM       MD       more  Scale base by value     1.500
PRINC      MC       SM       MD       more  Scale base by value     1.500
-------------------------------------------------------------------------
+-------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+-------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+-------------------------------------------------------+
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|E1        | 11.836    88 | 12.599    94 |    .763%       6 |
|E2        |  8.470    63 |  9.001    67 |    .532%       4 |
|E3        |  5.979    44 |  6.341    47 |    .362%       3 |
|E4        |  4.412    33 |  4.680    35 |    .268%       2 |
|A1        | 12.831    95 | 13.679   102 |    .848%       7 |
|A2        | 10.010    74 | 10.673    79 |    .663%       5 |
|A3        |  9.012    67 |  9.594    71 |    .582%       4 |
|A4        |  6.961    52 |  7.404    55 |    .443%       3 |
|MC        |  2.320    17 |  2.123    16 |  -.197%      -1 |
|SM        |  8.269    62 |  7.229    54 | -1.039%      -8 |
|MD        |  7.116    53 |  5.983    45 | -1.133%      -8 |
|UA        |  2.040    15 |  1.687    13 |  -.353%      -2 |
|UB        |  6.460    48 |  5.506    41 |  -.954%      -7 |
|LG        |  1.069     8 |   .886     7 |  -.182%      -1 |
|LX        |   .499     4 |   .270     2 |  -.229%      -2 |
|LC        |  1.482    11 |  1.326    10 |  -.156%      -1 |
|FD        |   .808     6 |   .649     5 |  -.160%      -1 |
|LT        |   .426     3 |   .368     3 |  -.058%       0 |
|Total     |100.000   743 |100.000   746 |   .000%       3 |
+----------+--------------+--------------+------------------+
```

Scenario 3

```
+------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+------------------------------------------------------+
------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type            Value
---------  ------------------------------  ------------------- ---------
PRICEZ     E1       E2       E3       more  Scale base by value     1.500
PRINC      E1       E2       E3       more  Scale base by value     1.500
------------------------------------------------------------------------
+-------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+-------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+-------------------------------------------------------+
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|E1        | 11.836    88 |  8.419    63 | -3.417%      -25 |
|E2        |  8.470    63 |  5.932    44 | -2.538%      -19 |
|E3        |  5.979    44 |  3.916    29 | -2.063%      -15 |
|E4        |  4.412    33 |  2.895    22 | -1.517%      -11 |
|A1        | 12.831    95 | 14.563   108 |  1.732%       13 |
|A2        | 10.010    74 | 11.349    84 |  1.338%       10 |
|A3        |  9.012    67 | 10.189    76 |  1.177%        9 |
|A4        |  6.961    52 |  7.870    59 |   .908%        7 |
|MC        |  2.320    17 |  2.656    20 |   .336%        3 |
|SM        |  8.269    62 |  9.458    70 |  1.189%        8 |
|MD        |  7.116    53 |  8.140    61 |  1.024%        8 |
|UA        |  2.040    15 |  2.332    17 |   .292%        2 |
|UB        |  6.460    48 |  7.387    55 |   .927%        7 |
|LG        |  1.069     8 |  1.222     9 |   .153%        1 |
|LX        |   .499     4 |   .565     4 |   .066%        0 |
|LC        |  1.482    11 |  1.697    13 |   .215%        2 |
|FD        |   .808     6 |   .924     7 |   .115%        1 |
|LT        |   .426     3 |   .488     4 |   .062%        1 |
|Total     |100.000   743 |100.000   745 |   .000%        2 |
+----------+--------------+--------------+------------------+
```

Scenario 4

```
+------------------------------------------------+
| Discrete Choice (One Level) Model              |
| Model Simulation Using Previous Estimates      |
| Number of observations             1259        |
+------------------------------------------------+
+---------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+---------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
PRICEZ     E1       E2       E3       more  Scale base by value      .500
PRINC      E1       E2       E3       more  Scale base by value      .500
-------------------------------------------------------------------------
+---------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+---------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+---------------------------------------------------------+
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|E1        | 11.836    88 | 16.127   120 |  4.290%       32 |
|E2        |  8.470    63 | 12.072    90 |  3.602%       27 |
|E3        |  5.979    44 |  9.653    72 |  3.674%       28 |
|E4        |  4.412    33 |  7.144    53 |  2.732%       20 |
|A1        | 12.831    95 | 10.225    76 | -2.606%      -19 |
|A2        | 10.010    74 |  8.000    60 | -2.010%      -14 |
|A3        |  9.012    67 |  7.245    54 | -1.767%      -13 |
|A4        |  6.961    52 |  5.597    42 | -1.365%      -10 |
|MC        |  2.320    17 |  1.817    14 |  -.504%       -3 |
|SM        |  8.269    62 |  6.491    48 | -1.778%      -14 |
|MD        |  7.116    53 |  5.585    42 | -1.531%      -11 |
|UA        |  2.040    15 |  1.603    12 |  -.437%       -3 |
|UB        |  6.460    48 |  5.072    38 | -1.388%      -10 |
|LG        |  1.069     8 |   .838     6 |  -.231%       -2 |
|LX        |   .499     4 |   .402     3 |  -.097%       -1 |
|LC        |  1.482    11 |  1.160     9 |  -.322%       -2 |
|FD        |   .808     6 |   .636     5 |  -.173%       -1 |
|LT        |   .426     3 |   .334     2 |  -.092%       -1 |
|Total     |100.000   743 |100.000   746 |   .000%        3 |
+----------+--------------+--------------+------------------+
```

Specification of scenario 2 is:

Attribute	Alternatives affected				Change type	Value
PRICEZ	MC	SM	MD	more	Scale base by value	1.500
PRINC	MC	SM	MD	more	Scale base by value	.500

This scenario is based on merged RP and SP data sets
Simulated Probabilities (shares) for this scenario:

Choice	Base		Scenario		Scenario - Base	
	%Share	Number	%Share	Number	ChgShare	ChgNumber
E1	11.836	88	13.176	98	1.340%	10
E2	8.470	63	9.429	70	.960%	7
E3	5.979	44	6.638	49	.659%	5
E4	4.412	33	4.900	36	.488%	3
A1	12.831	95	14.298	106	1.468%	11
A2	10.010	74	11.173	83	1.162%	9
A3	9.012	67	10.050	75	1.038%	8
A4	6.961	52	7.745	58	.783%	6
MC	2.320	17	1.964	15	-.356%	-2
SM	8.269	62	6.367	47	-1.902%	-15
MD	7.116	53	5.130	38	-1.985%	-15
UA	2.040	15	1.442	11	-.597%	-4
UB	6.460	48	4.744	35	-1.716%	-13
LG	1.069	8	.758	6	-.311%	-2
LX	.499	4	.121	1	-.378%	-3
LC	1.482	11	1.206	9	-.276%	-2
FD	.808	6	.536	4	-.273%	-2
LT	.426	3	.322	2	-.104%	-1
Total	100.000	743	100.000	743	.000%	0

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
E1	16.127	120	13.176	98	-2.950%	-22
E2	12.072	90	9.429	70	-2.642%	-20
E3	9.653	72	6.638	49	-3.016%	-23
E4	7.144	53	4.900	36	-2.244%	-17
A1	10.225	76	14.298	106	4.073%	30
A2	8.000	60	11.173	83	3.173%	23
A3	7.245	54	10.050	75	2.805%	21
A4	5.597	42	7.745	58	2.148%	16
MC	1.817	14	1.964	15	.147%	1
SM	6.491	48	6.367	47	-.124%	-1
MD	5.585	42	5.130	38	-.454%	-4
UA	1.603	12	1.442	11	-.160%	-1
UB	5.072	38	4.744	35	-.328%	-3
LG	.838	6	.758	6	-.080%	0
LX	.402	3	.121	1	-.281%	-2
LC	1.160	9	1.206	9	.046%	0
FD	.636	5	.536	4	-.100%	-1
LT	.334	2	.322	2	-.012%	0
Total	100.000	746	100.000	743	.000%	-3

Scenario 5

```
+------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type             Value
---------  ------------------------------  ------------------- ---------
ACCEVAF    E1       E2       E3       more  Scale base by value     1.500
-------------------------------------------------------------------------
+-------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+-------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+-------------------------------------------------------+
Simulated Probabilities (shares) for this scenario:
```

Choice	Base		Scenario		Scenario - Base	
	%Share	Number	%Share	Number	ChgShare	ChgNumber
E1	11.836	88	9.434	70	-2.402%	-18
E2	8.470	63	6.858	51	-1.611%	-12
E3	5.979	44	4.919	37	-1.061%	-7
E4	4.412	33	3.686	27	-.726%	-6
A1	12.831	95	13.896	103	1.065%	8
A2	10.010	74	10.839	81	.828%	7
A3	9.012	67	9.757	73	.745%	6
A4	6.961	52	7.538	56	.577%	4
MC	2.320	17	2.516	19	.195%	2
SM	8.269	62	8.970	67	.701%	5
MD	7.116	53	7.719	57	.603%	4
UA	2.040	15	2.213	16	.173%	1
UB	6.460	48	7.008	52	.548%	4
LG	1.069	8	1.159	9	.090%	1
LX	.499	4	.543	4	.044%	0
LC	1.482	11	1.607	12	.126%	1
FD	.808	6	.877	7	.069%	1
LT	.426	3	.462	3	.036%	0
Total	100.000	743	100.000	744	.000%	1

Scenario 6

```
+------------------------------------------------------+
|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   1259 observations.|
|RP and SP data are merged for this set of simulations.|
+------------------------------------------------------+
```

Specification of scenario 1 is:

Attribute	Alternatives affected				Change type	Value
YR2	E1	E2	E3	more	Fix at new value	.500
YR5	E1	E2	E3	more	Fix at new value	.250
YR10	E1	E2	E3	more	Fix at new value	.250

```
+-------------------------------------------------------+
|REVEALED PREFERENCE (RP) / STATED PREFERENCE (SP) DATA |
+-------------------------------------------------------+
| This scenario is based on merged RP and SP data sets  |
| The sample contains    494 observations marked as RP. |
| Data search located    744 SP scenarios that matched  |
| IDs with an RP observation and     21 SP scenarios    |
| with IDs that did not match any RP observation in the |
| full sample of   1259 total observations. Any remain- |
| ing observations were erroneous or unusable.          |
+-------------------------------------------------------+
```

Simulated Probabilities (shares) for this scenario:

Choice	Base		Scenario		Scenario - Base	
	%Share	Number	%Share	Number	ChgShare	ChgNumber
E1	11.836	88	8.108	60	-3.728%	-28
E2	8.470	63	8.201	61	-.269%	-2
E3	5.979	44	6.318	47	.339%	3
E4	4.412	33	5.558	41	1.146%	8
A1	12.831	95	8.815	66	-4.015%	-29
A2	10.010	74	9.988	74	-.023%	0
A3	9.012	67	8.921	66	-.091%	-1
A4	6.961	52	8.903	66	1.942%	14
MC	2.320	17	2.672	20	.351%	3
SM	8.269	62	9.526	71	1.258%	9
MD	7.116	53	8.222	61	1.106%	8
UA	2.040	15	2.359	18	.319%	3
UB	6.460	48	7.454	55	.994%	7
LG	1.069	8	1.230	9	.161%	1
LX	.499	4	.596	4	.097%	0
LC	1.482	11	1.704	13	.222%	2
FD	.808	6	.936	7	.127%	1
LT	.426	3	.490	4	.064%	1
Total	100.000	743	100.000	743	.000%	0

N23: The Multinomial Logit and Random Regret Models

N23.1 Introduction

In the multinomial logit model described in Chapter N16, there is a single vector of characteristics, which describes the individual, and a set of J parameter vectors. In the 'discrete choice' setting of this section, these are essentially reversed. The J alternatives are each characterized by a set of K 'attributes,' $\mathbf{x}_{ij}$. Respondent 'i' chooses among the J alternatives. There is a single parameter vector, $\boldsymbol{\beta}$. The model underlying the observed data is assumed to be the following random utility specification:

$$U(\text{choice } j \text{ for individual } i) = U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \varepsilon_{ij}, j = 1,...,J_i.$$

The random, individual specific terms, $(\varepsilon_{i1},\varepsilon_{i2},...,\varepsilon_{iJ})$ are assumed to be independently distributed, each with an extreme value distribution. Under these assumptions, the probability that individual i chooses alternative j is

$$\text{Prob}(U_{ij} > U_{iq}) \text{ for all } q \neq j.$$

It has been shown that for independent extreme value distributions, as above, this probability is

$$\text{Prob}(y_i = j) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij}\right)}{\sum_{m=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{im}\right)}$$

where y_i is the index of the choice made. 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 logit model described in Chapter N16. It does, however, make the very strong 'Independence from Irrelevant Alternatives' assumption which will be discussed below.

NOTE: The distinction made here between 'discrete choice' and 'multinomial logit' is not hard and fast. It is made purely for convenience in the discussion. As noted in Chapters N16 and N17, by interacting the characteristics with the alternative specific constants, the discrete choice model of this chapter becomes the multinomial logit model of Chapter N16. From this point, in the remainder of this reference guide for *NLOGIT*, we will refer to the model described in this chapter, with mathematical formulation as given above, as the 'multinomial logit model,' or MNL model as is common in the literature.

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 we do 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 arguments 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. *NLOGIT* 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. *NLOGIT* will accommodate these cases as well. All these variations are discussed Chapter N18.

N23.2 Command for the Multinomial Logit Model

The simplest form of the command for the discrete choice models is

NLOGIT ; 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 characteristics $

(With no qualifiers to indicate a different model, such as RPL or MNP, **NLOGIT** and **CLOGIT** are the same.) There are various ways to specify the utility functions – i.e., the right hand sides of the equations that underlie the model, and several different ways to specify the choice set. These are discussed in Chapter N20. The **; Rhs** specification may be replaced with an explicit definition of the utility functions, using **; 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 command builder for this model is found in Model:Discrete Choice/Discrete Choice. The Main and Options pages are both used to set up the model. The model and the choice set are defined in the Main page; the attributes are defined in the Options page. See Figure N23.1.

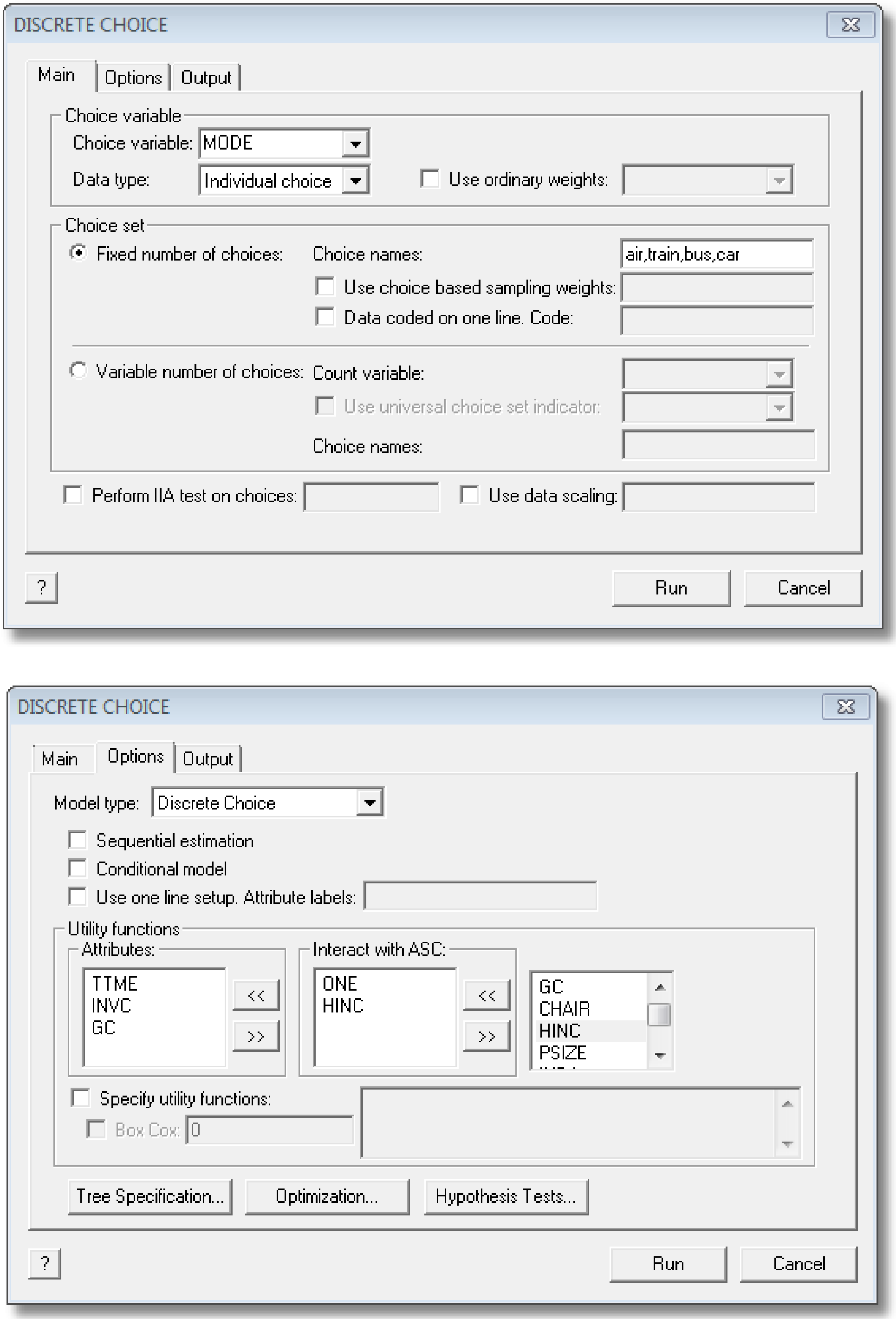


Figure N23.1 Command Builder for Multinomial Logit Model

N23.3 Results for the Multinomial Logit Model

Results for the multinomial logit model will consist of the standard model results and any additional descriptive output you have requested. The application below will display the full set of available results. 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.

In the *Last Model*, groups of coefficients for variables that are integrated 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            =     57.91928       |
| Prob. from Chi-squared[ 1] =      .00000       |
+-------------------------------------------------+
+---------+-------------+---------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+-------------+---------------+--------+---------+
 Fncn(1)      13.32858178       1.7513477    7.610   .0000
```

N23.4 Application

The MNL model based on the clogit data is estimated with the command

NLOGIT ; Lhs = mode
; Choices = air,train,bus,car
; Rhs = gc,ttme
; Rh2 = one,hinc
; Show Model
; Describe
; Crosstab
; Effects: gc(*)
; Full
; Ivb = incvlu
; Prob = pmnl
; List $

This requests all the optional output from the model. The **; Describe** specification detailed in Section N19.4.4 requests a set of descriptive statistics for the variables in the model, by choice. The leftmost set of results gives the coefficient estimates. Note that in this model, they are the same for the two generic coefficients, on *gc* and *ttme*, but they vary by choice for the alternative specific constant and its interaction with income. Also, since there is no ASC for *car* (it was dropped to avoid the dummy variable trap), there are no coefficients for the car grouping. The second set of values in the center section gives the mean and standard deviation for that attribute in that outcome for all observations in the sample. The third set of results gives the mean and variance for the particular attribute for the individuals that made that choice. The full set of results from the model is as follows. (The various parts of the output are described in Section N19.4.2.)

```
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| GC       TTME     A_AIR    AIR_HIN1 A_TRAIN    |
|        |Row  2| TRA_HIN2 A_BUS    BUS_HIN3                     |
+--------+------+------------------------------------------------+
|AIR     |     1| GC       TTME     Constant HINC     none       |
|        |     2| none     none     none                         |
|TRAIN   |     1| GC       TTME     none     none     Constant   |
|        |     2| HINC     none     none                         |
|BUS     |     1| GC       TTME     none     none     none       |
|        |     2| none     Constant HINC                         |
|CAR     |     1| GC       TTME     none     none     none       |
|        |     2| none     none     none                         |
+----------------------------------------------------------------+

Normal exit:   6 iterations. Status=0, F=    189.5252

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -189.52515
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    395.1 AIC/N =    1.881
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3321 .3235
Chi-squared[ 5]          =   188.46723
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
--------+--------------------------------------------------------------------
      GC|    -.01093**      .00459    -2.38  .0172     -.01992   -.00194
    TTME|    -.09546***     .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|    5.87481***     .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|    -.00537        .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|    5.54986***     .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|    -.05656***     .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|    4.13028***     .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|    -.02858*       .01544    -1.85  .0642     -.05885    .00169
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+---------------------------------------------------------------------------+
|              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.|
| -------------------   --------   | -------------------+------------------ |
| GC              -.0109  GC       |  102.648     30.575|  113.552    33.198 |
| TTME            -.0955  TTME     |   61.010     15.719|   46.534    24.389 |
| A_AIR           5.8748  ONE      |    1.000       .000|    1.000      .000 |
| AIR_HIN1        -.0054  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.|
| -------------------   --------   | -------------------+------------------ |
| GC              -.0109  GC       |  130.200     58.235|  106.619    49.601 |
| TTME            -.0955  TTME     |   35.690     12.279|   28.524    19.354 |
| A_TRAIN         5.5499  ONE      |    1.000       .000|    1.000      .000 |
| TRA_HIN2        -.0566  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.|
| -------------------   --------   | -------------------+------------------ |
| GC              -.0109  GC       |  115.257     44.934|  108.133    43.244 |
| TTME            -.0955  TTME     |   41.657     12.077|   25.200    14.919 |
| A_BUS           4.1303  ONE      |    1.000       .000|    1.000      .000 |
| BUS_HIN3        -.0286  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.|
| -------------------   --------   | -------------------+------------------ |
| GC              -.0109  GC       |   95.414     46.827|   89.085    49.833 |
| TTME            -.0955  TTME     |     .000       .000|     .000      .000 |
+---------------------------------------------------------------------------+
```

```
+------------------------------------------------------+
| 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|         33           7           4          14          58
   TRAIN|          7          39           5          12          63
     BUS|          3           6          15           6          30
     CAR|         15          11           6          27          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|         38           4           0          16          58
   TRAIN|          3          49           1          10          63
     BUS|          0           3          23           4          30
     CAR|          4          10           0          45          59
--------+-------------------------------------------------------------
   Total|         45          66          24          75         210

PREDICTED PROBABILITIES (* marks actual, + marks prediction.)
Indiv    AIR       TRAIN     BUS       CAR
    1   .0984     .3311     .1959     .3746*+
    2   .2566     .2262     .0530     .4641*+
    3   .1401     .1795     .1997     .4808*+
    4   .2732     .0297     .0211     .6759*+
    5   .3421     .1478     .0527     .4575*+
    6   .0831     .3962*+   .2673     .2534
    7   .6066*+   .0701     .0898     .2335
    8   .0626     .6059 +   .1925     .1390*
    9   .1125     .2932     .1995     .3947*+
   10   .1482     .0804     .1267     .6447*+
```

(Rows 11-210 are omitted.)

```
+---------------------------------------------------+
| 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|    -.80189***      .02645   -30.31  .0000     -.85374   -.75004
   TRAIN|     .31977***      .02326    13.75  .0000      .27419    .36536
     BUS|     .31977***      .02326    13.75  .0000      .27419    .36536
     CAR|     .31977***      .02326    13.75  .0000      .27419    .36536
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AIR|     .35343***      .02423    14.59  .0000      .30595    .40091
   TRAIN|   -1.06931***      .04923   -21.72  .0000    -1.16580   -.97282
     BUS|     .35343***      .02423    14.59  .0000      .30595    .40091
     CAR|     .35343***      .02423    14.59  .0000      .30595    .40091
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AIR|     .16787***      .01593    10.54  .0000      .13666    .19908
   TRAIN|     .16787***      .01593    10.54  .0000      .13666    .19908
     BUS|   -1.09159***      .03576   -30.52  .0000    -1.16168  -1.02149
     CAR|     .16787***      .01593    10.54  .0000      .13666    .19908
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in CAR
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AIR|     .29344***      .01845    15.90  .0000      .25727    .32961
   TRAIN|     .29344***      .01845    15.90  .0000      .25727    .32961
     BUS|     .29344***      .01845    15.90  .0000      .25727    .32961
     CAR|    -.74918***      .03057   -24.51  .0000     -.80909   -.68927
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |      AIR    TRAIN       BUS      CAR
--------+------------------------------------
     AIR|   -.8019    .3198     .3198    .3198
   TRAIN|    .3534  -1.0693     .3534    .3534
     BUS|    .1679    .1679   -1.0916    .1679
     CAR|    .2934    .2934     .2934   -.7492
```

N23.5 Partial Effects

We define the partial effects in the multinomial logit model as the derivatives of the probability of choice *j* with respect to attribute *k* in alternative *m*. This is

$$\frac{\partial P_j}{\partial x_{km}} = [\mathbf{1}(j=m) - P_m] P_j \beta_k ,$$

where the function $\mathbf{1}(j = m)$ equals one if *j* equals *m* and zero otherwise. These are naturally scaled since the probability is bounded. They are usually very small, so *NLOGIT* reports 100 times the value obtained, as in the example below, which is produced by

; Effects: gc[air]
; Full

```
+---------------------------------------------------+
| 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|    -.00134***   .6076D-04   -22.04  .0000     -.00146   -.00122
   TRAIN|     .00036***   .2132D-04    16.98  .0000      .00032    .00040
     BUS|     .00020***   .1406D-04    14.48  .0000      .00018    .00023
     CAR|     .00077***   .5266D-04    14.69  .0000      .00067    .00088
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Derivative wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN       BUS      CAR
--------+----------------------------------
     AIR|  -.0013    .0004     .0002    .0008
```

Derivatives and elasticities are obtained by averaging the observation specific values, rather than by computing them at the sample means. The listing reports the sample mean (average partial effect) and the sample standard deviation. Alternative approaches are discussed in Section N21.2.

It is common to report elasticities rather than the derivatives. These are

$$\frac{\partial \log P_j}{\partial \log x_{km}} = [\mathbf{1}(j=m) - P_m] x_{km} \beta_k .$$

The example below shows the counterpart to the preceding results produced by

> **; Effects: gc(air)**
> **; Full**

which requests a table of elasticities for the effect of changing *gc* in the *air* alternative.

```
+---------------------------------------------------+
| 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|    -.80189***      .02645   -30.31  .0000     -.85374   -.75004
   TRAIN|     .31977***      .02326    13.75  .0000      .27419    .36536
     BUS|     .31977***      .02326    13.75  .0000      .27419    .36536
     CAR|     .31977***      .02326    13.75  .0000      .27419    .36536
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+------------------------------------
     AIR|  -.8019    .3198    .3198    .3198
```

The difference between the two commands is the use of '[*air*]' for derivatives and '(*air*)' for elasticities. The full set of tables, one for each alternative, is requested with

alternative[*]

or **alternative(*)**.

Note that for this model, the elasticities take only two values, the 'own' value when *j* equals *m* and the 'cross' elasticity when *j* is not equal to *m*. The fact that the cross elasticities are all the same is one of the undesirable consequences of the IIA property of this model.

N23.6 Technical Details on Maximum Likelihood Estimation

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_{ij} = 1 if individual i makes choice j and 0 otherwise,

$$P_{ij} = \text{Prob}(y_i = j) = \text{Prob}(d_{ij} = 1) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ij})}{\sum_{m=1}^{J_i}\exp(\boldsymbol{\beta}'\mathbf{x}_{ij})},$$

$$\text{Log } L = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ij}\log P_{ij},$$

$$\bar{\mathbf{x}}_i = \sum_{j=1}^{J_i} P_{ij}\mathbf{x}_{ij},$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} d_{ij}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i),$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum_{i=1}^{n}\sum_{j=1}^{J_i} P_{ij}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'.$$

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 or (we have observed) with some particular data sets. 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 are such as to make estimation impossible, you should be able to estimate the model by using

; Alg = BFGS

as an alternative. If this method fails as well, you should conclude that your model is inestimable. Section N19.5 describes a constrained estimator that is computed to calibrate the parameters to a model computed previously. Newton's method is very sensitive to this exercise – it frequently breaks down when parameters are fixed in this fashion. In this case, *NLOGIT* automatically switches to the BFGS method. This is one of the effects of the **; Calibrate** specification.

You may provide your own starting values for the iterations 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. If you do not have alternative specific constants in the model (with **; Rh2 = one**), then the parameters will appear in the same order as the Rhs variables. If you have alternative specific constant terms but you have no other Rh2 variables, then regardless of where *one* appears in the Rhs list, the ASCs will be the last J-1 coefficients corresponding to that list.

For example, in our earlier application, if the model were specified with **; Rhs = gc,one,ttme**, then the following final arrangement of the parameters would result, and it is this order in which you would provide the starting values:

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence 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

If you have other Rh2 variables, the coefficients will be interleaved with the constants. The earlier application in Section N23.4 shows the result of **; Rhs = gc,ttme ; Rh2 = one,hinc**.

The log likelihood is somewhat different when the data consist of a set of ranks. The probability that enters the likelihood is as follows: Suppose there are a total of *J* ranks provided, and the outcomes are labeled (1), (2), ..., (*J*) where the sequencing indicates the ranking. (We continue to allow the number of alternatives to vary by individual.) Thus, alternative (1) is the most preferred, alternative (2) is second, and so on. For the present, assume that there are no ties. Then, the observation of a set of ranks is equivalent to the following compound event:

Alternative (1) is preferred to alternatives (2), ...(*J*),
Alternative (2) is preferred to alternatives (3), ...(*J*),
...
Alternative (*J*-1) is preferred to alternative (*J*).

The joint probability is the product of the probabilities of these events. There are, therefore, *J*-1 terms in the log likelihood, each of which is similar to the one shown above, but each has a different choice set. Combining terms, we have the following contribution of an individual to the log likelihood

$$\text{Log } L_i \;=\; \sum_{j=1}^{J_i-1} \log \frac{\exp\left[\boldsymbol{\beta}'\mathbf{x}_{ij}\right]}{\sum_{q=j}^{J_i} \exp\left[\boldsymbol{\beta}'\mathbf{x}_{iq}\right]}.$$

Note that the number of terms in the denominator is different for each *j* in the outer summation. The first and second derivatives can be constructed from results already given, and are not appreciably more complicated. They involve the same terms as given earlier, with an outer summation. If there are unranked alternatives, then the outer summation is from 1 to J_i - 1 - *nties*, where *nties* is the number of alternatives in the lowest ranked group less one. (E.g., 1,2,3,4,4,4 has *nties* = 2.)

N23.7 Random Regret Model

The random regret model begins from an assumption that when choosing between alternatives, decision makers seek to minimize anticipated random regret, where random regret consists of the sum of the familiar IID extreme value and a regret function defined below. Systematic regret for choice *i*, is R_i, which consists of the sum of the binary regrets associated with bilateral comparisons of the attributes of the chosen alternative and the available alternatives. (See Chorus (2010), and Chorus, Greene and Hensher (2013).)

Attribute level regret for the *k*th attribute for alternative *i* compared to available alternative *j* is

$$R_{ij}(k) = \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$

Systematic regret for choice *i* is the sum over the available alternatives of the systematic regret,

$$R_i = \sum_{j\neq i} \sum_{k=1}^{K} \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$

Random regret for alternative *i* is $R_i + \varepsilon_i$. Minimization of regret is equivalent to maximization of the negative of regret. This produces the familiar form for the probability,

$$P_i = \frac{\exp(-R_i)}{\sum_{j=1}^{J} \exp(-R_j)}.$$

We also consider a hybrid form, in which some attributes are treated in random regret form and others are contributors to random utility. The result is

$$R_i = \sum_{k=1}^{K} \beta_k x_{ik} - \sum_{j\neq i} \sum_{k=1}^{K} \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$

The maximum likelihood estimator is developed from this expression for the probabilities of the outcomes. Results produced by this model take the general form for multinomial choice models, as shown in the example below. Elasticities produced by **; Effects:…** are derived in Section N23.7.3.

N23.7.1 Commands for Random Regret

The command for the random regret model is

RRLOGIT ; Lhs = choice variable ; Choices = … specification of the choice set
; Rhs = attributes to be treated in the random regret form
; Rh2 = attributes interacted with ASCs, also in random regret form
; RUM = attributes that are treated in the random utility form
; … other options the same as used for CLOGIT … $

Note that for purposes of the functional form, the Rh2 variables are treated as if they were in the RR form. This is probably not a useful format, so the RUM list is provided for variables that should appear linearly in the utility function. For example, alternative specific constants should generally be explicit in the RUM list, rather than expanded in the Rh2 list. An example appears below.

N23.7.2 Application

In the specification below, the model is fit first in random utility form, including alternative specific constants. The second model treats the first three attributes in random regret form.

```
CLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = gc,ttme,invt,invc,aasc,tasc,basc,hinca
          ; Effects: gc(*)/invc(*) $
RRLOGIT   ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = gc,ttme,invt
          ; RUM = invc,aasc,tasc,basc,hinca
          ; Effects: gc(*)/invc(*) $
```

The models are not nested, so one cannot use a likelihood ratio test to search for the functional form. The noticeable increase in the log likelihood with the RR form below is suggestive of an improved fit, but it cannot be used formally as the basis for a test.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                 Choice
Log likelihood function      -182.33831
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    380.7 AIC/N =    1.813
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3574 .3492
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|     .07560***      .01825     4.14  .0000       .03983    .11137
    TTME|    -.10290***      .01099    -9.37  .0000     -.12444   -.08137
    INVT|    -.01435***      .00265    -5.41  .0000     -.01955   -.00915
    INVC|    -.08952***      .01995    -4.49  .0000     -.12863   -.05042
    AASC|    4.06574***     1.05260     3.86  .0001     2.00268   6.12881
    TASC|    4.27393***      .51214     8.35  .0000     3.27015   5.27772
    BASC|    3.71445***      .50856     7.30  .0000     2.71769   4.71121
   HINCA|     .02364**       .01155     2.05  .0407       .00100    .04628
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |     AIR    TRAIN       BUS      CAR
--------+-----------------------------------
     AIR|  5.4152  -2.3448  -2.3448  -2.3448
   TRAIN| -2.3946   7.4483  -2.3946  -2.3946
     BUS| -1.1512  -1.1512   7.5620  -1.1512
     CAR| -1.9584  -1.9584  -1.9584   5.2548
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
INVC    |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR| -5.2895   2.3425   2.3425   2.3425
   TRAIN|  1.0567  -3.5392   1.0567   1.0567
     BUS|   .4276    .4276  -2.5676    .4276
     CAR|   .4166    .4166    .4166  -1.4630

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -173.31398
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    362.6 AIC/N =    1.727
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3892 .3814
>>> Random Regret Form of MNL Model <<<
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|     .02634***      .00458     5.75  .0000      .01735    .03532
    TTME|    -.03606***      .00426    -8.46  .0000     -.04441   -.02771
    INVT|    -.00877***      .00121    -7.28  .0000     -.01113   -.00641
        |Attributes Attended to in Random Utility Form
    INVC|    -.05957***      .01049    -5.68  .0000     -.08012   -.03902
    AASC|    1.85720**       .86496     2.15  .0318      .16190   3.55250
    TASC|    2.59183***      .33957     7.63  .0000     1.92629   3.25736
    BASC|    1.99911***      .33786     5.92  .0000     1.33692   2.66130
   HINCA|     .02048**       .01021     2.01  .0448      .00047    .04048
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR|  1.6493  -1.0544  -1.0544  -1.0544
   TRAIN|  -.6910   2.7384   -.6910   -.6910
     BUS|  -.4518   -.4518   2.5840   -.4518
     CAR|  -.4492   -.4492   -.4492   2.0639

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
INVC    |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     AIR| -3.1053   1.9733   1.9733   1.9733
   TRAIN|   .5619  -2.4964    .5619    .5619
     BUS|   .3116    .3116  -1.6814    .3116
     CAR|   .1941    .1941    .1941  -1.0566
```

N23.7.3 Technical Details: Random Regret Elasticities

The definition of R_i is

$$R_i = \sum_{j \neq i} \sum_{k=1}^{K} \ln\{1 + \exp[\beta_k (x_{jk} - x_{ik})]\}$$

To simplify the expression, add back then subtract the ith term in the outer sum,

$$R_i = \left\{ \sum_{j=1}^{J} \sum_{k=1}^{K} \ln\{1 + \exp[\beta_k (x_{jk} - x_{ik})]\} \right\} - K\ln 2.$$

By the definition given earlier,

$$P_i = \frac{\exp[-R_i]}{\sum_{j=1}^{J} \exp[-R_j]}$$

Now, differentiate the probability. To obtain $\partial P_i/\partial x_{lk}$, use the result $\partial P_i/\partial x_{lk} = P_i \partial \ln P_i / x_{lk}$.

$$\ln P_i = -R_i - \ln \Sigma_{j=1}^{J} \exp(-R_j),$$

$$\frac{\partial \ln P_i}{\partial x_{lk}} = \frac{-\partial R_i}{\partial x_{lk}} - \frac{\partial \ln \Sigma_{j=1}^{J} \exp(-R_j)}{\partial x_{lk}}$$

$$= \frac{-\partial R_i}{\partial x_{lk}} - \frac{\Sigma_{j=1}^{J} \partial \exp(-R_j)/\partial x_{lk}}{\Sigma_{j=1}^{J} \exp(-R_i)}$$

$$= \frac{-\partial R_i}{\partial x_{lk}} - \frac{\Sigma_{j=1}^{J} \exp(-R_j) \partial(-R_j)/\partial x_{lk}}{\Sigma_{j=1}^{J} \exp(-R_i)}$$

$$= \frac{-\partial R_i}{\partial x_{lk}} - \sum_{j=1}^{J} P_j \frac{\partial(-R_j)}{\partial x_{lk}}$$

$$= \left(\sum_{j=1}^{J} P_j \frac{\partial R_j}{\partial x_{lk}} \right) - \frac{\partial R_i}{\partial x_{lk}}$$

We require $\partial R_i/\partial x_{lk}$.

$$\frac{\partial R_i}{\partial x_{lk}}(\text{where } l \neq i) \qquad = \beta_k \frac{\exp[\beta_k(x_{lk} - x_{ik})]}{1+\exp[\beta_k(x_{lk} - x_{ik})]} = \beta_k q(l,i,k)$$

$$\frac{\partial R_i}{\partial x_{ik}} R_i(\text{i.e., where } l = i) \quad = -\beta_k \sum_{j \neq \text{i}}^{J} \frac{\exp[\beta_k(x_{jk} - x_{ik})]}{1+\exp[\beta_k(x_{jk} - x_{ik})]}$$

$$= -\beta_k \sum_{j \neq i}^{J} q(l,j,k) = \beta_k (q(l,l,k) - 1)$$

because $\left[\sum_{j=1}^{J} q(l,j,k)\right] = 1.$

Remove the lth term from this sum to obtain

$$\left(\sum_{j=1}^{J} P_j \frac{\partial R_j}{\partial x_{lk}}\right) - \frac{\partial R_i}{\partial x_{lk}} = \left(\sum_{j \neq l} P_j \frac{\partial R_j}{\partial x_{lk}}\right) + P_l \frac{\partial R_l}{\partial x_{lk}} - \frac{\partial R_i}{\partial x_{lk}}$$

Now insert the expressions above. The alien terms in the first line go inside the brackets. The sum in the second one goes in the extra term

$$\left(\sum_{\text{j}^1\text{l}}^{\text{J}} P_j \frac{\partial R_j}{\partial x_{lk}}\right) + P_l \frac{\partial R_l}{\partial x_{lk}} - \frac{\partial R_i}{\partial x_{lk}}$$

$$= \left(\sum_{j \neq l}^{J} P_j \beta_k q(l,j,k)\right) + P_l\left(-\beta_k \sum_{j \neq l}^{J} q(l,j,k)\right) - \frac{\partial R_i}{\partial x_{lk}}$$

$$= \left[\sum_{j \neq l}^{J} (P_j - P_l)\beta_k q(l,j,k)\right] - \frac{\partial R_i}{\partial x_{lk}}$$

Now, restore the lth term, which will equal zero, since it contains $P_l - P_l$ to obtain the final result:

$$\text{Elasticity} = \frac{\partial \ln P_i}{\partial x_{lk}} = \left[\sum_{j=1}^{J} (P_j - P_l)\beta_k q(l,j,k)\right] - \frac{\partial R_i}{\partial x_{lk}}$$

For i not equal to l, i.e., the cross elasticity, this produces

$$\text{Elasticity} = \frac{\partial \ln P_i}{\partial x_{lk}} = \beta_k \left\{\left[\sum_{j=1}^{J} (P_j - P_l) q(l,j,k)\right] - q(l,i,k)\right\}$$

For i equal to l, i.e., the own elasticity,

$$\text{Elasticity} = \frac{\partial \ln P_i}{\partial x_{ik}} = \beta_k \left\{\left[\sum_{j=1}^{J} (P_j - P_l) q(i,j,k)\right] - [q(i,i,k) - 1]\right\}$$

N23.8 Fixed Effects Multinomial Logit Model

The multinomial logit (MNL) model extends the binary choice model of Section N9.5 to choice among multiple alternatives. There are $J + 1$ unordered outcomes denoted $j = 0,1,\ldots,J$. The probability that individual i makes choice j in choice situation t is

$$\text{Prob}(Y_{itj} = 1 \mid \mathbf{X}_i, \alpha_{ij}) = \frac{\exp(\alpha_{ij} + \boldsymbol{\beta}'\mathbf{x}_{itj})}{\Sigma_{m=0}^{J_{it}} \exp(\alpha_{im} + \boldsymbol{\beta}'\mathbf{x}_{itm})},\ j = 0,\ldots,J_{it},\ t = 1,\ldots,T_i.$$

(The model does not accommodate choice invariant variables as they would be indistinguishable from the fixed effects.) A normalization is also necessary. (The function is homogeneous of degree zero in α_{ij}. One of the α_{ij}'s for each i must be normalized. The solution, as suggested (indirectly) by Chamberlain (1984) is to define the choice outcome in terms of a base alternative (we use $j = 0$ as the base). The probability (model) is now

$$\text{Prob}(Y_{itj} = 1 \mid \mathbf{X}_i, \boldsymbol{\alpha}_i) = \frac{\exp(\alpha_{ij} + \boldsymbol{\beta}'\mathbf{x}_{itj})}{1 + \Sigma_{m=1}^{J_{it}} \exp(\alpha_{im} + \boldsymbol{\beta}'\mathbf{x}_{itm})},\ j = 1,\ldots,J_{it},\ t = 1,\ldots,T_i,$$

$$\text{Prob}(Y_{it0} = 1 \mid \mathbf{X}_i, \boldsymbol{\alpha}_i) = \frac{1}{1 + \Sigma_{m=1}^{J_{it}} \exp(\alpha_{im} + \boldsymbol{\beta}'\mathbf{x}_{itm})},\ j = 1,\ldots,J_{it},\ t = 1,\ldots,T_i.$$

N23.8.1 Estimation of the Fixed Effects Multinomial Logit Model

The command for estimation of the model is

```
FEMLOGIT    ; Lhs     = choice variable
            ; Choices = list of choice labels
            ; Rhs     = list of attributes (there is no ; Rh2 list)
            ; Panel or ; Pds specification $
```

The constants, α_{ij} will be conditioned out of the likelihood for estimation, and are not estimated. As such, it is not possible to compute predictions or partial effects.

N23.8.2 Application

In the following, we treat the 210 observations on mode choice as if it were a choice experiment in which each individual chose seven times.

```
FEMLOGIT    ; Lhs  = mode ; choices=air,train,bus,car
            ; Rhs  = gc,ttme,invc,invt
            ; Pds  = 7 $
```

The first two panels in the results describe the observed outcomes and the group sizes and notes that the last alternative, *car*, is used as the base. The second panel notes counts of observations that must be excluded from the separate estimations. If an individual never chooses a particular mode, then that individual must be excluded from the sample when estimating the equation for that mode. If an individual always chooses a particular mode, then they provide no information for estimating any of the equations are omitted from the sample altogether. Three sets of iterations are shown, one for each of the modes save for the base. The MDE is shown last.

```
-----------------------------------------------------------------
Fixed Effects Multinomial Logit Model
Sample contains     30 individuals.      MinT(i)=  7,  MaxT(i)=  7
Summary of Observed Choices                    CAR is the base choice
-----------------------------------------------------------------
Choice     % of Observed   Total   Choice     % of Observed   Total
--------------------------------   --------------------------------
AIR            27.62         58    TRAIN          30.00          63
BUS            14.29         30    CAR            28.10          59
-----------------------------------------------------------------
------------------------------------------------------
Sample Exclusions: Number of Choice Situations Dropped
Choice--   Number never chose      Number always chose
--------   ------------------      -------------------
AIR                 13                        2
TRAIN               13                        4
BUS                 15                        0
CAR       (base)    12                        1
If alt is never chosen, case not used for this choice.
If alt is always chosen, case not used for any choice.
------------------------------------------------------
Iterative procedure has converged
Normal exit:   6 iterations. Status=0, F=    .2261113D+02
Iterative procedure has converged
Normal exit:   7 iterations. Status=0, F=    .1746458D+02
Iterative procedure has converged
Normal exit:   7 iterations. Status=0, F=    .1430137D+02

-----------------------------------------------------------------------------
Fixed Effects  Multinomial Logit  Model
Dependent variable                  MODE
Log likelihood function        -54.37708
Restricted log likelihood     -122.17463
Chi squared [  4](P= .000)     135.59510
Significance level                .00000
McFadden Pseudo R-squared       .5549233
Estimation based on N =    840, K =   4
Inf.Cr.AIC  =    116.8 AIC/N =      .139
Estimator is Minimum Distance Wtd. Avrg
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|     .04399        .02737     1.61  .1080     -.00966    .09763
    TTME|    -.07150***     .01350    -5.30  .0000     -.09795   -.04505
    INVC|    -.04026        .02888    -1.39  .1633     -.09687    .01635
    INVT|    -.00905**      .00389    -2.32  .0202     -.01668   -.00141
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

N23.8.3 Technical Background for the FE MNL Estimator

The fixed effects (FE) binary logit model is

$$\text{Prob}(y_{it} = 1|\mathbf{x}_{it},\boldsymbol{\beta},\alpha_i) = \Lambda(\boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i),\ i = 1,\ldots,n;\ t = 1,\ldots,T_i, \tag{1}$$

where $\Lambda(z) = \exp(z)/(1 + \exp(z))$. Direct maximum likelihood estimation is straightforward in principle and can be done easily with *NLOGIT* even if n is very large – see Section N9.4 – but for two complications. There is no counterpart to the device (group mean deviations) used to partition the FE linear regression to estimate separately $\boldsymbol{\beta}$ then α_i. Here, all $n + K$ parameters are estimated simultaneously by maximizing the full log likelihood. The greater complication is the *incidental parameters problem*. As has been documented in numerous sources, plim $\hat{\boldsymbol{\beta}} = q(T)\boldsymbol{\beta}$ where $q(2) = 2$ and $q(T)$ appears to decline fairly rapidly as T increases.

Chamberlain (1984) developed a conditional estimator based on

$$\text{Prob}\left[Y_{i1} = y_{i1},\ldots,Y_{iT} = y_{iT} \mid \mathbf{X}_i, \left(\Sigma_{t=1}^{T} Y_{it}\right) = \left(\Sigma_{t=1}^{T} y_{it}\right)\right] = \frac{\exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T} y_{it}\mathbf{x}_{it}\right]}{\sum_{p=1}^{p=\binom{T}{S_i}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T} d_{it|p}\mathbf{x}_{it}\right]} \tag{2}$$

where in the denominator, S_i equals the sum of the actual realizations of Y_{it}, $\Sigma_{t=1}^{T} y_{it}$ while $P = \binom{T}{S_i}$ equals the number of different permutations of the elements (ones and zeros) in the vector $\mathbf{y}_i' = (y_{i1}, y_{i2},\ldots y_{iT})$. The pth vector in the set of P possible vectors is denoted $\mathbf{d}_{i|p}' = (d_{i1|p},\ldots,d_{iT|p})$. The conditional log likelihood (that is free of α_i) is

$$\log L_C = \sum_{i=1}^{n} \boldsymbol{\beta}'\Sigma_{t=1}^{T} y_{it}\mathbf{x}_{it} - \log \sum_{p=1}^{p=\binom{T}{S_i}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T} d_{it|p}\mathbf{x}_{it}\right] \tag{3}$$

When S_i equals 0 or T, then $P = 1$, and the conditional probability equals one, so the contribution of this observation group to the log likelihood is zero. This is the estimator shown in Section N9.5.

Chamberlain notes, 'Since L is in the form of a conditional logit log-likelihood function, it can be maximized by standard programmes.' Strictly speaking, this is correct. For practical purposes, however, the observation understates the complexity of the calculations. The number of terms in the denominator of the conditional probability is $P = \binom{T}{S_i}$. When $T = 2$ and $S = 1$, there are 2; when $T = 3$ and $S = 1$ or 2, there are 3. The number grows rapidly; if T equals 50, the number of terms when S is 25 is $P = 1.26 \times 10^{14}$. A remarkable result presented in Krailo and Pike (1984) allows quick computation of this estimator even for large T. For example, the following experiment,

```
ROWS      ; 50000 $
CREATE    ; x1 = Rnn(0,1); x2 = Rnn(0,1) ; y = Rnd(2)-1 $
TIMER $
LOGIT     ; Lhs = y; Rhs = x1,x2 ; Pds = 50 $
```

will estimate a fixed effects binary logit model using the conditional estimator for 1,000 individuals and 50 periods in less than a second.

Conditional Maximum Likelihood Estimation

The remainder of the treatment in Chamberlain motivates an otherwise conventional maximum likelihood estimator that has now been installed in *NLOGIT* and other programs. (We have dwelt on this now apparently moot complication as it will reappear in the discussion of the multinomial logit model.)

The fixed effects multinomial logit model is

$$\text{Prob}(Y_{itj} = 1 | \mathbf{X}_i, \alpha_{ij}) = \frac{\exp(\alpha_{ij} + \boldsymbol{\beta}'\mathbf{z}_{itj})}{1 + \Sigma_{m=1}^{J} \exp(\alpha_{im} + \boldsymbol{\beta}'\mathbf{z}_{itm})}, \; j = 1,\ldots, J, \tag{4}$$

where $\mathbf{z}_{itj} = \mathbf{x}_{itj} - \mathbf{x}_{it0}$. The practical complexity and the incidental parameters problems both reappear at this point. The conditional probability is

$$\text{Prob}\left[\begin{pmatrix} Y_{i11} = y_{i11}, Y_{i12} = y_{i12}, \ldots, Y_{i1J} \\ \ldots \\ Y_{iT1} = y_{iT1}, Y_{iT2} = y_{iT2}, \ldots, Y_{iTJ} \end{pmatrix} \middle| \; \mathbf{X}_i, \begin{matrix} \left(\Sigma_{t=1}^{T} Y_{itj}\right) = \left(\Sigma_{t=1}^{T} y_{itj}\right) \\ \Sigma_{j=1}^{J} Y_{itj} = 1 \end{matrix}\right]$$

$$= \frac{\exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J} y_{itj}\mathbf{x}_{itj}\right]}{\sum_{p=1}^{p=\binom{JT}{\Sigma_{j=1}^{J} S_{ij}}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J} d_{itj|p}\mathbf{x}_{it}\right]} \tag{5}$$

The complexity of the computation has increased enormously. The number of permutations in the denominator is over the elements in a *JT*-element vector, rather than a *T*-element vector. Chamberlain's statement of the result does not hint at the complexity; '$B_i = \{\mathbf{d}=(d_{11},\ldots,d_{TJ}) | d_{tj} = 0$ or 1, $\Sigma_j d_{tj}=1$, $\Sigma_t d_{tj}=S_{ij}\}$.' He states: 'This is in the form of a conditional logit log-likelihood function and can be maximized by standard programmes.'

The number of terms would seem to be even more daunting than before. The number of permutations in a *JT*-vector is vastly more than in a *T*-vector, even for moderate *J* (such as 4 or 5 which would be typical). (We did not find even mention of the multinomial logit model in contemporary textbooks, applications, or commercial software.)

The vectorization of **d** in (5) and Chamberlain's definition of the set of permutations has obscured a vast simplification of the calculation. Rather than write B_i as a set of *JT*-element vectors, it is more convenient to define B_i as a set of $T \times J$ matrices $\mathbf{D}_i$ such that each row in the matrix contains exactly one 1 (this from $\Sigma_j d_{tj} = 1$) and whose *j*th column sums to S_{ij}, the sum of the outcomes for the binary choice indicators y_{itj}. In (5), reverse the order of summation in the numerator to obtain

$$\begin{aligned} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J} y_{itj}\mathbf{x}_{itj}\right] &= \exp\left[\boldsymbol{\beta}'\Sigma_{j=1}^{J}\Sigma_{t=1}^{T} y_{itj}\mathbf{x}_{itj}\right] \\ &= \exp\left[\boldsymbol{\beta}'\left(\Sigma_{t=1}^{T} y_{it1}\mathbf{x}_{it1}\right) + \boldsymbol{\beta}'\left(\Sigma_{t=1}^{T} y_{it2}\mathbf{x}_{it2}\right) + \ldots\right] \\ &= \prod_{j=1}^{J} \exp\left[\boldsymbol{\beta}'\left(\Sigma_{t=1}^{T} y_{itj}\mathbf{x}_{itj}\right)\right]. \end{aligned}$$

The numerator in (5) is the product of the numerators in (2) for the J binary choices defined by y_{itj} = 1 or 0. By similarly imposing the constraint that every row in $\mathbf{D}_i$ contains exactly one 1, the denominator is reconstructed as

$$\sum_{p=1}^{p=\binom{JT}{\Sigma_{j=1}^{J}S_{ij}}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J}d_{itj|p}\mathbf{x}_{it}\right] = \sum_{j=1}^{J} \sum_{p|j=1}^{p|j=\binom{T}{S_{ij}}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J}d_{itj|p}\mathbf{x}_{it}\right]$$
$$= \sum_{p|j=1}^{p|j=\binom{T}{S_{ij}}} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{T}\Sigma_{j=1}^{J}d_{itj|p}\mathbf{x}_{it}\right]$$

The conditional probability in Chamberlain's MNL estimator is equal to the product of the conditional binary choice probabilities, where each alternative is treated as a binary choice of it vs. the base alternative. The log likelihood is, therefore, the sum of the individual log likelihoods for these J binary choices. (Each term is based on the set of observations that chose either the base or that specific alternative.) Maximization of the log likelihood for the FE MNL model can, at least in principle, use tools already available in standard software.

Minimum Distance Estimation

In fact, the arrangement of the data and computations for this estimator remains devilishly complicated, albeit less so than prescribed by Chamberlain. However, there is a far simpler estimator that preserves nearly all the efficiency of the full conditional MLE based on the minimum distance principle. Returning to the original model, consider the binary choice between $y_{itj} = m$ or 0. This would imply the original treatment for the binary case. We could, in principle, estimate $\boldsymbol{\beta}$ by using the subsample of individuals who choose either outcome m or the base outcome, 0. This would be an application of the simpler conditional binary choice estimator. Let $\hat{\boldsymbol{\beta}}_m, m = 1,...,J$, denote these estimators, and let $\mathbf{V}_m$ denote the associated estimated asymptotic covariance matrix for each estimator. We now have J separate estimators of the same $\boldsymbol{\beta}$. We will combine these into a single estimator by minimizing the estimation criterion

$$q = \sum_{m=1}^{J} \left(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right) \mathbf{V}_m^{-1}\left(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right)$$

The estimator that results is a matrix weighted average,

$$\hat{\boldsymbol{\beta}}_{MDE} = \left[\sum_{m=1}^{J}\mathbf{V}_m^{-1}\right]^{-1} \sum_{m=1}^{J}\mathbf{V}_m^{-1}\hat{\boldsymbol{\beta}}_m .$$

The estimated asymptotic covariance matrix is the first term,

$$\left[\sum_{m=1}^{J}\mathbf{V}_m^{-1}\right]^{-1} .$$

N24: The Scaled Multinomial Logit Model

N24.1 Introduction

The scaled multinomial logit (SML) model incorporates individual heterogeneity in the multinomial logit model. The model is a particular form of the generalized mixed multinomial logit model discussed in Chapters N29 and N33. The general form of the scaled MNL derives from a random utility model with heteroscedasticity across individuals, rather than across choices;

$$U_{it,j} = \beta' x_{it,j} + (1/\sigma_i)\varepsilon_{it,j}.$$

where $\varepsilon_{it,j}$ has the usual type I extreme value distribution. Note that the scaling is choice invariant but varies across individuals. The model is equivalent to the multinomial logit model of Chapter N17 with individual specific parameter vector, $\boldsymbol{\beta}_i = \sigma_i \boldsymbol{\beta}$**;**

$$\text{Prob}(\text{choice}_{i,t} = j) = \frac{\exp(\boldsymbol{\beta}_i' \mathbf{x}_{ij,t})}{\sum_{j=1}^{J} \exp(\boldsymbol{\beta}_i' \mathbf{x}_{ij,t})},\ j = 1,...,J;\ i = 1,...,n;\ t = 1,...,T,$$

where $\boldsymbol{\beta}_i = \sigma_i \boldsymbol{\beta}$.

When the variation across individuals is modeled as due to unobserved heterogeneity, we specify

$$\sigma_i = \exp(-\tau^2/2 + \tau w_i).$$

The term, w_i in the scale factor is random variation across individuals. The structural parameter, τ, carries the model. With $\tau = 0$, the model reverts to the original multinomial logit model. It is not possible to identify a separate location parameter in σ_i – this would correspond to the overall constant scale factor for the variance, which is already present; $\text{Var}[\varepsilon_{it,j}] = \gamma^0 = \pi^2/6$. The constant $-\tau^2/2$ is chosen so that $E[\sigma_i] = 1$ if $w_i \sim N[0,1]$. Note that if w_i is normally distributed, which is assumed, then σ_i has a lognormal distribution with mean equal to 1. The model thus far treats the heterogeneity in σ_i as all unobserved. The specification can be extended to allow observed heterogeneity in the scale factor as well, as in

$$\sigma_i = \exp(-\tau^2/2 + \tau w_i + \boldsymbol{\delta}' \mathbf{z}_i).$$

The model takes some aspects of the random parameters logit (RPLOGIT) model discussed in Chapter N29. The formulation above suggests a panel data – or stated choice experiment form for repeated choice situations. The assumption is that σ_i is constant through time. This can be relaxed, as shown below, if one treats the panel as if it were a cross section.

N24.2 Command for the Scaled MNL Model

The general command form for this model is

SMNLOGIT ; Lhs = choice variable
; Choices = choice set specification
; Rhs = attributes …
; Rh2 = interactions with ASCs $

Utility functions may be specified using the explicit form shown in Chapter N20. The scaling is applied to the full coefficient vector regardless of which way it is specified. Several variations on this basic form will be useful. The heteroscedasticity in observable variables is specified with

; Hft = variables in z (does not contain a constant term, one).

All random parameters models in *NLOGIT* can be fit with 'panel' or repeated choice experiment data. The panel is specified as always, with

; Pds = number of choice situations …

See Chapters N18, N29 and N33 for further discussion of panel data sets. The model is fit by maximum simulated likelihood. You can control two important aspects of this computation. Use

; Pts = number of random draws for the simulations
and **; Halton**

to specify using Halton sequences rather than random draws (samples) to do the integration.

Elasticities, saved probabilities, and other optional features associated with the MNL model are all provided the same way as in the simpler formulations.

N24.3 Application

Two applications below illustrate the estimator. The first modifies the basic MNL

SMNLOGIT ; Lhs = Mode ; Choices = air,train,bus,car
; Rhs = gc,ttme,invc,invt,one
; Halton
; Pts = 25 $

The second adds observed heterogeneity, household income, to the model for the variance. To illustrate the estimator, we have specified that the sample is composed of groups of three choice situations (this is purely artificial – the sample is actually a cross section).

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  MODE
Log likelihood function      -184.50669
Estimation based on N =    210, K =   7
Inf.Cr.AIC  =    383.0 AIC/N =    1.824
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3498 .3425
Chi-squared[ 4]          =    198.50415
Prob [ chi squared > value ] =   .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      GC|     .06930***      .01743     3.97  .0001      .03513    .10346
    TTME|    -.10365***      .01094    -9.48  .0000     -.12509   -.08221
    INVC|    -.08493***      .01938    -4.38  .0000     -.12292   -.04694
    INVT|    -.01333***      .00252    -5.30  .0000     -.01827   -.00840
   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
--------+--------------------------------------------------------------------
Scaled Multinomial Logit Model
Log likelihood function      -170.10469
McFadden Pseudo R-squared      .4156924
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    356.2 AIC/N =    1.696
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4157 .4082
Constants only   -283.7588  .4005 .3928
At start values  -184.0543  .0758 .0639
Response data are given as ind. choices
Replications for simulated probs. =  25
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|     .12856*        .07138     1.80  .0717     -.01135    .26847
    TTME|    -.24605***      .05469    -4.50  .0000     -.35324   -.13885
    INVC|    -.15957*        .08376    -1.91  .0568     -.32375    .00460
    INVT|    -.02319**       .01134    -2.05  .0408     -.04541   -.00097
   A_AIR|    13.1526***     3.87351     3.40  .0007      5.5607   20.7445
 A_TRAIN|    9.64084***     2.50951     3.84  .0001     4.72228  14.55939
   A_BUS|    8.35466***     2.08397     4.01  .0001     4.27015  12.43917
        |Variance parameter tau in GMX scale parameter
TauScale|    1.11114***      .12892     8.62  .0000      .85846   1.36381
        |Weighting parameter gamma in GMX model
GammaMXL|        0.0     .....(Fixed Parameter).....
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     .99942       1.48264      .67  .5003    -1.90650   3.90534
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

These are the estimated elasticities of the probabilities with respect to the generalized cost of travel. (This seems not to be a very good specification. The elasticity appears to have the wrong sign. The signs of the other variables that involve cost and time of travel have expected negative signs.) The elasticities for the unscaled multinomial logit model are shown first.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|   4.9664  -2.1466  -2.1466  -2.1466
   TRAIN|  -2.1912   6.8310  -2.1912  -2.1912
     BUS|  -1.0547  -1.0547   6.9321  -1.0547
     CAR|  -1.8020  -1.8020  -1.8020   4.8097

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|   3.2009   -.9987   -.8911  -1.2166
   TRAIN|  -1.1497   4.6046  -1.1497  -1.6580
     BUS|   -.6515   -.6636   3.4699   -.7324
     CAR|  -1.6106  -1.7145  -1.1722   3.1863
```

The second example adds observed heterogeneity to the scale factor.

```
SMNLOGIT   ; Lhs = Mode ; Choices = air,train,bus,car
           ; Rhs = gc,ttme,invc,invt,one
           ; Hft = hinc
           ; Pds = 3
           ; Halton
           ; Pts = 25 $
```

```
-----------------------------------------------------------------------------
Scaled Multinomial Logit Model
Dependent variable                MODE
Log likelihood function     -175.97384
Restricted log likelihood   -291.12182
Chi squared [   9 d.f.]      230.29595
Significance level              .00000
McFadden Pseudo R-squared     .3955319
Estimation based on N =    210, K =   9
Inf.Cr.AIC  =    369.9 AIC/N =    1.762
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3955 .3868
Constants only   -283.7588  .3798 .3709
At start values  -183.9030  .0431 .0292
Response data are given as ind. choices
Replications for simulated probs. =  25
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Number of obs.=   210, skipped    0 obs
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|     .13113*         .07373     1.78  .0753     -.01338    .27565
    TTME|    -.23196**        .09485    -2.45  .0145     -.41786   -.04606
    INVC|    -.16089*         .08297    -1.94  .0525     -.32351    .00173
    INVT|    -.02384**        .01186    -2.01  .0443     -.04708   -.00061
   A_AIR|    12.1298**       5.24495     2.31  .0207      1.8499   22.4097
 A_TRAIN|    8.92354**       3.84111     2.32  .0202     1.39511  16.45197
   A_BUS|    8.09167**       3.53386     2.29  .0220     1.16543  15.01791
        |Variance parameter tau in GMX scale parameter
TauScale|    1.19427***       .33152     3.60  .0003      .54451   1.84404
        |Heterogeneity in tau(i)
 TauHINC|    -.00243          .00535     -.45  .6500     -.01292    .00806
        |Weighting parameter gamma in GMX model
GammaMXL|        0.0     .....(Fixed Parameter).....
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|    1.04732         1.51238      .69  .4886    -1.91688   4.01152
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+----------------------------------
     AIR|   3.3612  -1.0638  -1.0117  -1.3494
   TRAIN|  -1.1292   4.7327  -1.1091  -1.5165
     BUS|   -.7581   -.7294   3.7079   -.8212
     CAR|  -1.6461  -1.6820  -1.2590   3.2577
```

N24.4 Technical Details

The model is estimated using maximum simulated likelihood. The full likelihood function is that of the generalized mixed logit model in Chapters N29 and N33. The restrictions used to produce this model are $\mathbf{\Gamma} = \mathbf{0}$ in the mixed logit part of the model – see Chapter N29 and $\gamma = 0$ in the GMXLOGIT formulation – see Chapter N33. (All of the other parameters that produce the random parameters model are also suppressed.) The scaled MNL thus adds a single new parameter to the MNL model, τ.

The value of σ_i reported in the final model results is the sample average value, where the average is taken in two directions. The value of σ_i is obtained as the average over the random draws or Halton draws. Then, the average reported (with the sample standard deviation) is averaged over the individuals in the sample. The model specifies that the population expected value of σ_i equals one. The average reported is near one (in accordance with the law of large numbers) but differs slightly because of sampling variability.

N25: Latent Class and 2^K Multinomial Logit Model

N25.1 Introduction

The latent class model is similar to the random parameters model of Chapter N29. In the latent class formulation, parameter heterogeneity across individuals is modeled with a discrete distribution, or set of 'classes.' The situation can be viewed as one in which the individual resides in a 'latent' class, *c*, which is not revealed to the analyst. There are a fixed number of classes, *C*. Estimates consist of the class specific parameters and for each person, a set of probabilities defined over the classes. Individual *i*'s choice among *J* alternatives at choice situation *t* given that they are in class *c* is the one with maximum utility, where the utility functions are

$$U_{jit} = \boldsymbol{\beta}_c'\mathbf{x}_{jit} + \varepsilon_{jit},$$

where U_{jit} = utility of alternative *j* to individual *i* in choice situation *t*,

$\mathbf{x}_{jit}$ = union of all attributes and characteristics that appear in all utility functions. For some alternatives, $x_{jit,k}$ may be zero by construction for some attribute *k* which does not enter their utility function for alternative *j*,

ε_{jit} = unobserved heterogeneity for individual *i* and alternative *j* in choice situation *t*,

$\boldsymbol{\beta}_c$ = class specific parameter vector.

Within the class, choice probabilities are assumed to be generated by the multinomial logit model.

As noted, the class membership is not observed. (Unconditional class probabilities are specified by the multinomial logit form.) The class specific probabilities may be a set of fixed constants if no observable characteristics that help in class separation are observed. In this case, the class probabilities are simply functions of *C* parameters, θ_c, the last of which is fixed at zero. You will specify the number of classes, *C*, from two to five. This model does not impose the IIA property on the observed unconditional probabilities (though it does within each class.) For a given individual, the model's estimate of the probability of a specific choice is the expected value (over classes) of the class specific probabilities. See technical details in Section N25.11.

N25.2 Model Command

The latent class model is a one level (nonnested) model. To request it, use

LCLOGIT	**; Lhs = ... ; Choices = ...**
	; Rhs = ...
or	**; Model: U(...)=... / U(...) = ... all as usual**
	; ... any other options
	; Pds = number of choice situations fixed or variable (omit if one)
	; Pts = C, the number of classes or ; Nclass = C $

(The model command **NLOGIT ; LCM** may also be used.) The preceding format assumes that the latent class probabilities are constants. If you have variables that are person specific, and constant across choices and choice situations (such as age or income), then you can build them into the model with

; LCM = list of variables

(Do not include *one* in the list.) Other common options include

; Prob = name	to use for estimated probabilities
; Utility = name	to use for estimated utilities

and the usual other options for output, technical output, elasticities, descriptive statistics, etc. (See Chapters N19-N22 for details.) Note that for this estimator,

- Choice based sampling is not supported, though you can use ordinary weights with **; Wts**.
- Data may be individual or proportions.

As in the mixed logit model (Chapter N29), the number of choice situations may vary across individuals. This model may be fit with cross section or repeated choice situation (panel) data. If you do not specify the **; Pds = setting** or **; Panel** specification, it will be assumed that you are using a cross section. In principle, this works, but estimates may have large standard errors. The estimator becomes sharper as the number of observations per person increases.

The number of latent classes must be specified on the command. There is no theory for the right number of classes. If you specify too many, some parameters will be estimated with huge standard errors, or after estimation, the estimated asymptotic covariance matrix will not be positive definite. If you observe either of these conditions, try reducing *C* in the command.

There is no command builder for this version of the choice model. The command must be provided in text form as shown above. The following general options are *not* available for the latent class model:

; Ivb = name	No inclusive values are computed.
; IAS = list	IIA is not testable here, since it is not imposed.
; Cprob = name	Conditional and unconditional probabilities are the same.
; Ranks	This estimator may not be based on ranks data.
; Scale ...	Data scaling is only for the nested logit model.

The remainder of the setup is identical to the multinomial logit model.

N25.3 Individual Specific Results

Denote the class probabilities by π_{ic} and the conditional choice probabilities by $P_{jit}|c$. Within the class, the individual choices from one situation to the next are assumed to be independent. Thus, the conditional probability for the observed sequence of choices for person *t* is

$$P_{ji}|c = \prod_{m=1}^{T_i} P_{jim} \mid c\,,$$

where T_i denotes the number of choice situations for person *i* – this may vary by person; you provide this in your command with the **; Pds = setting** specification. The unconditional probability for the sequence of choices is the expected value,

$$P_{ji} = \sum_{c=1}^{C} \pi_{ic} \prod_{m=1}^{T_i} P_{jim} \mid c = \sum_{c=1}^{C} \text{Prob}(class = c)\text{Prob}(choices \mid c)\,.$$

This is the term that enters the log likelihood for estimation of the model. In this formula, it is implied that the '*j*' indicates the choice that the individual actually makes. We can use Bayes theorem to obtain a 'posterior' estimate of the individual specific class probabilities,

$$\text{Prob(class} = c \mid \text{choices, data)} = \frac{\pi_{ic} \prod_{m=1}^{T_i} P_{jim} \mid c}{\sum_{c=1}^{C} \pi_{ic} \prod_{m=1}^{T_i} P_{jim} \mid c}\,.$$

This provides a person specific set of conditional (posterior) estimates of the class probabilities, $\hat{\pi}_{ic}^*$. With this in hand, we can obtain an individual specific posterior estimate of the parameters,

$$\hat{\boldsymbol{\beta}}_i = \sum_{c=1}^{C} \hat{\pi}_{ic}^* \hat{\boldsymbol{\beta}}_c\,.$$

You can request *NLOGIT* to construct a matrix named *beta_i* containing these individual specific estimates by adding

; Parameters

to the model command. This will create a matrix named *beta_i* that has number of rows equal to the number of individuals (not the number of observations, as you are using a panel) and number of columns equal to the number of elements in $\boldsymbol{\beta}$. Each row will contain $\hat{\boldsymbol{\beta}}_i'$. A second matrix, *classp_i*, that is $N \times C$ will contain the estimated conditional class probabilities, $\hat{\pi}_{ic}^*$, for each individual. An example appears in Section N25.8.

The elements in *classp_i* may also be saved as variables in the data area. You must create the set of variables first, then define a namelist for them. The setting

; Classp = namelist

will then save the variables. For example, for a three class LC model, you could use

```
NAMELIST    ; (new) ; pclass = pc1, pc2, pc3 $
LCLOGIT     ; … ; Nclass = 3 ; Classp = pclass $
```

N25.4 Constraining the Model Parameters

You may specify that certain parameters are to be the same in all classes. Use

; Fix = names of variables if you use ; Rhs
or names of parameters if you use ; Model:...

For example, the model fit in the next section uses the command

```
LCLOGIT   ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,ttme ; Rh2   = one
          ; Pts = 2
          ; ... $
```

This is a two class model. When we add the specification

; Fix = gc

the estimates for the model parameters appear as below. The coefficient on *gc* is the same in the two classes. (We have artificially grouped the observation into 30 groups of seven for the illustration.)

```
LCLOGIT   ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,ttme ; Rh2   = one
          ; Pts = 2
          ; Fix = gc
          ; LCM = hinc ; Pds = 7 $
```

```
-----------------------------------------------------------------------------
Latent Class Logit Model
Dependent variable                    MODE
Log likelihood function         -158.60029
Restricted log likelihood       -291.12182
Chi squared [  11 d.f.]          265.04305
Significance level                  .00000
McFadden Pseudo R-squared         .4552099
Estimation based on N =    210, K =   11
Inf.Cr.AIC  =    339.2 AIC/N =    1.615
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4552 .4455
Constants only   -283.7588  .4411 .4311
At start values  -199.9800  .2069 .1928
Response data are given as ind. choices
Number of latent classes =             2
Average Class Probabilities
     .573  .427
LCM model with panel has      30 groups
Fixed number of obsrvs./group=         7
Number of obs.=   210, skipped    0 obs
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Utility parameters in latent class -->> 1
    GC|1|    -.01366***      .00491    -2.78  .0054     -.02329   -.00404
  TTME|1|    -.18606***      .02726    -6.82  .0000     -.23949   -.13263
 A_AIR|1|    9.68918***     1.76652     5.48  .0000     6.22686  13.15150
A_TRAI|1|    5.36413***      .96114     5.58  .0000     3.48033   7.24793
 A_BUS|1|    6.01580***     1.00863     5.96  .0000     4.03892   7.99268
        |Utility parameters in latent class -->> 2
    GC|2|    -.01366***      .00491    -2.78  .0054     -.02329   -.00404
  TTME|2|    -.04828***      .01660    -2.91  .0036     -.08082   -.01573
 A_AIR|2|    6.24727***     1.31891     4.74  .0000     3.66225   8.83229
A_TRAI|2|    5.52786***     1.06461     5.19  .0000     3.44127   7.61446
 A_BUS|2|    3.62508***     1.13892     3.18  .0015     1.39283   5.85733
        |This is THETA(01) in class probability model.
Constant|     .54095        1.48777      .36  .7162    -2.37503   3.45693
 _HINC|1|    -.00672         .03534     -.19  .8492     -.07597    .06254
        |This is THETA(02) in class probability model.
Constant|        0.0    .....(Fixed Parameter).....
 _HINC|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.
-----------------------------------------------------------------------------
```

A possibly more flexible method of constraining the model parameters is to use

; Rst = list

This can be used generally to impose fixed value and equality constraints on the latent class model in *NLOGIT*. You must provide the full set of specifications for all *J* classes. No specifications are provided for the class probability model, which must be unrestricted. If you have *K* variables including constants in the utility model, and *J* classes, then you must provide *JK* specifications here. Note also, if you use *one* to set up the constants, keep in mind, these are put at the end of the parameter vector. If you use **; Rh2 = list**, the variables are expanded and multiplied by the ASCs. In general, it will be useful to fit the model without the **; Rst** restrictions to see how the parameters are arranged.

An example that illustrates this would be

```
LCLOGIT    ; Lhs  = mode
           ; Choices = air,train,bus,car
           ; Rhs = one,gc,ttme
           ; LCM
           ; Pts = 2 $
```

To force the coefficients on *gc* and *ttme* to be the same in both classes, you could use

```
; Rst = bgc,bttme,aa1,at1,ab1,
        bgc,bttme,aa2,at2,ab2
```

This sets up the parameter vector shown in the results below. Note that the first two coefficients are the same in the two classes.

```
-----------------------------------------------------------------------------
Latent Class Logit Model
Dependent variable                    MODE
Log likelihood function       -174.42942
Restricted log likelihood     -291.12182
Chi squared [   9 d.f.]        233.38480
Significance level                .00000
McFadden Pseudo R-squared       .4008370
Estimation based on N =    210, K =   9
Inf.Cr.AIC  =    366.9 AIC/N =    1.747
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4008 .3922
Constants only   -283.7588  .3853 .3764
At start values  -199.9272  .1275 .1149
Response data are given as ind. choices
Number of latent classes =             2
Average Class Probabilities
     .612  .388
LCM model with panel has      30 groups
Fixed number of obsrvs./group=         7
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Utility parameters in latent class -->> 1
    GC|1|    -.00859*        .00498    -1.72  .0846     -.01836    .00117
  TTME|1|    -.10408***      .01704    -6.11  .0000     -.13748   -.07068
 A_AIR|1|    7.83473***     1.04467     7.50  .0000     5.78720   9.88225
A_TRAI|1|    5.71646***      .71747     7.97  .0000     4.31024   7.12268
 A_BUS|1|    3.88956***      .76829     5.06  .0000     2.38373   5.39539
        |Utility parameters in latent class -->> 2
    GC|2|    -.00859*        .00498    -1.72  .0846     -.01836    .00117
  TTME|2|    -.10408***      .01704    -6.11  .0000     -.13748   -.07068
 A_AIR|2|    4.36673***     1.09525     3.99  .0001     2.22007   6.51339
A_TRAI|2|    1.69393**       .79868     2.12  .0339      .12855   3.25932
 A_BUS|2|    2.90232***      .71358     4.07  .0000     1.50372   4.30092
        |Estimated latent class probabilities
 PrbCls1|     .61159***      .14705     4.16  .0000      .32337    .89980
 PrbCls2|     .38841***      .14705     2.64  .0083      .10020    .67663
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This would be the same as **; Fix = gc,ttme**. However, **; Rst = list** allows for more general constraints, and allows you to fix coefficients at particular values as well. For a two class model, rather little is gained over the **; Fix** specification. However, when the model contains more than two classes, it becomes possible to force coefficients to be equal across a subset of the classes, but not all of them.

N25.5 Accessing Class Specific Information

This section will describe post estimation analysis of the specific classes that your latent class model has identified. This would include, for example, computing elasticities and doing simulations. This will involve three steps:

1. Forming the partition of the data set into the estimated classes
2. Obtaining the parameters needed for the analysis from the larger set of estimated parameters
3. Computing the desired features of the data, such as the elasticities

It is not known which individual is in which class. (If it were known, the classes would not be 'latent.') The best estimate of which class an individual resides in would be taken from the posterior probabilities derived in Section N25.3. Based on these probabilities, we would use

$$c_i^* \;=\; c_i \text{ such that } \hat{\pi}^*_{c^*} = \text{maximum}_c\,(\hat{\pi}^*_1,\ldots,\hat{\pi}^*_C)\,.$$

That is, the class with the maximum posterior probability. This number is computed automatically. Use

CREATE ; classi $ (You may use any name desired.)
LCLOGIT ; … ; classp = classi $

For each row of data that applies to individual *i*, the class number with the maximum probability will appear in the indicated variable. (Note that **; classp = namelist** is also used to save the latent class probabilities. You can save both the class indicator and the probabilities by using

; classp = namelist, class variable

i.e., including an additional variable in the list. If you have only one variable name in the *classp* list, it is obvious that only the identifier is to be saved because there will always be two or more classes.

The second step in this exercise is to extract the necessary coefficients from the overall estimates from the model. If the model command is

LCLOGIT ; Lhs = …
; Rhs = K variables, possibly including a constant, one
; Pts = J ; … $

Then the vector *b* saved by the model will contain ($\mathbf{b}_1$, $\mathbf{b}_2$, …, $\mathbf{b}_J$, other). The other parameters, not needed here, will be the class probabilities or possibly coefficients used to compute the prior class probabilities. A second matrix, *b_lc* will be saved as well. The rows of *b_lc* are the class specific coefficient vector. You can extract rows of a matrix simply by using

MATRIX ; name = matrix(j,*) $

where *j* is the row you wish to extract. For your latent class model,

MATRIX ; bc1 = b_lc (1,*) $

will extract row 1 of the matrix.

The last step is to analyze the class you specify. Do this with

```
CLOGIT      ; Lhs = … (same as the LCLOGIT command)
            ; Start = the vector ; Maxit = 0
            ; … Effects: … $
```

To do a simulation analysis, use a second **CLOGIT** command after this one, but omit the **; Start…** specification.

The following command set combines the steps and illustrates the procedure:

```
CREATE      ; classid $
LCLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = gc,invc,invt,ttme ; Rh2 = one
            ; Pts = 3  ; Pds = 7
            ; classp = classid $
MATRIX      ; bc1 = b_lc(1,*) $
CLOGIT      ; If [classid = 1] ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = gc,invc,invt,ttme ; Rh2 = one
            ; Start = bc1 ; Maxit = 0
            ; Effects: invt(*)  $
SAMPLE      ; All $
MATRIX      ; bc2 = b_lc(2,*) $
CLOGIT      ; If [classid = 2] ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = gc,invc,invt,ttme ; Rh2 = one
            ; Start = bc2 ; Maxit = 0
            ; Effects: invt(*) $
```

N25.6 An Application

A latent class model based on the clogit data is estimated with the commands

```
NLOGIT      ; Lhs = mode
            ; Choices = air,train,bus,car
            ; Rhs = gc,ttme ; Rh2   = one
            ; Effects: gc(air)
            ; Crosstab
            ; Pts = 2 ; Pds = 7
            ; LCM = hinc
            ; Parameters ; List $
```

Note that we have artificially grouped the sample into 30 groups of seven observations. This is the model that was fit as an MNL model in Chapter N17. Results are shown below. The MNL model is fit first to obtain the starting values for the iterations. The results for the latent class model are given next.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -199.97662
Estimation based on N =    210, K =   5
Inf.Cr.AIC  =    410.0 AIC/N =    1.952
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2816
Chi-squared[ 2]          =    167.56429
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
--------+--------------------------------------------------------------------
    GC|1|    -.01578***      .00438    -3.60  .0003     -.02437   -.00719
  TTME|1|    -.09709***      .01044    -9.30  .0000     -.11754   -.07664
 A_AIR|1|    5.77636***      .65592     8.81  .0000     4.49078   7.06193
A_TRAI|1|    3.92300***      .44199     8.88  .0000     3.05671   4.78929
 A_BUS|1|    3.21073***      .44965     7.14  .0000     2.32943   4.09204
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:  20 iterations. Status=0, F=    158.5813

-----------------------------------------------------------------------------
Latent Class Logit Model
Dependent variable                 MODE
Log likelihood function      -158.58128
Restricted log likelihood    -291.12182
Chi squared [  12 d.f.]       265.08108
Significance level               .00000
McFadden Pseudo R-squared      .4552752
Estimation based on N =    210, K =  12
Inf.Cr.AIC  =    341.2 AIC/N =    1.625
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4553 .4447
Constants only   -283.7588  .4411 .4303
At start values  -199.9783  .2070 .1916
Response data are given as ind. choices
Number of latent classes =            2
Average Class Probabilities
     .573  .427
LCM model with panel has      30 groups
Fixed number of obsrvs./group=        7
Number of obs.=   210, skipped    0 obs
```

```
---------+---------------------------------------------------------------------
         |                  Standard            Prob.      95% Confidence
     MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+----------------------------------------------------------------------
         |Utility parameters in latent class -->> 1
     GC|1|    -.01480*         .00764    -1.94  .0528     -.02977    .00018
   TTME|1|    -.18597***       .02737    -6.79  .0000     -.23961   -.13233
  A_AIR|1|    9.67515***      1.77945     5.44  .0000     6.18750  13.16280
 A_TRAI|1|    5.39833***       .98043     5.51  .0000     3.47672   7.31995
  A_BUS|1|    6.02787***      1.01332     5.95  .0000     4.04181   8.01394
         |Utility parameters in latent class -->> 2
     GC|2|    -.01286**        .00635    -2.02  .0429     -.02531   -.00041
   TTME|2|    -.04842***       .01652    -2.93  .0034     -.08080   -.01605
  A_AIR|2|    6.25612***      1.31406     4.76  .0000     3.68061   8.83163
 A_TRAI|2|    5.51199***      1.06768     5.16  .0000     3.41938   7.60461
  A_BUS|2|    3.62297***      1.13691     3.19  .0014     1.39467   5.85126
         |This is THETA(01) in class probability model.
Constant|     .53114          1.47670      .36  .7191    -2.36313   3.42542
  _HINC|1|    -.00653          .03508     -.19  .8524     -.07529    .06224
         |This is THETA(02) in class probability model.
Constant|        0.0    .....(Fixed Parameter).....
  _HINC|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.
-------------------------------------------------------------------------------
+---------------------------------------------------------+
| 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|           40             12             4             3            58
   TRAIN|           12             44             4             3            63
     BUS|            2              4            20             4            30
     CAR|            5              4             6            44            59
--------+-----------------------------------------------------------------------
   Total|           59             64            34            53           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|           43             15             0             0            58
   TRAIN|           11             52             0             0            63
     BUS|            0              6            23             1            30
     CAR|            0              0             0            59            59
--------+-----------------------------------------------------------------------
   Total|           54             73            23            60           210
```

N25.7 The 2^K Model

Section N18.9 describes a situation in which some individuals in a sample explicitly indicate that they ignored certain attributes. To consider a simple example (using our clogit data as a backdrop), assume the model were

$$U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \beta_1\, gc + \beta_2\, invt + \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$$

This defines the utility functions for an individual in the sample. Suppose some individuals indicate that they did not consider the in-vehicle time, *invt*, in their decision. Then, for those individuals, the appropriate utility functions are

$$U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \beta_1\, gc + \qquad \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$$

That is, the appropriate adjustment is to force the coefficient on *invt* to equal zero for those individuals. That is what *NLOGIT* does internally when the -888 value is used as described in Section N18.9.

We now consider the possibility that individuals do ignore certain attributes, but we do not know explicitly who ignores which one or both, or neither. Suppose that attributes *gc* and *invt* are involved. (See Hensher, Rose and Greene (2011).) The description suggests a latent class model such as

Class 1 $U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \beta_1\, gc + \beta_2\, invt + \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$
Class 2 $U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \qquad \beta_2\, invt + \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$
Class 3 $U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \beta_1\, gc + \qquad \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$
Class 4 $U(air,train,bus,car) = <\alpha a,\alpha t,\alpha b,0> + \qquad \beta_3\, invc + <\varepsilon_a\ \varepsilon_t\ \varepsilon_b\ \varepsilon_c>.$

If there are K attributes being treated this way, then the latent class model has 2^K classes – hence the name of the model.

The command structure for this model modifies the **LCLOGIT** command as follows:

```
LCLOGIT    ; Lhs = choice variable
           ; Choices = choice set definition
           ; Rhs = x1, x2, …, xK, …, other xs
           ; Rh2 = variables interacted with ASCs
           ; LCM or ; LCM = list of variables
           ; Pds = panel data setup if any
           ; Pts = 102 or 103 or 104 $
```

The number of classes is set up from the **; Pts** specification, which specifies K as the third digit. This is also the number of variables at the beginning of the Rhs list that will be analyzed in this model. The number of such variables may be 2, 3, or 4. With four attributes, there will be 16 classes. The following specifies a $2^2 = 4$ class model:

```
LCLOGIT    ; Lhs = mode
           ; Choices = air,train,bus,car
           ; Rhs = gc,invt,invc
           ; Rh2 = one,hinc
           ; Pts = 102 $
```

The results below illustrate the estimator. Note that the coefficients are assumed to be the same across classes. The results suggest that the data do not contain evidence that individuals ignored only *gc*, however quite a large fraction appeared to have ignored both *gc* and *invt*. (That is how the results would be interpreted. Since we have artificially grouped the observations, the results are only illustrative.)

```
-----------------------------------------------------------------------------
Endog. Attrib. Choice LC Model
Dependent variable                   MODE
Log likelihood function      -223.79636
LCM model with panel has      30 groups
Fixed number of obsrvs./group=        7
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Utility parameters in latent class -->> 1
    GC|1|     .04888***      .01633     2.99  .0028      .01687    .08089
  INVT|1|    -.01255***      .00117   -10.74  .0000     -.01484   -.01026
  INVC|1|    -.04514***      .00726    -6.22  .0000     -.05937   -.03091
 A_AIR|1|    -.85055         .53993    -1.58  .1152    -1.90881    .20770
A_TRAI|1|    1.14329***      .23175     4.93  .0000      .68908   1.59751
 A_BUS|1|     .01526         .29490      .05  .9587     -.56273    .59325
        |Utility parameters in latent class -->> 2
    GC|2|     .04888***      .01633     2.99  .0028      .01687    .08089
  INVT|2|        0.0    .....(Fixed Parameter).....
  INVC|2|    -.04514***      .00726    -6.22  .0000     -.05937   -.03091
 A_AIR|2|    -.85055         .53993    -1.58  .1152    -1.90881    .20770
A_TRAI|2|    1.14329***      .23175     4.93  .0000      .68908   1.59751
 A_BUS|2|     .01526         .29490      .05  .9587     -.56273    .59325
        |Utility parameters in latent class -->> 3
    GC|3|        0.0    .....(Fixed Parameter).....
  INVT|3|    -.01255***      .00117   -10.74  .0000     -.01484   -.01026
  INVC|3|    -.04514***      .00726    -6.22  .0000     -.05937   -.03091
 A_AIR|3|    -.85055         .53993    -1.58  .1152    -1.90881    .20770
A_TRAI|3|    1.14329***      .23175     4.93  .0000      .68908   1.59751
 A_BUS|3|     .01526         .29490      .05  .9587     -.56273    .59325
        |Utility parameters in latent class -->> 4
    GC|4|        0.0    .....(Fixed Parameter).....
  INVT|4|        0.0    .....(Fixed Parameter).....
  INVC|4|    -.04514***      .00726    -6.22  .0000     -.05937   -.03091
 A_AIR|4|    -.85055         .53993    -1.58  .1152    -1.90881    .20770
A_TRAI|4|    1.14329***      .23175     4.93  .0000      .68908   1.59751
 A_BUS|4|     .01526         .29490      .05  .9587     -.56273    .59325
        |Estimated latent class probabilities
 PrbCls1|     .39047**       .15245     2.56  .0104      .09167    .68927
 PrbCls2|        0.0         .18530      .00 1.0000 -.36318D+00  .36318D+00
 PrbCls3|     .14715         .13433     1.10  .2733     -.11613    .41042
 PrbCls4|     .46238***      .15199     3.04  .0023      .16448    .76028
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

N25.8 Individual Results

The components of the latent class model are the prior class probabilities, π_{ic} and the conditional choice probabilities, P(j|c). The posterior estimates of the class probabilities are

$$\hat{\pi}^*_{ic} = \frac{\hat{\pi}_{ic}\hat{P}_i(j\,|\,c)}{\sum_{c=1}^{C}\hat{\pi}_{ic}\hat{P}_i(j\,|\,c)}.$$

These revised probabilities are used to compute individual specific estimates of β as well as the elasticities and willingness to pay measures. The model below is estimated with the commands

```
CREATE      ; p1,p2 $
NAMELIST    ; pc = p1,p2 $
LCLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = invc,invt,gc ; Rh2 = one,hinc
            ; Effects: invc(*) ; Full
            ; Pts = 2 ; Pds = 7
            ; WTP = invt/invc ; par ; Classp = pc $
```

```
-----------------------------------------------------------------------------
Latent Class Logit Model
Dependent variable                     MODE
Log likelihood function          -188.36102
LCM model with panel has          30 groups
Fixed number of obsrvs./group=            7
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Utility parameters in latent class -->> 1
  INVC|1|    -.22612**        .09582    -2.36  .0183     -.41393   -.03832
  INVT|1|    -.03557***       .01340    -2.65  .0079     -.06183   -.00931
    GC|1|     .18821**        .09323     2.02  .0435      .00548    .37094
 A_AIR|1|   -7.88152***      2.87482    -2.74  .0061   -13.51607  -2.24696
AIR_HI|1|    -.01646          .04148     -.40  .6915     -.09776    .06484
A_TRAI|1|    2.60857***       .65694     3.97  .0001     1.32100   3.89615
TRA_HI|1|    -.03867**        .01650    -2.34  .0191     -.07101   -.00634
 A_BUS|1|     .80457          .75708     1.06  .2879     -.67928   2.28842
BUS_HI|1|    -.02065          .01994    -1.04  .3003     -.05974    .01843
        |Utility parameters in latent class -->> 2
  INVC|2|     .00105          .03516      .03  .9762     -.06787    .06997
  INVT|2|    -.00869*         .00471    -1.84  .0654     -.01793    .00055
    GC|2|     .01163          .03222      .36  .7181     -.05152    .07478
 A_AIR|2|   -2.30014**       1.14299    -2.01  .0442    -4.54036   -.05992
AIR_HI|2|     .01813          .02112      .86  .3905     -.02326    .05952
A_TRAI|2|    1.60981*         .82931     1.94  .0522     -.01562   3.23523
TRA_HI|2|    -.02850          .02289    -1.24  .2132     -.07336    .01637
 A_BUS|2|    1.31031          .88693     1.48  .1396     -.42804   3.04865
BUS_HI|2|    -.02545          .02508    -1.01  .3103     -.07461    .02372
        |Estimated latent class probabilities
 PrbCls1|     .52595***       .09375     5.61  .0000      .34219    .70970
 PrbCls2|     .47405***       .09375     5.06  .0000      .29030    .65781
--------+--------------------------------------------------------------------
```

N25.8.1 Parameters

A best guess of the parameter vector for each individual can be computed using

$$E[\boldsymbol{\beta}|\text{choices}] = \sum_{c=1}^{C} \hat{\pi}_{ic}^{*} \hat{\boldsymbol{\beta}}_c$$

The results for the model estimated above are shown in Figure N25.1. (Note that we have artificially grouped the sampled individuals into panels of seven observations for this example.)

Matrix - BETA_I

[30, 9] Cell: -0.0214444

	1	2	3	4	5	6	7	8	9
1	-0.225462	-0.0354912	0.187698	-7.86528	-0.0163575	2.60567	-0.0386445	0.80604	-0.0206689
2	-0.225916	-0.035545	0.188051	-7.87645	-0.0164267	2.60767	-0.0386649	0.805028	-0.0206593
3	-0.225962	-0.0355504	0.188087	-7.87757	-0.0164337	2.60787	-0.0386669	0.804926	-0.0206584
4	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
5	-0.216539	-0.0344353	0.180763	-7.64606	-0.0149988	2.56644	-0.0382448	0.825904	-0.0208572
6	0.00104903	-0.00868708	0.0116315	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
7	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
8	0.00104893	-0.00868709	0.0116315	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
9	-0.226122	-0.0355694	0.188212	-7.88151	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
10	-0.226122	-0.0355694	0.188212	-7.88152	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
11	-0.226121	-0.0355693	0.188211	-7.88149	-0.0164579	2.60857	-0.0386741	0.804572	-0.020655
12	-0.226122	-0.0355694	0.188212	-7.88151	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
13	0.000167329	-0.00879141	0.0123168	-2.3218	0.0179992	1.61369	-0.0285353	1.30835	-0.0254308
14	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
15	-0.226122	-0.0355694	0.188212	-7.88152	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
16	-0.226122	-0.0355694	0.188212	-7.88152	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
17	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
18	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
19	-0.226122	-0.0355694	0.188212	-7.88152	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
20	-0.226122	-0.0355694	0.188212	-7.88152	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
21	-0.226122	-0.0355694	0.188212	-7.88151	-0.0164581	2.60857	-0.0386741	0.804569	-0.020655
22	0.00104545	-0.0086875	0.0116343	-2.30023	0.0181329	1.60983	-0.0284959	1.3103	-0.0254493
23	0.00104906	-0.00868707	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
24	-0.223058	-0.0352068	0.18583	-7.80623	-0.0159915	2.5951	-0.0385368	0.811391	-0.0207197
25	-0.225969	-0.0355512	0.188093	-7.87775	-0.0164347	2.6079	-0.0386673	0.804911	-0.0206582
26	0.00104721	-0.00868729	0.0116329	-2.30018	0.0181331	1.60982	-0.0284959	1.3103	-0.0254494
27	0.00104863	-0.00868713	0.0116318	-2.30015	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
28	0.00104881	-0.0086871	0.0116316	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
29	0.00104906	-0.00868708	0.0116314	-2.30014	0.0181334	1.60981	-0.0284958	1.31031	-0.0254494
30	-0.188718	-0.0311432	0.159138	-6.96254	-0.0107626	2.44413	-0.0369983	0.887839	-0.0214444

Matrix - CLASSP_I

[30, 2] Cell: 0.997092

	1	2
1	0.997092	0.00290835
2	0.999092	0.000908088
3	0.999294	0.000706096
4	1.98699e-016	1
5	0.957814	0.0421859
6	1.27782e-007	1
7	9.5902e-024	1
8	5.69014e-007	0.999999
9	1	1.31234e-007
10	1	1.71134e-008
11	0.999995	4.98321e-006
12	1	2.17162e-007
13	0.00388135	0.996119
14	6.86801e-012	1
15	1	3.01304e-009
16	1	4.78307e-009
17	2.926e-015	1
18	8.93698e-010	1
19	1	1.08105e-008
20	1	1.20904e-009
21	0.999999	5.24265e-007
22	1.59112e-005	0.999984
23	1.03712e-022	1
24	0.986511	0.0134887
25	0.999325	0.000675306
26	8.14358e-006	0.999992
27	1.91005e-006	0.999998
28	1.09056e-006	0.999999
29	2.91805e-009	1
30	0.835349	0.164651

Figure N25.1 Estimated Posterior Probabilities and Parameters

N25.8.2 Willingness to Pay

The latent class model can also compute the estimated willingness to pay measure for each individual in the sample based on the preceding estimates of their parameters. The model request is identical to that used for the random parameters model. With the **LCLOGIT** command, use

; Par ; WTP = parameter1 / parameter2

where the two parameters are identified by variable name if you have used **; Rhs = list** to specify the utility functions or parameter names if you have used **; Model:** to specify utility functions. The latent class estimator computes the mean, *wtp_i*. In general, the WTP calculation will have an attribute level coefficient in the numerator and a cost or income measure in the denominator.

In the example above, we have specified

; Par ; WTP = invt / invc

to estimate the willingness to pay for a shorter trip. Results are shown below. The WTP values are shown in the rightmost column. The posterior probabilities are shown at the left and the posterior estimates of β are shown in the center. Note that WTP appears to have the wrong sign for some of the individuals. This is a consequence of the *invc* parameter having the wrong sign in class 2 in the estimated model. When the posterior probability is high (or one) for class 2, this estimate gets a dominant weight in the result. This suggests the consequence of a badly specified model, which our numerical illustration here seems to exemplify.

Matrix - CLASSP_I   [30, 2] Cell: 0.997092

	1	2
1	0.997092	0.00290835
2	0.999092	0.000908088
3	0.999294	0.000706096
4	1.98699e-016	1
5	0.957814	0.0421859
6	1.27782e-007	1
7	9.5902e-024	1
8	5.69014e-007	0.999999
9	1	1.31234e-007
10	1	1.71134e-008
11	0.999995	4.98321e-006
12	1	2.17162e-007
13	0.00388135	0.996119
14	6.86801e-012	1
15	1	3.01304e-009
16	1	4.78307e-009
17	2.926e-015	1
18	8.93698e-010	1
19	1	1.08105e-008
20	1	1.20904e-009
21	0.999999	5.24265e-007
22	1.59112e-005	0.999984
23	1.03712e-022	1
24	0.986511	0.0134887
25	0.999325	0.000675306
26	8.14358e-006	0.999992
27	1.91005e-006	0.999998
28	1.09056e-006	0.999999
29	2.91805e-009	1
30	0.835349	0.164651

Matrix - BETA_I   [30, 9] Cell: 0.804569

	1	2
1	-0.225462	-0.0354912
2	-0.225916	-0.035545
3	-0.225962	-0.0355504
4	0.00104906	-0.00868707
5	-0.216539	-0.0344353
6	0.00104903	-0.00868708
7	0.00104906	-0.00868707
8	0.00104893	-0.00868709
9	-0.226122	-0.0355694
10	-0.226122	-0.0355694
11	-0.226121	-0.0355693
12	-0.226122	-0.0355694
13	0.000167329	-0.00879141
14	0.00104906	-0.00868707
15	-0.226122	-0.0355694
16	-0.226122	-0.0355694
17	0.00104906	-0.00868707
18	0.00104906	-0.00868707
19	-0.226122	-0.0355694
20	-0.226122	-0.0355694
21	-0.226122	-0.0355694
22	0.00104545	-0.0086875
23	0.00104906	-0.00868707
24	-0.223058	-0.0352068
25	-0.225969	-0.0355512
26	0.00104721	-0.00868729
27	0.00104863	-0.00868713
28	0.00104881	-0.0086871
29	0.00104906	-0.00868708

Matrix - WTP_I   [30, 1] Cell: 0.157416

	1
1	0.157416
2	0.157337
3	0.157329
4	-8.28081
5	0.159026
6	-8.28104
7	-8.28081
8	-8.28184
9	0.157302
10	0.157302
11	0.157302
12	0.157302
13	-52.5396
14	-8.28081
15	0.157302
16	0.157302
17	-8.28081
18	-8.28081
19	0.157302
20	0.157302
21	0.157302
22	-8.30985
23	-8.28081
24	0.157837
25	0.157328
26	-8.29564
27	-8.28428
28	-8.28279
29	-8.28081
30	0.165025

Figure N25.2 Willingness to Pay Values and Posterior Probabilities

The WTP values are saved in the matrix *wtp_i* as shown in Figure N25.2. You may also expand the matrix into variable(s) in the data set as follows:

1. Use **CREATE** or **NAMELIST ; (New) ; …** to create the variable or variables if more than one.
2. Change **; WTP = definition** to **; WTP (variable or namelist) = definition**.

For example, the following will create a new variable, *invtwtp* with the matrix *wtp_i*:

```
CREATE       ; invtwtp $
RPLOGIT …    ; WTP (invtwtp) = invt / invc
```

N25.8.3 Elasticities

Elasticities and partial effects are computed using the posterior estimate of β_i as shown above. The IIA assumptions apply within the classes. However, the mixed model has a different posterior estimate of β for each individual, so the assumptions do not extend to the latent class model as averaged across individuals. The elasticities for the corresponding MNL model are shown below for comparison.

```
--------------------------------------------------------------------------------
Average elasticity      of prob(alt) wrt INVC      in AIR
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|   Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AIR|   -3.64460***      .52294    -6.97  .0000    -4.66954  -2.61966
   TRAIN|     .45876***      .10100     4.54  .0000      .26080    .65673
     BUS|     .72156***      .20005     3.61  .0003      .32946   1.11365
     CAR|    1.12303***      .31908     3.52  .0004      .49765   1.74840
--------+-----------------------------------------------------------------------
Average elasticity      of prob(alt) wrt INVC      in TRAIN
     AIR|     .48297***      .10649     4.54  .0000      .27426    .69169
   TRAIN|   -4.77019***      .39592   -12.05  .0000    -5.54618  -3.99419
     BUS|    1.66479***      .13855    12.02  .0000     1.39322   1.93635
     CAR|    2.20779***      .17851    12.37  .0000     1.85791   2.55768
--------+-----------------------------------------------------------------------
Average elasticity      of prob(alt) wrt INVC      in BUS
     AIR|     .33890***      .07881     4.30  .0000      .18444    .49337
   TRAIN|     .78575***      .08516     9.23  .0000      .61884    .95265
     BUS|   -3.77555***      .27281   -13.84  .0000    -4.31024  -3.24086
     CAR|    1.13704***      .13530     8.40  .0000      .87186   1.40223
--------+-----------------------------------------------------------------------
Average elasticity      of prob(alt) wrt INVC      in CAR
     AIR|     .43901***      .07427     5.91  .0000      .29345    .58457
   TRAIN|     .94205***      .08747    10.77  .0000      .77061   1.11348
     BUS|    1.14673***      .13308     8.62  .0000      .88590   1.40756
     CAR|   -2.50537***      .27334    -9.17  .0000    -3.04111  -1.96962
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

```
--------+------------------------------------
INVC    |      AIR     TRAIN       BUS       CAR
--------+------------------------------------
     AIR| -3.6446     .4588     .7216    1.1230
   TRAIN|   .4830   -4.7702    1.6648    2.2078
     BUS|   .3389     .7857   -3.7755    1.1370
     CAR|   .4390     .9420    1.1467   -2.5054
```

(Multinomial Logit Model)

```
--------+------------------------------------
INVC    |      AIR     TRAIN       BUS       CAR
--------+------------------------------------
     AIR| -2.7340    1.1983    1.1983    1.1983
   TRAIN|   .5536   -1.8144     .5536     .5536
     BUS|   .2104     .2104   -1.3328     .2104
     CAR|   .2241     .2241     .2241    -.7443
```

N25.8.4 Accessing Individual Coefficients After Estimation

The estimated matrix *beta_i* contains a row for each individual containing the ‘posterior’ estimates of the model parameters,

$$E[\boldsymbol{\beta}|\text{choices}] = \sum_{c=1}^{C} \hat{\pi}^*_{ic}\hat{\boldsymbol{\beta}}_c$$

An example appears in Figure N25.1. You can access *beta_i* like any other matrix. One use might be to compute individual specific results using the posterior (conditional) estimates of the parameters, $\hat{\boldsymbol{\beta}}_i$. If the sample contains N individuals in total and each individual has T_i choice situations, then *beta_i* will contain N rows, one for each individual and one column for each attribute in the model. The same estimated parameter vector, $\hat{\boldsymbol{\beta}}_i$, would be used for any of the T_i choice situations. For example, the following estimates a four class latent class model with each respondent having three choice situations. There are 210 choice situations in the sample, so $N = 210/3 = 70$.

```
CREATE      ; ahinc = aasc*hinc ; thinc = tasc*hinc ; bhinc = basc*hinc $
NAMELIST    ; xc = invc,invt,gc,aasc,ahinc,tasc,thinc,basc,bhinc $
LCLOGIT     ; Lhs = mode ; Choices = air,train,bus,car ; Rhs = xc
            ; Pts = 2 ; Pds = 7 ; Parameters $
```

(The interactions are computed separately because the choice models cannot use constructed variables in the attribute labels. The Rhs part of the preceding could be replaced with **; Rhs = invc,invt,gc ; Rh2 = one,hinc.**)

The resulting matrix, *beta_i* contains 70 rows and seven columns. The **CREATE** function,

Mbx(matrix, variable, namelist) = index function

is provided. The function works as follows: The variable *i* indicates all the rows of the data set that provide utility functions for individual *i*. Here, there are 7×4 = 28, so

```
CREATE      ; i = Trn(28,0) $
```

The variable *u* uses the Mbx(…) function to create the utility values. Each row of **x** data for individual is multiplied by the *i*th row of *beta_i*. The variables are defined by the namelist, here *xc*.

CREATE ; u = Mbx(beta_i, i, xc) $

Now that the utilities are created, we use them to compute multinomial logit probabilities. In this command, the choice set size is fixed at 4. If it varies by individual, then the '4' would be replaced with the name of a variable that gives for each choice situation, the number of choices in that choice situation.

CREATE ; lcp = Mnl_Probs(u,Set=4) $

Note, finally, the preceding shows an example that uses the Mbx(…) and Mnl_Probs(…) functions. For the particular application, in the latent class model setting, you would get the identical results with

LCLOGIT ; Lhs = mode ; Choices = air,train,bus,car ; Rhs = xc
; Pts = 2 ; Pds = 7 ; Parameters ; Prob = lcp ; Utility = u $

You might be interested in using the conditional, observation specific parameter estimates from the latent class model. You can request them to be placed in the data area with the following sequence of steps:

1. Create a new set of variables, one for each parameter that will be saved.
2. Create a namelist for these variables.
3. Use **; Par = namelist name** in the command.

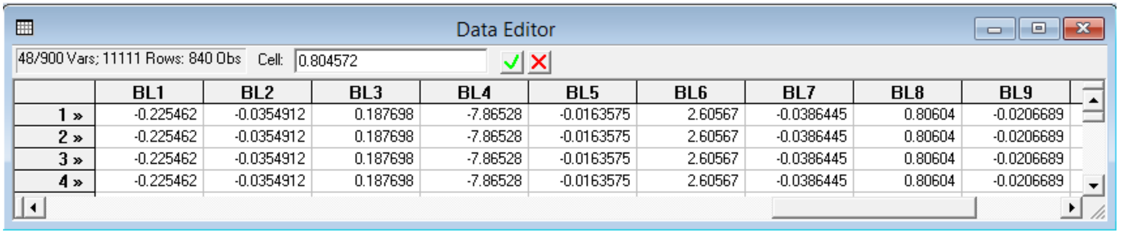
The following example demonstrates: The model in the preceding section contains nine variables on the Rhs/Rh2,

invc, *invt*, *gc*, *air_asc*, *air_asc*hinc*, *trainasc*, *trainasc*hinc*, *busasc*, *busasc*hinc*.

(The interactions are created by the Rh2 specification.) The following saves the nine coefficients for each individual – they are repeated for each row of data for the individual).

NAMELIST ; (new) ; betalcm = bl1,bl2,bl3,bl4,bl5,bl6,bl7,bl8,bl9 $
LCLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = invc,invt,gc ; Rh2 = one,hinc ; Pts = 2 ; Pds = 7
; Par = betalcm $

The resulting new data for the first individual are shown below. Note that since **; Pds = 7**, there are 4*7 = 28 identical rows of coefficients. The first four are shown in Figure N25.3.


Data Editor

48/900 Vars; 11111 Rows: 840 Obs Cell: 0.804572

	BL1	BL2	BL3	BL4	BL5	BL6	BL7	BL8	BL9
1 »	-0.225462	-0.0354912	0.187698	-7.86528	-0.0163575	2.60567	-0.0386445	0.80604	-0.0206689
2 »	-0.225462	-0.0354912	0.187698	-7.86528	-0.0163575	2.60567	-0.0386445	0.80604	-0.0206689
3 »	-0.225462	-0.0354912	0.187698	-7.86528	-0.0163575	2.60567	-0.0386445	0.80604	-0.0206689
4 »	-0.225462	-0.0354912	0.187698	-7.86528	-0.0163575	2.60567	-0.0386445	0.80604	-0.0206689

Figure N25.3 Individual Specific Coefficients

We will now use these commands to compute the choice probabilities

```
NAMELIST    ; xlc = invc,invt,gc,aasc,aasc*hinc,tasc,tasc*hinc,basc,basc*hinc $
CREATE      ; utility = betalcm ' xlc $
CREATE      ; lcmprob = Mnl_Probs(utility,Set=4) $
```

(Note that these probabilities would be obtained by simply adding **; Prob = lcmprob** in the model command. The preceding illustrates how to access the coefficients.)

N25.9 Scaled Latent Class Model

The latent class models defined thus far differentiate the classes based on the taste coefficients, **β**. A variant of the model might be specified in which the classes are also defined on the basis of the scaling of the random component of the utility functions. Thus,

$$U_{ij}|(\text{class} = c) = \beta_c' x_{ij} + \sigma_c \varepsilon_{ij}.$$

In the MNL case, the parameter σ_c is not identified, and is normalized at 1 – i.e., not estimated. The same would be true in the latent class model. The model as stated is not identified. However, it is estimable if the taste coefficient vector is restricted. For example, in a two class model, if any of the βs are equal across the classes, then one of the σ_c is estimable. In that instance, only one of them is normalized. A pure 'scaling' model would have $\beta_1 = \beta_2 = \ldots = \beta_c$ and the σ parameters would vary freely. The implied model would be

$$U_{ij}|(\text{class} = c) = \beta_c' x_{ij} + \sigma_c \varepsilon_{ij}.$$

You can construct the scaling model from the LC model by using

; SLC

then providing the restrictions on the βs. The **; SLC** specification adds the scale parameters to the model (normalizing the last one to 1.0), then the restrictions are provided with

; Rst = list.

For example a pure scaling, two class model would appear as follows:

```
LCLOGIT     ; Lhs = mode ; Choices = air,train,bus,car ; Rhs = gc,invc,invt,ttme
            ; SLC ; Rst = b1,b2,b3,b4, b1,b2,b3,b4
            ; Pts = 2 ; Pds = 3 $
```

(We specify the **; Pds = 3** situations for this example only.)

N25.10 Random Regret Latent Class Model

The latent class model specified to this point defines a multinomial logit, random utility model in each class. You can switch to a random regret formulation (see Section N23.7.1) by changing the model command to

LCRRLOGIT ; … as before for the latent class model $

Each segment will now be constructed using the random regret formulation. The model command

LCLOGIT ; Rrm ; … as before $

is the same. With this construction, you can also build a latent class model in which some classes are built around the random utility (RUM) framework and others are built around the random regret framework. To specify a hybrid model of this sort, use

LCLOGIT ; …
; Pts = number of classes in total, j
; Regret = number of random regret classes, q $

Note that *j* must be strictly larger than *q*. For example, an interesting model that could reveal a partitioning of the population would be implied by

; Pts = 2
; Regret = 1.

N25.11 Technical Details

The log likelihood function for this model is the sum of the logs of P_{ji} as given in Section N25.3. The log likelihood function is maximized directly using *NLOGIT*'s general optimization package. Applications in the literature have suggested the EM method as a preferable approach, but we have not found this to be the case.

(In addition, the EM algorithm does not allow the imposition of cross class restrictions, such as those used to form the 2^K model.) The estimated asymptotic covariance matrix is based on the second derivatives. If the latent class parameters are not precisely estimated, because of rounding error, this matrix may fail to be positive definite. In this case, the BHHH estimator is used instead. Starting values for the iterations are obtained by assuming the classes are equally probable, but the class specific (bold beta) vectors differ slightly from the MNL estimates. If they and the class probabilities are assumed to be equal, then all derivatives of the log likelihood will equal zero. This is a local maximizer of the log likelihood. To avoid this point, the starting MNL values are perturbed slightly.

Within the class, choice probabilities are assumed to be generated by the multinomial logit model.

$$\text{Prob}(y_{it} = j \mid \text{class} = c) = \frac{\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}{\sum_{j=1}^{J_i}\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}.$$

As noted, the class membership is not observed. Class probabilities are specified by the multinomial logit form,

$$\text{Prob(class} = c) \;=\; Q_{ic} \;=\; \frac{\exp\left(\boldsymbol{\theta}_c'\mathbf{z}_i\right)}{\sum_{c=1}^{C}\exp\left(\boldsymbol{\theta}_c'\mathbf{z}_i\right)},\; \boldsymbol{\theta}_C \;=\; \mathbf{0}.$$

where $\mathbf{z}_i$ is an optional set of person, situation invariant characteristics. The class specific probabilities may be a set of fixed constants if no such characteristics are observed. In this case, the class probabilities are simply functions of C parameters, θ_c, the last of which is fixed at zero. This model does not impose the IIA property on the observed probabilities.

For a given individual, the model's estimate of the probability of a specific choice is the expected value (over classes) of the class specific probabilities. Thus,

$$\begin{aligned}\text{Prob}(y_{it} = j) &= E_c\left[\frac{\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}{\sum_{j=1}^{J_i}\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}\right]\\ &= \sum_{c=1}^{C}\text{Prob}[class = c]\left[\frac{\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}{\sum_{j=1}^{J_i}\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}\right].\end{aligned}$$

When there are T_i choice situations, the choices are independent conditioned on the class, so

$$\begin{aligned}\text{Prob}(y_{i1} = j_1,...,y_{iT_i} = j_{T_i}) &= E_c\prod_{t=1}^{T_i}\left[\frac{\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}{\sum_{j=1}^{J_i}\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}\right]\\ &= \sum_{c=1}^{C}\frac{\exp\left(\boldsymbol{\theta}_c'\mathbf{z}_i\right)}{\sum_{s=1}^{C}\exp\left(\boldsymbol{\theta}_s'\mathbf{z}_i\right)}\prod_{t=1}^{T_i}\left[\frac{\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}{\sum_{j=1}^{J_i}\exp\left(\boldsymbol{\beta}_c'\mathbf{x}_{jit}\right)}\right].\end{aligned}$$

N26: Heteroscedastic Extreme Value Model

N26.1 Introduction

The main virtues of the heteroscedastic extreme value (HEV) model are its freedom from the IIA assumption and its allowance of differential cross elasticities among all pairs of alternatives. (See Bhat (1995) and Allenby and Ginter (1995). The algorithm and interpretation adopted in *NLOGIT* are those in Bhat's paper.) Unlike the nested logit model, the HEV model does not require prior partitioning of the choice set into mutually exclusive branches to achieve this result. The model is a random utility formulation as usual,

$$U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \varepsilon_{ij}$$

$$= V_{ij} + \varepsilon_{ij},$$

Choice j is made if $U_{ij} > U_{iq}$ for all q not equal to j.

The CDF for each ε_{ij} is the type 1 extreme value distribution with precision parameter θ_j – the scale parameter is $\sigma_j = 1/\theta_j$,

$$F(\varepsilon_{ij}) = \exp(-\exp(-\theta_j\varepsilon_{ij})).$$

The ε_{ij}s are independent, but not identically distributed – they have mean zero, but variance $\pi^2/(6\theta_j^2)$. Thus, each one has a different scale factor. For identification purposes, one of the θs is set to one. In *NLOGIT*'s estimator, this is the last one. This model does not have the IIA property of the multinomial logit model. The derivatives and elasticities of the probabilities differ across all alternatives and attributes. Elasticities and derivatives are computed with the evaluation of

$$\partial P_{ij} / \partial x_{k,iq} = (\partial P_{ij} / \partial V_{iq}) (\partial V_{iq} / \partial x_{k,iq})$$

$$= (\partial P_{ij} / \partial V_{iq}) \beta_k,$$

in which P_{ij} is the probability of the jth alternative and $x_{k,iq}$ is the kth attribute in the qth utility function (q and j may be unequal). These derivatives are discussed in the technical notes in Section N26.6.

N26.2 Command for the HEV Model

The command for this model is

HLOGIT	**; Lhs = dependent variable**
	; Choices = ... specification of the choice set
	; ... specification of utility functions
	; ... any other options $

(The alternative format, **NLOGIT ; Heteroscedastic** may be used instead.) The model is setup otherwise exactly as described in Chapters N17-N22 – this is a modification of the MNL model described in Chapter N17.

The command builder may also be used for this model by selecting Model:Discrete Choice/Multinomial Probit, HEV, RPL. The discrete choice model is defined on the Main page and the HEV format of the model is selected on the Options page. See Figures N26.1 and N26.2 for the setup of the model shown in the application in Section N26.3.

The following features of *NLOGIT* are not available for this model:

; Cprob = name	Conditional and unconditional probabilities are the same.
; Ranks	This estimator may not be based on ranks data.
; Scale ...	Data scaling is only for the nested logit model.
; IIA = list	IIA is not testable here, since it is not imposed.

In principle, one could test IIA as a restriction on the HEV model, since the restriction $\theta_j = 1$ does produce the MNL. However, this test is rather indirect, since IIA relates to more than just heteroscedasticity. The remainder of the setup is identical to the multinomial logit model. All other options are available, including

; Probs = name	to retain the predicted probabilities
; Utility = name	to retain the predicted systematic utilities

and so on.

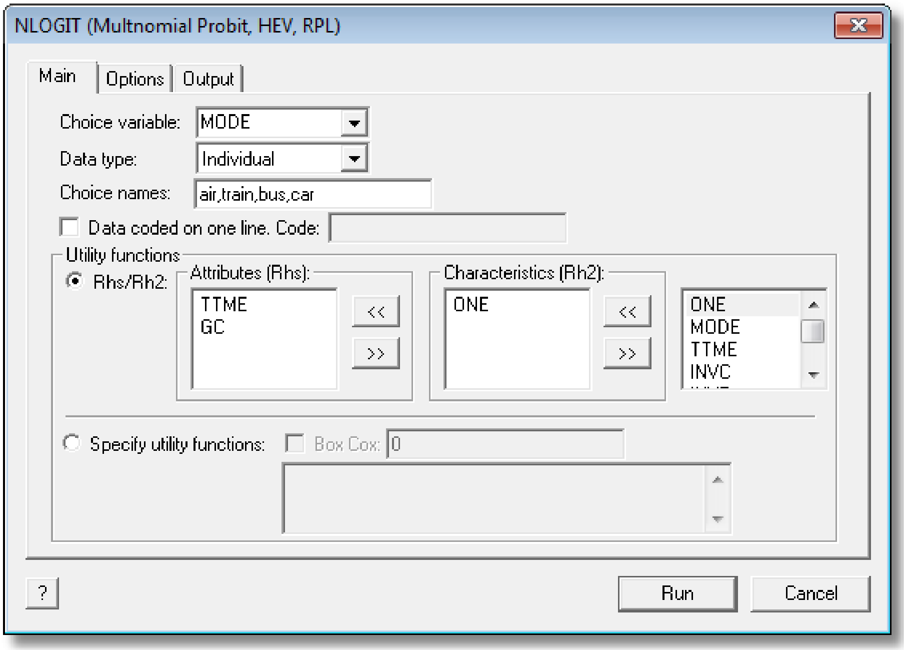


Figure N26.1 Main Page of Command Builder for the HEV Model

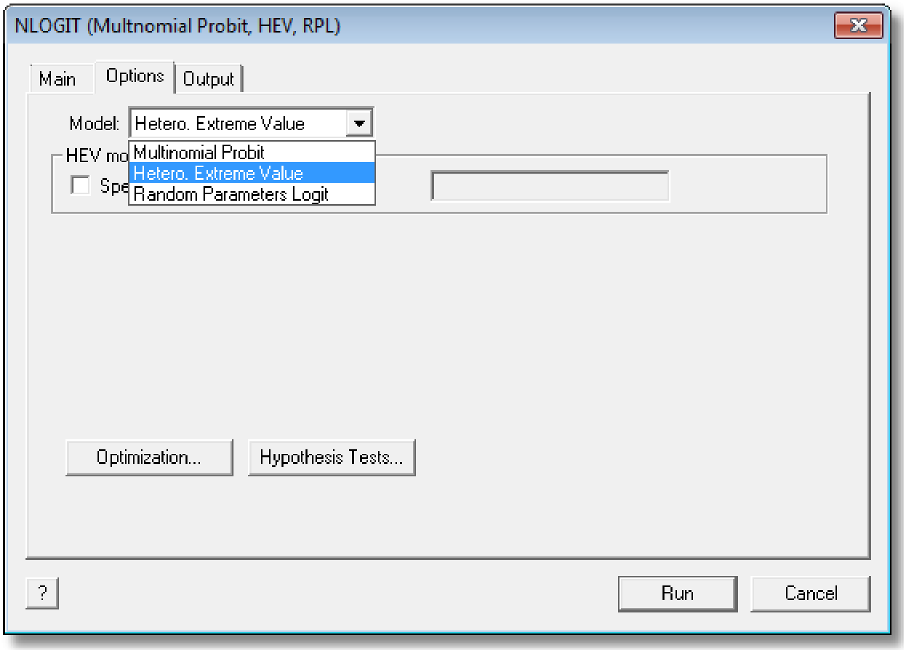


Figure N26.2 Options Page of Command Builder for the HEV Model

N26.3 Application

The HEV model based on the clogit data is estimated with the command

HLOGIT ; Lhs = mode
; Choices = air,train,bus,car
; Rhs = gc,ttme
; Rh2 = one,hinc
; Effects: gc(air) ; Lpt = 60 $

This is the model that was fit as an MNL model in Chapter N17. We have now relaxed the equal variances assumption. Results are shown below. The MNL model is fit first to obtain the starting values for the iterations. The results for the HEV model are given next.

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function      -189.52515
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    395.1 AIC/N =    1.881
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3321 .3202
Chi-squared[ 5]          =    188.46723
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
--------+--------------------------------------------------------------------
      GC|    -.01093**       .00459    -2.38  .0172     -.01992   -.00194
    TTME|    -.09546***      .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|    5.87481***      .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|    -.00537         .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|    5.54986***      .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|    -.05656***      .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|    4.13028***      .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|    -.02858*        .01544    -1.85  .0642     -.05885    .00169
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the estimates for the HEV model. Note, the scale parameters are normalized to 1.0, so the reported results show the departure from the MNL model – zero values here imply scale factors of 1.0, which are the values for MNL. The additional set of derived parameters show the implied estimates of the standard deviations of ε_j in the random utility model. The value 1.28255 is the standard deviation under the MNL assumption.

```
-----------------------------------------------------------------------------
Heteroscedastic Extreme Value Model
Dependent variable                  MODE
Log likelihood function       -181.14819
Restricted log likelihood     -291.12182
Chi squared [  11 d.f.]        219.94725
Significance level                .00000
McFadden Pseudo R-squared       .3777581
Estimation based on N =    210, K =  11
Inf.Cr.AIC  =    384.3 AIC/N =    1.830
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3778 .3667
Constants only   -283.7588  .3616 .3503
At start values  -193.7765  .0652 .0486
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
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.16389         .33857     -.48  .6283     -.82749    .49970
    TTME|   -1.03949        2.12090     -.49  .6241    -5.19638   3.11740
   A_AIR|    49.8163       102.0271      .49  .6254   -150.1531  249.7858
AIR_HIN1|     .04693         .15650      .30  .7643     -.25981    .35368
 A_TRAIN|    48.9298       99.63430      .49  .6234   -146.3499  244.2094
TRA_HIN2|    -.51323        1.16507     -.44  .6596    -2.79672   1.77025
   A_BUS|    35.1788       72.62915      .48  .6281   -107.1717  177.5293
BUS_HIN3|    -.09161         .25306     -.36  .7173     -.58759    .40437
        |Scale Parameters of Extreme Value Distns Minus 1.0
   s_AIR|    -.94107***      .11924    -7.89  .0000    -1.17477   -.70736
 s_TRAIN|    -.94110***      .13093    -7.19  .0000    -1.19771   -.68449
   s_BUS|    -.89553***      .20698    -4.33  .0000    -1.30121   -.48985
   s_CAR|        0.0    .....(Fixed Parameter).....
        |Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
   s_AIR|    21.7632       44.03379      .49  .6211    -64.5415  108.0678
 s_TRAIN|    21.7758       48.40609      .45  .6528    -73.0984  116.6500
   s_BUS|    12.2767       24.32362      .50  .6138    -35.3967   59.9501
   s_CAR|    1.28255    .....(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.
-----------------------------------------------------------------------------
```

These results compare the HEV model to the MNL. The HEV elasticities show that the IIA assumption has been relaxed. At the same time, the predictions from the two models are roughly the same.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+----------------------------------
     AIR|   -.8034     .2257     .4483     .8478
   TRAIN|    .2599  -1.0425     .4369     .9638
     BUS|    .1578     .1596  -1.6786    1.2149
     CAR|    .3800     .3701     .5630  -2.8586
```

(These are the estimated elasticities from the MNL model in Chapter N24.)

```
--------+----------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+----------------------------------
     AIR|   -.8019     .3198     .3198     .3198
   TRAIN|    .3534  -1.0693     .3534     .3534
     BUS|    .1679     .1679  -1.0916     .1679
     CAR|    .2934     .2934     .2934    -.7492
```

N26.4 Constraining the Precision Parameters

You may constrain the precision parameters to fixed values or equality. Equating groups of them to each other produces a hybrid of the heteroscedastic model and multinomial logit model. The **; Ivset:** parameter can be used for this purpose, the same as if the parameters were inclusive value parameters (see Chapter N29). The general form of the specification is

; Ivset: (group of names) = [value] /
(group of names) = [value] and so on

You may specify as many groups as desired. Of course, the lists of names must not overlap. Also, the **= [value]** is optional. If you omit it, then the precision parameters are forced to equal each other within each set, but the value is free. If **= [value]** is included, then the set of precision parameters are all forced to equal that specific value (and are not estimated.) For example, in a four outcome model, [*air*,*train*,*bus*,*car*], one might be interested in examining a partition of *private*(*air*,*car*) and *public*(*bus*,*train*) Since the fourth precision parameter (*train*) is going to be set to one (for identification), one might proceed as follows:

; Ivset: (air,car) / (bus) = [1] $

One of the precision parameters in the model must be normalized at 1.0. At the outset, *NLOGIT* does this by constraining the last variance to equal 1.0. Since your **; Ivset:** specification sets a different variance to 1.0, *NLOGIT* accepts this as renormalizing the model on this alternative instead of the last one. In this instance, given this specification, the normalized choice becomes *bus* instead of *car*. This is shown in the example below, which is produced by this specification. The crucial point is that for identification, at least one restriction must be placed on the variances in the HEV model. If you specify a restriction, then the model is automatically identified by your restriction, so you can, as we did above, remove the initial normalization.

```
-----------------------------------------------------------------------------
Heteroscedastic Extreme Value Model
Dependent variable                 MODE
Log likelihood function      -188.33965
Restricted log likelihood    -291.12182
Chi squared [  10 d.f.]       205.56434
Significance level               .00000
McFadden Pseudo R-squared      .3530555
Estimation based on N =    210, K =  10
Inf.Cr.AIC  =    396.7 AIC/N =    1.889
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3531 .3426
Constants only   -283.7588  .3363 .3256
At start values  -193.7765  .0281 .0124
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
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.02138**        .01044     -2.05  .0405     -.04184   -.00093
    TTME|    -.14690***       .04848     -3.03  .0024     -.24192   -.05188
   A_AIR|    9.15848***      3.22179      2.84  .0045     2.84389  15.47308
AIR_HIN1|    -.01124          .02544      -.44  .6587     -.06111    .03863
 A_TRAIN|    9.34066***      3.05853      3.05  .0023     3.34605  15.33527
TRA_HIN2|    -.10305***       .03912     -2.63  .0084     -.17973   -.02636
   A_BUS|    7.40705**       2.96948      2.49  .0126     1.58698  13.22712
BUS_HIN3|    -.04341*         .02595     -1.67  .0944     -.09428    .00745
        |Scale Parameters of Extreme Value Distns Minus 1.0
   s_AIR|    -.49213***       .18989     -2.59  .0096     -.86430   -.11996
 s_TRAIN|    -.47456**        .20992     -2.26  .0238     -.88599   -.06313
   s_BUS|        0.0     .....(Fixed Parameter).....
   s_CAR|    -.49213***       .18989     -2.59  .0096     -.86430   -.11996
        |Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
   s_AIR|    2.52534***       .94419      2.67  .0075      .67476   4.37591
 s_TRAIN|    2.44089**        .97514      2.50  .0123      .52964   4.35214
   s_BUS|    1.28255     .....(Fixed Parameter).....
   s_CAR|    2.52534***       .94419      2.67  .0075      .67476   4.37591
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |     AIR    TRAIN       BUS       CAR
--------+------------------------------------
     AIR|  -.8535    .4320     .7673     .3032
   TRAIN|   .3659  -1.2871     .8742     .3440
     BUS|   .2208    .2258   -2.5936     .2199
     CAR|   .2675    .2849     .5769    -.7783
```

In principle, one should be able to use this device to reproduce the MNL model. For our application, we would use

; Ivset: (air,train,bus,car) = [1]

The results are reasonably close. They are not exact because even with 60 quadrature points, there is some rounding error in the Laguerre quadrature approximation to the integrals.

```
-----------------------------------------------------------------------------
Heteroscedastic Extreme Value Model
Dependent variable                  MODE
Log likelihood function       -191.32689
Restricted log likelihood     -291.12182
Chi squared [   8 d.f.]        199.58985
Significance level                .00000
McFadden Pseudo R-squared       .3427944
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    398.7 AIC/N =    1.898
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3428 .3343
Constants only   -283.7588  .3257 .3171
At start values  -193.7765  .0126-.0001
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
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.01067**       .00424    -2.52  .0119     -.01898   -.00236
    TTME|    -.08300***      .00597   -13.90  .0000     -.09470   -.07130
   A_AIR|    5.18885***      .69095     7.51  .0000     3.83462   6.54309
AIR_HIN1|    -.00608         .01289     -.47  .6369     -.03135    .01918
 A_TRAIN|    5.24358***      .61076     8.59  .0000     4.04651   6.44065
TRA_HIN2|    -.05933***      .01271    -4.67  .0000     -.08425   -.03442
   A_BUS|    3.77023***      .71256     5.29  .0000     2.37363   5.16682
BUS_HIN3|    -.03053*        .01764    -1.73  .0835     -.06512    .00405
        |Scale Parameters of Extreme Value Distns Minus 1.0
   s_AIR|        0.0    .....(Fixed Parameter).....
 s_TRAIN|        0.0    .....(Fixed Parameter).....
   s_BUS|        0.0    .....(Fixed Parameter).....
   s_CAR|        0.0    .....(Fixed Parameter).....
        |Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
   s_AIR|    1.28255    .....(Fixed Parameter).....
 s_TRAIN|    1.28255    .....(Fixed Parameter).....
   s_BUS|    1.28255    .....(Fixed Parameter).....
   s_CAR|    1.28255    .....(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.
-----------------------------------------------------------------------------
Multinomial Logit Estimates
--------+--------------------------------------------------------------------
      GC|    -.01093**       .00459    -2.38  .0172     -.01992   -.00194
    TTME|    -.09546***      .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|    5.87481***      .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|    -.00537         .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|    5.54986***      .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|    -.05656***      .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|    4.13028***      .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|    -.02858*        .01544    -1.85  .0642     -.05885    .00169
-----------------------------------------------------------------------------
```

Elasticity wrt change of X in row choice on Prob[column choice]

GC	AIR	TRAIN	BUS	CAR
AIR	-.8617	.4630	.4446	.3185
TRAIN	.4085	-1.2802	.3932	.3698
BUS	.1680	.1655	-1.2823	.1632
CAR	.2836	.2966	.2924	-.7596

These are the elasticities from the multinomial logit model.

GC	AIR	TRAIN	BUS	CAR
AIR	-.8019	.3198	.3198	.3198
TRAIN	.3534	-1.0693	.3534	.3534
BUS	.1679	.1679	-1.0916	.1679
CAR	.2934	.2934	.2934	-.7492

There is an alternative way to fix the precision parameters. Use the specification

; Sdv = list of symbols and values

This specification operates the same as **; Rst = list**. To impose fixed values, put that value in the list. For example, the preceding example could also be done with

; Sdv = 1,1,1,1

To allow a parameter to be unrestricted, just insert a name for it. For example, the original model is specified with

; Sdv = s1, s2, s3, 1.0

Finally, to force parameters to be equal, give them the same name. For example,

; Ivset: (air,car) / (bus) = [1]

and **; Sdv = s_aircar, s_train, 1, s_aircar**

are the same. To illustrate,

```
HLOGIT  ; Lhs  = mode
        ; Rhs  = gc,ttme ; Rh2 = one,hinc
        ; Choices = air,train,bus,car
        ; Sdv  = 1,1,v3,v4 ; Lpt = 60 $
```

produces the following results:

```
-----------------------------------------------------------------------------
Heteroscedastic Extreme Value Model
Dependent variable                  MODE
Log likelihood function       -181.12685
Restricted log likelihood     -291.12182
Chi squared [  10 d.f.]        219.98994
Significance level                .00000
McFadden Pseudo R-squared       .3778314
Estimation based on N =    210, K =  10
Inf.Cr.AIC  =    382.3 AIC/N =    1.820
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3778 .3678
Constants only   -283.7588  .3617 .3514
At start values  -193.7765  .0653 .0502
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
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.00980***      .00247    -3.96  .0001     -.01465   -.00495
    TTME|    -.06114***      .00643    -9.50  .0000     -.07375   -.04853
   A_AIR|    2.95197***      .54997     5.37  .0000     1.87405   4.02989
AIR_HIN1|     .00226         .00791      .29  .7751     -.01324    .01776
 A_TRAIN|    2.86278***      .41544     6.89  .0000     2.04853   3.67704
TRA_HIN2|    -.02996***      .00594    -5.04  .0000     -.04161   -.01831
   A_BUS|    2.06693***      .33521     6.17  .0000     1.40993   2.72393
BUS_HIN3|    -.00493         .00858     -.57  .5655     -.02175    .01188
        |Scale Parameters of Extreme Value Distns Minus 1.0
   s_AIR|        0.0    .....(Fixed Parameter).....
 s_TRAIN|        0.0    .....(Fixed Parameter).....
      V3|     .79409*        .45379     1.75  .0801     -.09531   1.68349
      V4|    15.9977       22.60142      .71  .4791    -28.3003   60.2957
        |Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
   s_AIR|    1.28255    .....(Fixed Parameter).....
 s_TRAIN|    1.28255    .....(Fixed Parameter).....
      V3|     .71487***      .18082     3.95  .0001      .36048   1.06927
      V4|     .07545         .10033      .75  .4520     -.12119    .27210
--------+--------------------------------------------------------------------
```

N26.5 Individual Heterogeneity in the Variances

The variances in the HEV model may be specified to be individually heterogeneous of the form

$$\theta_{ij} = \theta_j \exp(\boldsymbol{\gamma}'\mathbf{h}_i),$$

(save for the last one, in which $\theta_{ij} = 1$). This estimator is requested with

HLOGIT ; ... as before
; Hfn = list

For example, respecifying the earlier application with

HLOGIT ; Lhs = mode
; Choices = air,train,bus,car
; Rhs = gc,ttme ; Rh2 = one
; Hfn = hinc
; Crosstab ; Effects: gc(air) ; Lpt = 60 $

produces the results below.

```
-----------------------------------------------------------------------------
Heteroscedastic Extreme Value Model
Dependent variable                MODE
Log likelihood function     -190.28652
Restricted log likelihood   -291.12182
Chi squared [   9 d.f.]      201.67059
Significance level              .00000
McFadden Pseudo R-squared     .3463681
Estimation based on N =    210, K =   9
Inf.Cr.AIC  =    398.6 AIC/N =    1.898
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3464 .3369
Constants only   -283.7588  .3294 .3197
At start values  -217.1216  .1236 .1109
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
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.17091         .34905     -.49  .6244     -.85504     .51321
    TTME|    -.83099        1.70270     -.49  .6255    -4.16823    2.50624
   A_AIR|    40.0347       81.38851      .49  .6228   -119.4839   199.5532
 A_TRAIN|    26.0510       50.83392      .51  .6083    -73.5816   125.6837
   A_BUS|    25.6262       52.25164      .49  .6238    -76.7851   128.0375
        |Scale Parameters of Extreme Value Distributions
   s_AIR|     .05344         .11014      .49  .6275     -.16243     .26931
 s_TRAIN|     .05971         .13053      .46  .6474     -.19612     .31554
   s_BUS|     .10324         .20895      .49  .6212     -.30630     .51278
   s_CAR|        1.0    .....(Fixed Parameter).....
        |Heterogeneity in Scales of Ext.Value Distns.
    HINC|     .00492         .00387     1.27  .2029     -.00265     .01250
--------+--------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------
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 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|           29            10             6            13            58
   TRAIN|           12            33             6            12            63
     BUS|            5             6            15             4            30
     CAR|           15            13             5            26            59
--------+----------------------------------------------------------------
   Total|           62            62            32            54           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|           40             2             2            14            58
   TRAIN|            3            50             1             9            63
     BUS|            0             3            23             4            30
     CAR|            5            11             1            42            59
--------+----------------------------------------------------------------
   Total|           48            66            27            69           210

Elasticity wrt change of X in row choice on Prob[column choice]
+---------------------------------------------------+
| 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|    -.92370***      .04727   -19.54  .0000    -1.01635   -.83105
   TRAIN|     .38685***      .02448    15.80  .0000      .33887    .43483
     BUS|     .83207***      .05967    13.94  .0000      .71511    .94902
     CAR|     .45277***      .02885    15.70  .0000      .39623    .50931
--------+--------------------------------------------------------------
```

N26.6 Technical Details

The probability that choice *j* is made is

$$P_j \;=\; \text{Prob}[U_j > U_q] \text{ for all } q \text{ not equal to } j.$$

$$= \int_{-\infty}^{\infty} \prod_{q \neq j} F[\theta_q(V_j - V_q + \varepsilon_j)]\theta_j f(\theta_j \varepsilon_j) d\varepsilon_j,$$

where $f(t)$ is the density, $f(t) = \exp(-t)\exp(-\exp(-t)) = -F(t)\log(F(t))$. The probabilities and derivatives must be evaluated numerically, as there is no closed form for the integral. As Bhat notes, they can be approximated using Gauss-Laguerre quadrature. The method is discussed below.

To compute the probabilities, first make the change of variable $u_j = \exp[-\theta_j\varepsilon_j]$. Then, the probability becomes

$$P_j \;=\; \int_{-\infty}^{\infty} \prod_{q \neq j} F[\theta_q(V_j - V_q - (\log u_j)/\theta_j)]\exp(-u_j) du_j$$

$$= \int_{-\infty}^{\infty} \prod_{q \neq j} F[t(q \mid j)]\exp(-u_j) du_j$$

where, again, $F(t) = \exp(-\exp(-t))$ and $t(q|j) = \theta_q [V_j - V_q - (\log u_j)/\theta_j]$. There is no closed form for this integral. However, it can be approximated using Gauss-Laguerre quadrature. Thus, we use

$$\int_{-\infty}^{\infty} \prod_{q \neq j} F[t(q \mid j)]\exp(-u_j) du_j \approx \sum_{l=1}^{L} w_l F[\theta_q (V_j \text{-} V_q \text{ - } (\log h_l)/\theta_l)]$$

where w_l is the weight and h_l is the abscissa of the Gauss-Laguerre polynomial. We have used a 60 point approximation. (The weights and abscissas may be found in Abramovitz and Stegun (1972).) You can set the number of points in your command with **; Lpt = n**, where *n* is from 2 to 64. The commands in the examples include **; Lpt = 60**.

The derivatives of the probabilities must also be approximated. These are, for cross terms in which *m* is not equal to *j*,

$$\frac{\partial P_j}{\partial V_q} = \int_{-\infty}^{\infty} \prod_{s \neq j} F[t(s \mid j)]\theta_q \log F[t(q \mid j)]\exp(-u_j) du_j,$$

$$\frac{\partial P_j}{\partial \theta_q} = \int_{-\infty}^{\infty} \prod_{s \neq j} F[t(s \mid j)](-t(q \mid j)/\theta_q)\log F[t(q \mid j)]\exp(-u_j) du_j,$$

and, for the own terms,

$$\frac{\partial P_j}{\partial V_j} = \int_{-\infty}^{\infty} \left\{\prod_{s \neq j} F[t(s \mid j)\right\} \left\{\sum_{s \neq i} \left[-\theta_s \log F[t(s \mid j)]\right]\right\} \exp(\text{-}u_j) du_j,$$

$$\frac{\partial P_j}{\partial V_j} = \int_{-\infty}^{\infty} \left\{\prod_{s \neq j} F[t(s \mid j)\right\} \left\{\sum_{s \neq j} \left[-\theta_s \log u_j / \theta_j^2\right] \log F[t(s \mid j)]\right\} \exp(\text{-}u_j) du_j.$$

All of these are evaluated using the quadrature method. The derivatives are then used in constructing the log likelihood and the elasticities and partial (marginal) effects.

The model with heterogeneous variances,

$$\theta_{ij} = \theta_j \exp(\boldsymbol{\gamma}'\mathbf{h}_i),$$

is a straightforward extension. The functions are assembled for the purpose of computing the log likelihood and the derivatives. Then,

$$\frac{\partial P_{ij}}{\partial \theta_q} = \frac{\partial P_{ij}}{\partial \theta_{iq}} \exp(\boldsymbol{\gamma}'\mathbf{h}_i)\,,$$

where $\partial P_{ij}/\partial\theta_{iq}$ is evaluated using the expression given earlier for $\partial P_j/\partial\theta_q$. Finally,

$$\frac{\partial P_{ij}}{\partial \boldsymbol{\gamma}} = \sum_{q=1}^{J_i} \frac{\partial P_{ij}}{\partial \theta_{iq}} \theta_{iq}\mathbf{h}_i\,.$$

N27: Multinomial Probit Model

N27.1 Introduction

In the multinomial probit (MNP) model, the individual's choice among J alternatives is the one with maximum utility, where the utility functions are

$$U_{ji} = \boldsymbol{\beta}'\mathbf{x}_{ji} + \varepsilon_{ji},$$

where U_{ji} = utility of alternative j to individual i,

$\mathbf{x}_{ji}$ = union of all attributes that appear in all utility functions. For some alternatives, $x_{i,tk}$ may be zero by construction for some attribute k which does not enter their utility function for alternative j,

ε_{ji} = unobserved heterogeneity for individual i and alternative j.

The multinomial logit model specifies that ε_{ji} are draws from independent extreme value distributions (which induces the IIA condition). In the multinomial probit model, we assume that ε_{ji} are normally distributed with standard deviations Sdv[ε_{ji}] = σ_j and correlations Cor[ε_{ji}, ε_{mi}] = ρ_{jm} (the same for all individuals). Observations are independent, so Cor[ε_{ji},ε_{ms}] = 0 if i is not equal to s, for all j and m. A variation of the model allows the standard deviations and covariances to be scaled by a function of the data, which allows some heteroscedasticity across individuals.

The correlations ρ_{jm} are restricted to $-1 < \rho_{jm} < 1$, but they are otherwise unrestricted save for a necessarily normalization. The correlations is that the last row of the correlation matrix must be fixed at zero. The standard deviations are unrestricted with the exception of a normalization – two standard deviations are fixed at 1.0 – *NLOGIT* fixes the last two. In principle, up to 20 alternatives may be in the model, but our experience thus far is that this model is extremely difficult to estimate, and will usually not be estimable with a completely free correlation matrix even with only five alternatives. The difficulty increases greatly with the number of alternatives. (Imposition of constraints which may improve this situation is discussed below.)

This model may also be fit with panel data. In this case, the utility function is modified as follows:

$$U_{ji,t} = \boldsymbol{\beta}'\mathbf{x}_{jt,t} + \varepsilon_{ji,t} + v_{ji,t},$$

where 't' indexes the periods or replications. There are two formulations for $v_{ji,t}$,

Random effects $v_{ji,t} = v_{ji,s}$ (the same in all periods),

First order autoregressive $v_{ji,t} = \alpha_j v_{ji,t-1} + a_{ji,t}$.

N27.2 Model Command

This is a one level (nonnested) model. The setup is identical to the multinomial logit model with one level. To request it, use

MNPROBIT ; Lhs = ... ; Choices = ...
; Rhs = ... or ; Model: U (...) =... / U (...) = ... all as usual
; ... any other options $

(The alternative model command used in earlier versions of *NLOGIT*, **NLOGIT ; MNP** is equivalent and may be used instead.)

Options include

; Prob = name to use for estimated probabilities
; Utility = name to use for estimated utilities

and the usual other options for output, technical output, elasticities, descriptive statistics, etc. (See Chapters N17-N22 for details.) There are some special cases for this estimator:

- The number of alternatives must be fixed – it may not vary across observations.
- The choice set must be fixed.
- Choice based sampling is not supported, though you can use ordinary weights.
- Data may be individual, proportions, or frequencies.

(The second derivatives matrix is not computed for this model, so it is not possible to compute a robust covariance matrix estimator.) An additional option is

; Pts = number of replications to compute multivariate normal probabilities

Computation of multivariate normal probabilities is discussed in Section N27.9.

The following features of *NLOGIT* are *not* available for this model:

; Tree ... This is not a nested logit model.
; Ivb = name, **; Ivl = name**, **; Ivt = name** No inclusive values are computed.
; IIA = list IIA is not testable here, since it is not imposed.
; Cprob = name Conditional and unconditional probabilities are the same.
; Ranks This estimator may not be based on ranks data.
; Scale ... Data scaling is only for the nested logit model.

The command builder may also be used for this model by selecting Model/Discrete Choice/Multinomial Probit, HEV, RPL. The choice set and utility functions for the model are defined on the Main page and the MNP format of the model is selected on the Options page.

N27.3 An Application

The multinomial probit model based on the clogit data is estimated with the command

MNPROBIT ; Lhs = mode
; Choices = air,train,bus,car
; Rhs = gc,ttme
; Rh2 = one,hinc
; Effects: gc(air)
; Pts = 10 $

This is the model that was fit as an MNL model in Chapter N17. We have now relaxed the equal variances assumption and replaced the four independent extreme value distributions with a multivariate (four variate) normal distribution. The probabilities are computed with 20 replications, which is fairly small; we do this for purposes of a simple illustration. Results are shown below. The MNL model is fit first to obtain the starting values for the iterations. The results for the MNP model are given next. The two sets of results are merged in the display below.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -189.52515
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    395.1 AIC/N =    1.881
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3321 .3202
Chi-squared[ 5]          =    188.46723
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
--------+--------------------------------------------------------------------
      GC|    -.01093**       .00459    -2.38  .0172     -.01992   -.00194
    TTME|    -.09546***      .01047    -9.11  .0000     -.11599   -.07493
   A_AIR|    5.87481***      .80209     7.32  .0000     4.30275   7.44688
AIR_HIN1|    -.00537         .01153     -.47  .6412     -.02797    .01722
 A_TRAIN|    5.54986***      .64042     8.67  .0000     4.29465   6.80507
TRA_HIN2|    -.05656***      .01397    -4.05  .0001     -.08395   -.02917
   A_BUS|    4.13028***      .67636     6.11  .0000     2.80464   5.45593
BUS_HIN3|    -.02858*        .01544    -1.85  .0642     -.05885    .00169
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

These are the estimates for the multinomial probit model:

```
-----------------------------------------------------------------------------
Multinomial Probit Model
Dependent variable                 MODE
Log likelihood function      -188.52929
Restricted log likelihood    -291.12182
Chi squared [  13 d.f.]       205.18505
Significance level               .00000
McFadden Pseudo R-squared      .3524041
Estimation based on N =    210, K =  13
Inf.Cr.AIC  =    403.1 AIC/N =    1.919
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3524 .3388
Constants only   -283.7588  .3356 .3216
At start values  -214.6841  .1218 .1033
Response data are given as ind. choices
Replications for simulated probs. =  10
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.02164**       .00857    -2.52  .0116     -.03843   -.00484
    TTME|    -.09385**       .03695    -2.54  .0111     -.16626   -.02144
   A_AIR|    5.00370**      2.01840     2.48  .0132     1.04771   8.95968
AIR_HIN1|     .00522         .02788      .19  .8516     -.04942    .05985
 A_TRAIN|    6.03988***     1.93044     3.13  .0018     2.25629   9.82347
TRA_HIN2|    -.06621***      .02340    -2.83  .0047     -.11207   -.02035
   A_BUS|    4.46541***     1.20839     3.70  .0002     2.09701   6.83382
BUS_HIN3|    -.01989         .01777    -1.12  .2629     -.05472    .01493
        |Std. Devs. of the Normal Distribution.
  s[AIR]|    2.58879**      1.20019     2.16  .0310      .23646   4.94112
s[TRAIN]|    2.14401**      1.05964     2.02  .0430      .06716   4.22086
  s[BUS]|        1.0    .....(Fixed Parameter).....
  s[CAR]|        1.0    .....(Fixed Parameter).....
        |Correlations in the Normal Distribution
rAIR,TRA|     .11088        1.04655      .11  .9156    -1.94032   2.16208
rAIR,BUS|    -.10316        1.21174     -.09  .9322    -2.47813   2.27181
rTRA,BUS|     .66132         .46589     1.42  .1558     -.25180   1.57445
rAIR,CAR|        0.0    .....(Fixed Parameter).....
rTRA,CAR|        0.0    .....(Fixed Parameter).....
rBUS,CAR|        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 table below compares the elasticities from the MNP model to the MNL model. The MNL results appear first. They are clearly similar, but the specification does make a difference.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.8019    .3198    .3198    .3198

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR| -1.0001    .3754    .4357    .4619
```

N27.4 Modifying the Covariance Structure

In the base case, the covariance and correlation matrix of the utility functions in the model is assumed to be of the following form, where we use a four choice model to illustrate:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ \rho_{12} & \sigma_2 & & \\ \rho_{13} & \rho_{23} & 1 & \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(Correlations instead of covariances are shown below the diagonal – this is schematic, not a covariance matrix as such.) The last row and the second to last variance must be restricted as shown (or equivalent restrictions must appear elsewhere in the matrix). (See the results in the preceding section for an illustration of these constraints.) However, at least in principle, there remain three free correlations in the matrix, those enclosed in parentheses. You can modify the structure of this matrix to change the standard deviations and to allow other correlations to be nonzero.

If you are not going to use the program default specification of the covariance matrix, then you must be cognizant of the identification problem in this model. The issue of identification concerns a limit on which and how many parameters can be estimated with the model, no matter how much data are in hand or how good those data are. In general, this model identifies a total of J-2 free standard deviations and (J-1)(J-2)/2 free correlations. You can restrict these two components of the model, so long as the counting rule is satisfied in the main. The usual way to do so will be to specify the standard deviations and the correlations separately, while maintaining identification. The standard deviations are straightforward, but you will have to be careful with the correlations. It is easy to specify an unidentified model, and *NLOGIT* cannot prevent you from doing so. You will know that the model you have specified has too many free parameters specified if the solver reaches maximum iterations without finding a solution, or it claims to reach a solution but the estimated standard errors are huge.

N27.4.1 Specifying the Standard Deviations

The standard deviations in the model are restricted in that two of them (the last two as *NLOGIT* formulates the model) must be set equal to 1.0. You may specify the vector of standard deviations with

; Sdv = list

You must provide exactly *J* specifications (*J* is the number of alternatives). Note that the last two specifications that you give will be redundant, since the $\sigma(J\text{-}1) = \sigma(J) = 1$ regardless. Nonetheless, you must provide the full set of *J* values (this is an internal consistency check). Names are used to specify free parameters or to impose equality constraints. Values are given to specify fixed parameters. All specified standard deviations must be strictly positive. For an example, to specify that only the first standard deviation in our four choice example is free, we might use

; Sdv = sigma1, 1, 1, 1

You may specify a homoscedastic model with

; Sdv = a single value or name

for a single specification. But, two of the standard deviations, $\sigma(J\text{-}1)$ and $\sigma(J)$, are already fixed at 1.000. So, if all standard deviations are to be equal, then all must equal 1.000. As such, in a homoscedastic model, all standard deviations must be fixed at 1.000. To specify this variant of the model, you may use any value, but this will then be the same as

; Sdv = 1

One useful way to specify these parameters will be to use named scalars. You might want to experiment with different values for some correlation or variance parameter. But, if your list **; Sdv = list** contains the name of a scalar that you created with **CALC**, then this is a fixed value, not a free parameter. Thus,

```
CALC        ; sd = 1.23 $
MNPROBIT    ; ... ; Sdv = sd,sd,1.0 $ (There are three choices.)
```

imposes the restriction that all three standard deviations are fixed (not to be estimated). The first two will be fixed at 1.23. But, if *sd* is not the name of an existing scalar, then the preceding will specify a model in which there is one free standard deviation parameter, which applies to both the first and second alternatives.

To illustrate this feature, we have fit the MNP model estimated earlier while imposing homoscedasticity. The command is

```
MNPROBIT    ; Lhs  = mode
            ; Choices = air,train,bus,car
            ; Rhs  = gc,ttme
            ; Rh2 = one,hinc
            ; Effects: gc(air)
            ; Pts = 10
            ; Sdv = 1,1,1,1 $
```

Results for this model are shown below. The imposition of the restriction actually has a minimal effect on the results, as can be seen in the results below, compared with those given earlier. Nonetheless, the log likelihood falls from -189.52929 to -191.67856. The chi squared for this test of homoscedasticity is only 4.299, which does not exceeds 5.99. The hypothesis of homoscedasticity and independence would not be rejected, in contrast to Chapter N26 by comparing the MNL and HEV models. The corresponding chi squared there was 16.754 with three degrees of freedom – the critical value is 7.815.)

```
-----------------------------------------------------------------------------
Multinomial Probit Model
Dependent variable                 MODE
Log likelihood function      -191.67856
Restricted log likelihood    -291.12182
Chi squared [  11 d.f.]       198.88651
Significance level               .00000
McFadden Pseudo R-squared      .3415864
Estimation based on N =    210, K =  11
Inf.Cr.AIC  =    405.4 AIC/N =    1.930
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3416 .3299
Constants only   -283.7588  .3245 .3125
At start values  -214.6841  .1072 .0913
Response data are given as ind. choices
Replications for simulated probs. =  10
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.01178***      .00319    -3.69  .0002     -.01803   -.00553
    TTME|    -.05537***      .01085    -5.10  .0000     -.07663   -.03411
   A_AIR|    3.16417***      .72595     4.36  .0000     1.74134   4.58701
AIR_HIN1|     .00107         .01392      .08  .9387     -.02622    .02836
 A_TRAIN|    3.68996***      .55807     6.61  .0000     2.59617   4.78376
TRA_HIN2|    -.04330***      .00987    -4.39  .0000     -.06265   -.02395
   A_BUS|    2.79244***      .45752     6.10  .0000     1.89572   3.68916
BUS_HIN3|    -.02220*        .01146    -1.94  .0528     -.04466    .00026
        |Std. Devs. of the Normal Distribution.
  s[AIR]|        1.0    .....(Fixed Parameter).....
s[TRAIN]|        1.0    .....(Fixed Parameter).....
  s[BUS]|        1.0    .....(Fixed Parameter).....
  s[CAR]|        1.0    .....(Fixed Parameter).....
        |Correlations in the Normal Distribution
rAIR,TRA|    -.93899        1.72238     -.55  .5856    -4.31480   2.43682
rAIR,BUS|    -.17167         .80366     -.21  .8308    -1.74681   1.40346
rTRA,BUS|     .55039*        .28791     1.91  .0559     -.01390   1.11467
rAIR,CAR|        0.0    .....(Fixed Parameter).....
rTRA,CAR|        0.0    .....(Fixed Parameter).....
rBUS,CAR|        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.
-----------------------------------------------------------------------------
```

Elasticities for the homoscedastic model are shown in the top panel of the table below.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+--------------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+--------------------------------------
     AIR| -1.0448    .2400     .7513     .5672
Elasticity wrt change of X in row choice on Prob[column choice]
--------+--------------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+--------------------------------------
     AIR| -1.0001    .3754     .4357     .4619
```

N27.4.2 Specifying the Correlation Matrix

Unless your model is fairly small (generally not more than five choices) a completely unrestricted correlation matrix is usually going to cause convergence problems. (Keep in mind, you are estimating a correlation matrix for a set of variables that is unobserved.) You can specify the correlation matrix in two ways. You may impose both fixed value and equality constraints with

; Cor = list of specifications

where the list of specifications defines either a free parameter or the name of a previous parameter, or a fixed value. The setup has the same form as that for **; Sdv = list** described above. The list is for the lower triangle of the correlation matrix, *not including the elements on the diagonal*. For example, suppose the alternatives are *air,train,bus,car*. The correlation part of the disturbance covariance matrix (below the diagonal) is

$\rho(train,air)$

$\rho(bus,air)$ $\rho(bus,train)$

$\rho(car,air)$ $\rho(car,train)$ $\rho(car,bus)$.

Then,

; Cor = Rta, Rba, 0.5, Rc, Rc, Rc

imposes one fixed value constraint and two equality constraints. There are three free parameters. Note in the general specification for a four choice model, identification allows only three free correlations, so the preceding merely rearranges the free correlations. This will change the parameter values, but it will not change the log likelihood.

In this specification, you must specify the full list of $J(J\text{-}1)/2$ symbols, where J is the number of alternatives (including repetitions if you are imposing equality constraints). Symbols may be any alphanumeric character string you desire. Numeric values which fix correlations must be strictly between -1 and +1. Note once again the warning noted earlier. *The name of an existing scalar provides a fixed value.*

> **NOTE:** Although you are providing $J(J\text{-}1)/2$ symbols for the correlation matrix, in fact, the model allows only $(J\text{-}1)(J\text{-}2)/2$ free parameters in the correlation matrix. You will normally satisfy the identification restriction by placing zeros in the matrix, but this is not strictly necessary. Having two correlations free but equal to each other is the same (for identification purposes) as having one free correlation and one set equal to zero. Note the application of this result in the example above – the equality of the last three correlations imposes two restrictions.

You can fix certain pairwise equalities of the correlations with the following shortcut:

; Eqc = choice, choice, …, choice.

This forces all pairwise correlations for the group of outcomes to be equal. For example,

; Eqc = air,train,car

imposes the restriction ρ(**train**,**air**) = ρ(**train**,**car**) = ρ(**air**,**car**). You may further impose this equality to a fixed value by adding the value in parentheses after the list. For example,

; Eqc = air,train,car (.75).

Finally, you may force *all* pairwise correlations in the model to be equal by giving a single specification. Use

; Cor = value

to fix all correlations at the value. For example, **; Cor = 0** would be typical – this would fix all correlations at zero. (This would produce a version of the HEV model, with normally distributed disturbances rather than extreme value.) Or, you may specify that there be a single correlation coefficient to be estimated, with

; Cor = name.

For our four choice example, you might specify **; Cor = r** which would force all six correlations to be equal, and there would be one parameter to be estimated. Note that the default option here is a free, unrestricted correlation matrix. (Note, **; Cor = rho** would *fix* all correlations at the current value of the scalar *rho*.)

To illustrate this feature, we now fit a true counterpart to the MNL model. The command would be

```
MNPROBIT   ; Lhs  = mode
           ; Choices = air,train,bus,car
           ; Rhs  = gc,ttme
           ; Rh2 = one,hinc
           ; Effects: gc(air)
           ; Pts = 10
           ; Sdv = 1,1,1,1
           ; Cor  = 0  $
```

The results are shown below. The log likelihood function now falls to -197.46059. The value in the unrestricted model was -188.52929. Thus, the chi squared statistic for testing this most restrictive model against the unrestricted model is twice the difference, or 17.863. The critical value is 11.07, so the five restrictions are rejected, albeit, not decisively. Note, also, that the restriction of no cross correlation, once homoscedasticity is assumed, produces a change in the log likelihood from -191.67856 to -197.46059, which is also significant.

```
-----------------------------------------------------------------------------
Multinomial Probit Model
Dependent variable                 MODE
Log likelihood function      -197.46059
Restricted log likelihood    -291.12182
Chi squared [   8 d.f.]       187.32244
Significance level               .00000
McFadden Pseudo R-squared      .3217252
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    410.9 AIC/N =    1.957
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3217 .3130
Constants only   -283.7588  .3041 .2952
At start values  -216.9267  .0897 .0780
Response data are given as ind. choices
Replications for simulated probs. =  20
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.00826***      .00298    -2.77  .0055     -.01409   -.00242
    TTME|    -.05773***      .00456   -12.66  .0000     -.06667   -.04879
   A_AIR|    3.70565***      .52264     7.09  .0000     2.68129   4.73000
AIR_HIN1|    -.00444         .00946     -.47  .6386     -.02298    .01410
 A_TRAIN|    3.73707***      .43113     8.67  .0000     2.89206   4.58207
TRA_HIN2|    -.04227***      .00860    -4.91  .0000     -.05914   -.02541
   A_BUS|    2.58935***      .47092     5.50  .0000     1.66636   3.51233
BUS_HIN3|    -.02058*        .01135    -1.81  .0699     -.04283    .00167
        |Std. Devs. of the Normal Distribution.
  s[AIR]|        1.0    .....(Fixed Parameter).....
s[TRAIN]|        1.0    .....(Fixed Parameter).....
  s[BUS]|        1.0    .....(Fixed Parameter).....
  s[CAR]|        1.0    .....(Fixed Parameter).....
        |Correlations in the Normal Distribution
rAIR,TRA|        0.0    .....(Fixed Parameter).....
rAIR,BUS|        0.0    .....(Fixed Parameter).....
rTRA,BUS|        0.0    .....(Fixed Parameter).....
rAIR,CAR|        0.0    .....(Fixed Parameter).....
rTRA,CAR|        0.0    .....(Fixed Parameter).....
rBUS,CAR|        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 table below compares the elasticities from the most restrictive model in the top panel to those from the least restrictive one, in the bottom. Once again, the effect is substantive, but not radical.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR     TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.8984    .4086    .4462    .3444
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR     TRAIN      BUS      CAR
--------+-----------------------------------
     AIR| -1.0001    .3754    .4357    .4619
```

N27.5 Testing IIA with a Multinomial Probit Model

A multinomial probit model with all standard deviations equal to one and uncorrelated random terms specifies a model that is comparable to the multinomial logit model. This suggests that you could test the IIA property by using an LR or LM test of the assumption that all of the standard deviations in a model *with uncorrelated disturbances* are equal. The test would be carried out as follows:

```
CALC        ; Ran (seed for generator) $
MNPROBIT    ; ... specify the choices and utility functions
            ; Cor = 0 $
CALC        ; lu = logl $
CALC        ; Ran (same seed for generator) $
MNPROBIT    ; ... specify the choices and utility functions
            ; Sdv = 1
            ; Cor = 0 $
CALC        ; lr = logl
            ; List
            ; lrstat = 2 * (lu - lr) $
```

We applied this procedure in passing in the preceding section. The log likelihoods for the three models estimated were

Most restrictive:	$\sigma_j = 1$, $\rho_{jm} = 0$	Log likelihood = -197.46059
Restrictive:	$\sigma_j = 1$	Log likelihood = -191.67856
Unrestricted:		Log likelihood = -189.52515.

In principle, a test of the first assumption as the null hypothesis against the alternative of the second is sufficient to reject IIA. We found the chi squared to be 11.564 with two degrees of freedom. The critical value is 5.99, so the hypothesis is rejected. A test of the third model against the null of the first produced a chi squared of 15.871 with five degrees of freedom. The critical value is 11.07, so once again the hypothesis is rejected. Which test should be preferred is uncertain. Under the null hypothesis, the estimated parameters in the second model are more precisely estimated, so this may favor it. We are unaware of any other evidence on the question.

N27.6 A Model of Covariance Heterogeneity

You can add a form of individual heterogeneity to the disturbance covariance matrix. The model extension is

$$\text{Var}[\boldsymbol{\varepsilon}_i] = \exp[\boldsymbol{\gamma}'\mathbf{h}(t)] \times \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is the matrix defined earlier (the same for all individuals), and $\mathbf{h}(i)$ is an individual (not alternative) specific set of variables that *does not include a constant*. The new parameters to be estimated are $\gamma_1,...,\gamma_H$. Request this feature with

; Hfn = list of variables in h.

The parameters in $\boldsymbol{\gamma}$ can be restricted like those in $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$**,** using

; Rst = list of specifications for γ (only).

In the same fashion as **; Sdv** and **; Cor**, **; Rst = a single value or symbol** will constrain all parameters in $\boldsymbol{\gamma}$ to equal each other, and, if a value is given, to be fixed at that value.

N27.7 Panel Data – The Multinomial Multiperiod Probit Model

The multinomial probit model may be estimated with a panel of data. In this case, the utility function is modified as follows:

$$U_{ji,t} = \boldsymbol{\beta}'\mathbf{x}_{ji,t} + \varepsilon_{ji,t} + v_{ji,t},$$

where 't' indexes the periods or replications. There are two formulations for $v_{jt,p}$,

Random effects	$v_{ji,t} = v_{ji,s}$ (the same in all periods),
First order autoregressive	$v_{ji,t} = \alpha_j\, v_{ji,t-1} + a_{ji,t}$.

It is assumed that you have a total of T_i observations (choice situations) for person i. Two situations might lend themselves to this treatment. If the individual is faced with a set of choice situations that are similar and occur close together in time, then the random effects formulation is likely to be appropriate. However, if the choice situations are fairly far apart in time, or if habits or knowledge accumulation are likely to influence the latter choices, then the autoregressive model might be the better one.

The data set for individual 'i' consists of T_i sets of observations. Each 'set' is a choice situation. Consider, for example, a four choice model. If individual 't' has 10 choice situations in their data set, then for that person, your physical data set for this person contains 10 times four, or 40 rows of data. As suggested, the number of situations may vary by person though the number of choices in the choice set in each situation must be the same, and the same for all individuals. The number of choice situations is specified as usual for panel data with

; Pds = the specification.

Again, 'specification' gives either the fixed T or a variable which contains the fixed T_i for that person. Do note, however, that the count here is a count of groups, not a count of rows of data. To continue our example, with four choices, and 10 situations, you would have 40 lines of data for this person, but would use **; Pds = 10** not **; Pds = 40**. Likewise, if you were using a count variable, your count variable for this person would equal 10.0 on each of the 40 lines of data. This feature cannot be specified in the command builder; it must be part of the command.

The default specification is the random effects model. This is specified simply by specifying the number of periods. The AR(1) model is specified by adding **; AR1** to the model command. You can restrict the autoregression parameters by using

; AR1 = list of symbols

in the same fashion as the correlations and standard deviations discussed in the preceding section.

There are some important restrictions that constrain this model. First, this is for very small panels. The reason is that the full data set for the individual must be used in the integration. Thus, if you have a four choice model, and four periods, then it is necessary to evaluate 16 variate integrals to compute the log likelihood (actually 12-variate as the differences enter the computations). This will tightly restrict the size of model that this can apply to. The limit in the simulator is 20. Second, in this model, only J-1 random effects are identified, so the last row of the covariance matrix and the last autocorrelation coefficient are fixed at zero.

N27.8 Technical Details

The log likelihood function for this model is formulated as follows: Suppose alternative j is chosen. Let the matrix

$$\mathbf{S} = \begin{bmatrix} 1 & & \\ r_{21} & 1 & \\ r_{31} & r_{32} & 1 \end{bmatrix},$$

(with appropriate zeros inserted and larger for a model with more than three choices) be the $J\times J$ correlation matrix for the J disturbances. Then, by construction,

$$U_{ji} > U_{qi} \text{ for all } q \text{ not equal to } j.$$

The probability of this outcome occurring is

$$\begin{aligned} &\text{Prob} (\varepsilon_{1i} - \varepsilon_{ji} < \boldsymbol{\beta}'(\mathbf{x}_{1i} - \mathbf{x}_{ji}), \\ &\quad \ldots \\ &\quad \varepsilon_{qi} - \varepsilon_{ji} < \boldsymbol{\beta}'(\mathbf{x}_{qi} - \mathbf{x}_{ji}) \text{ for the } J\text{-1 alternatives that are not } j). \end{aligned}$$

This is a (J-1) variate integral for the normal CDF with covariance matrix $\mathbf{V} = \mathbf{TST}'$, where $\mathbf{T}$ has J-1 rows, [1 0 0 ... -1 0 0 / 0 1 0 ... -1 0 ... /...] and where in the qth row, the +1 appears in the qth position and the -1 appears in the jth position. Row j is all zeros, and is dropped. The J-1 fold integral for the normal CDF with zero mean vector, covariance matrix $\mathbf{V}$, lower limits $-\infty$ and upper limit $\boldsymbol{\beta}'(\mathbf{x}_{jt} - \mathbf{x}_{qt})$ is the probability that enters the log likelihood.

All derivatives are computed numerically, so added to the time consumption of the function evaluation is the need to compute the probability many times for each observation. As a general rule, this time will be long. Estimation of the MNP model is the most time consuming among those supported by *NLOGIT*.

N27.9 Multivariate Normal Probabilities

NLOGIT uses the GHK (Geweke, Hajivassiliou, Keane) simulation methodology to approximate the multivariate normal CDF. (See Greene (2012) for details.) The technique produces relatively fast and accurate approximations to the M fold integral

$$P = \int_{A(M)}^{B(M)} \ldots \int_{A(1)}^{B(1)} f(x_1,\ldots,x_M)dx_1,,,dx_{M.}$$

where $f(\ldots)$ is the M-variate normal density function for $\mathbf{x}$ with mean vector zero and $M\times M$ positive definite covariance matrix, $\mathbf{\Omega}$. The approximation is obtained by averaging a set of R replications obtained by transforming draws produced by a random number generator. The simulation estimator of P is consistent in R. Further details may be found in Greene (2012) and in the symposium in the November, 1994, *Review of Economics and Statistics* and the references cited there. Usage, including how to set R is discussed below. M may be up to 20, though the accuracy for a given R declines with M, though for any M, it increases with R. Again, the estimated P is consistent in R.

The value of R, the number of replications, is set globally, at the time you start *NLOGIT*, at 100. Authors differ on how large R must be to get good approximations. The default 100 is a compromise. Some have mentioned 500. You may change R, but be aware that higher R leads to greatly increased amounts of computation; estimators which use this technique are slow. The ways to set R are with **CALC** and in the estimation commands. To set R permanently, use

CALC ; Rep (r) $ (for example, **CALC ; Rep (100) $**).

To set the number of replications in the command, use

MNPROBIT ; ... ; Pts = the desired value of R $

The full method of computing the integrals is detailed in Greene (2012). We will provide only a sketch here. The desired probability is Prob[$a_i \le x_i \le b_i$, i = 1,...,K], where the K variables have zero means and covariance matrix $\mathbf{\Sigma}$. (Nonzero means are accommodated just by transformation to simple deviations.) The probability is approximated by

$$P = \frac{1}{R}\sum_{r=1}^{R}\prod_{k=1}^{K} Q_{rk} ,$$

where R is the number of points used in the simulation. The Cholesky factorization of $\mathbf{\Sigma}$ is $\mathbf{LL}'$ where $\mathbf{L} = [l]_{km}$ is lower triangular. Note $l_{km} = 0$ if $m > k$. The recursive computation of P is begun with $Q_{r1} = \Phi(b_1/l_{11}) - \Phi(a_1/l_{11})$, where $\Phi(t)$ is the standard normal CDF evaluated at t. Using the random number generator, ε_{r1} is a random draw from the standard normal distribution truncated in the range $A_{r1} = a_1/l_{11}$ to $B_{r1} = b_1/l_{11}$. The draw from this distribution is obtained using Geweke's method. For a draw from the $N[\mu,\sigma^2]$ distribution truncated in the range A to B, we obtain u = a draw from the U[0,1] distribution. Then, the desired draw is

$$z = \mu + \sigma\Phi^{-1}[(1-u)\Phi((B-\mu)/\sigma) + u\Phi((A-\mu)/\sigma)].$$

For $k = 2,...,K$, use the recursion

$$A_{rk} = \left[a_k - \sum_{m=1}^{k-1} l_{km}\varepsilon_{rm} \right] / l_{kk}\ , \quad B_{rk} = \left[b_k - \sum_{m=1}^{k-1} l_{km}\varepsilon_{rm} \right] / l_{kk}\ ,$$

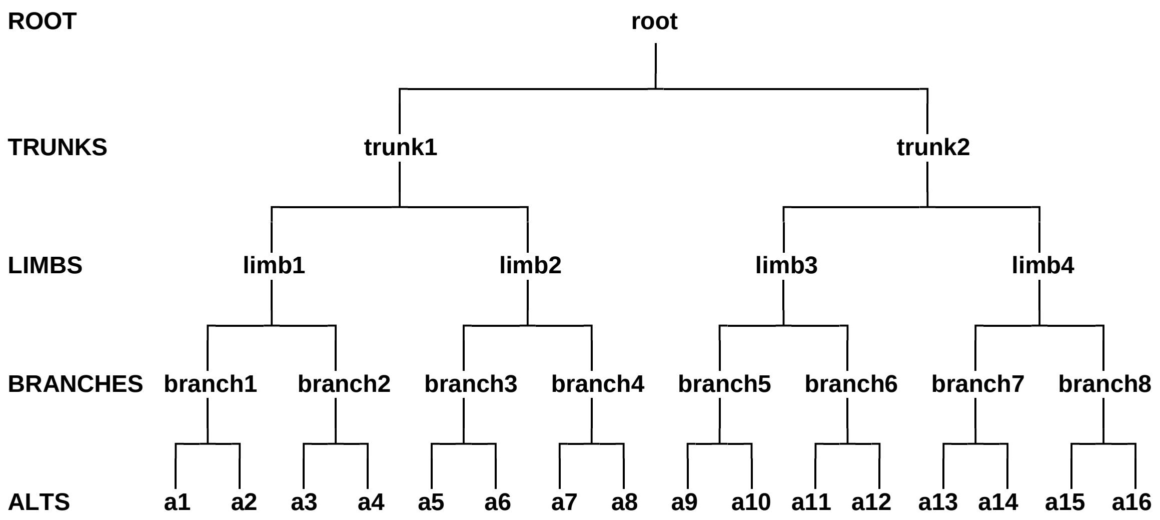
$$Q_{rk} = \Phi(B_{rk}) - \Phi(A_{rk}).$$

Then, P is the average of the R draws of products of K probabilities. Numerical properties and efficiency of this simulator are discussed at many places in the literature. References are given in Greene (2012).

N28: Nested Logit and Covariance Heterogeneity Models

N28.1 Introduction

The nested logit model is an extension of the multinomial model presented in Chapter N17. The models described here are based on variations of a four level tree structure such as the following:



Individuals are assumed to make a choice among NALT = J alternatives (alts) in a choice set. The 'twigs' in the tree are the elemental alternatives in the choice set. There may be up to 500 alternatives in the model, a total of 25 branches throughout the tree, 10 limbs, and five trunks. The model may contain one or more limbs. Each limb may contain one or more branches, and each branch may contain one or more twigs (choices). If there is only one trunk and one limb, the model is, by implication, a two level model. As for single level models, choice sets may vary by individual. However, in order to construct a tree for such a setting, a universal choice set, as described in Section N20.2.1, is necessary. The variable sized choice set is then indicated by setting up the full tree structure, and indicating that certain choices are unavailable for the particular individual.

The command for fitting nested logit models is the same as described in Chapters N19-N20 for one level models, save for the addition of the tree definition in the command and, optionally, the specification of additional utility functions for choices made at higher levels in the tree. The nested logit model is limited to four level models for full information maximum likelihood (FIML) estimation. It also allows estimation of two and higher level models by sequential, or two step estimation.

Utility functions can be specified for trunks the same as for limbs and branches (though it is unlikely that there will be very many attributes at this level in a tree). All options are available, including logs, Box-Cox transformation, fixed values, starting values, trunk specific constants, interaction terms, and so on. Utility functions for the trunks may include up to 10 variables including the set of constant terms if used. Since the command structure and options for the nested logit model are the same as those for the one level model, we will present in this chapter only the parts of the command setup that are specific to nested models. All users of this program should read Chapters N18-N22 before proceeding.

Most of the discussion to follow concerns full information maximum likelihood estimation of the nested logit model. The 'standard' (nonnormalized) model is discussed in Sections N28.2-N28.6. Two important variants on the model are discussed in Section N28.7. After setting up the model, users will generally want to use one of the alternative specifications discussed here. Section N28.9 presents a method of sequential, limited information maximum likelihood estimation. There are ever fewer settings in which this is a preferable estimator to FIML, but they do arise occasionally. The last three sections present two extensions of the nested logit model, one that accommodates observed individual heterogeneity and the second, that relaxes the assumption that each alternative is limited to appear in a single branch.

N28.2 Mathematical Specification of the Model

Individuals are assumed to choose one of the alternatives at the lowest level of the tree. Thus, they also choose a branch, a limb and a trunk. We denote by $j|b,l,r$ the choice of alternative j in branch b in limb l in trunk r. The number of alternatives in the branch/limb/trunk, $N_{b|l,r}$, can vary in every branch, limb, and trunk, and the number of branches in the l,rth limb/trunk, $N_{l|r}$ is likely to vary across limbs and trunks as well. No assumption of equal choice set sizes is made at any point in the following. (Note that for ease of presentation, we have dropped the observation subscript.)

The choice probability defined in Chapter N17 is now redefined to be the conditional probability of alternative j in branch b, limb l, and trunk r, $j|b,l,r$:

$$P(j|b,l,r) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r})}{\sum_{q|b,l,r}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l,r})} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r})}{\exp(J_{b|l,r})},$$

where $J_{b|l,r}$ is the *inclusive value* for branch b in limb l, trunk r, $J_{b|l,r} = \log \Sigma_{q|b,l,r} \exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l,r})$. At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l, trunk r,

$$P(b|l,r) = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \tau_{b|l,r}J_{b|l,r})}{\sum_{s|l,r}\exp(\boldsymbol{\alpha}'\mathbf{y}_{s|l,r} + \tau_{s|l,r}J_{s|l,r})} = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \tau_{b|l,r}J_{b|l,r})}{\exp(I_{l|r})},$$

where $I_{l|r}$ is the inclusive value for limb l in trunk r, $I_{l|r} = \log \Sigma_{s|l,r} \exp(\boldsymbol{\alpha}'\mathbf{y}_{s|l,r} + \tau_{s|l,r}J_{s|l,r})$. The probability of choosing limb l in trunk r is

$$P(l|r) = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_{l|r} + \sigma_{l|r}I_{l|r})}{\sum_{s|r}\exp(\boldsymbol{\delta}'\mathbf{z}_{s|r} + \sigma_{s|r}I_{s|r})} = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_{l|r} + \sigma_{l|r}I_{l|r})}{\exp(H_r)},$$

where H_r is the inclusive value for trunk r, $H_r = \log \Sigma_{s|r} \exp(\boldsymbol{\delta}'\mathbf{z}_{s|r} + \sigma_{s|r} I_{s|r})$. Finally, the probability of choosing a particular limb, r, is

$$P(r) = \frac{\exp(\boldsymbol{\theta}'\mathbf{h}_r + \phi_r H_r)}{\sum_s \exp(\boldsymbol{\theta}'\mathbf{h}_s + \phi_s H_s)}.$$

By the laws of probability, the unconditional probability of the observed choice made by an individual is

$$P(j,b,l,r) = P(j|b,l,r) \times P(b|l,r) \times P(l|r) \times P(r).$$

This is the contribution of an individual observation to the likelihood function for the sample.

The 'nested logit' aspect of the model arises when any of the $\tau_{j|i,l}$ or $\sigma_{i|l}$ or ϕ_l differ from 1.0. If all of these deep parameters are set equal to 1.0, the unconditional probability specializes to

$$\mathrm{P}(j,bj,l,r) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l,r} + \boldsymbol{\alpha}'\mathbf{y}_{b|l,r} + \boldsymbol{\delta}'\mathbf{z}_{l|r} + \boldsymbol{\theta}'\mathbf{h}_r)}{\sum_r \sum_l \sum_b \sum_j \exp(\boldsymbol{\beta}'\mathbf{x}_{jmbml,r} + \boldsymbol{\alpha}'\mathbf{y}_{b,l,r} + \boldsymbol{\delta}'\mathbf{z}_{l,r} + \boldsymbol{\theta}'\mathbf{h}_r)},$$

which is the probability for a one level model. The model is written in a very general form. The parameters of the model are, in exactly this order:

$$\beta_1,\beta_2,...,\beta_{nx},\alpha_1,\alpha_2,...,\alpha_{ny},\delta_1,\delta_2,...\delta_{nz},\theta_1,\theta_2,...,\theta_{nh},\tau_1...\tau_B,\sigma_1...,\sigma_L,\phi_1,...,\phi_R$$

where B is the total number of branches in the model, L is the number of limbs, and R is the number of trunks in the model. The **x**, **y**, **z**, and **h** vectors in the formulation above include all basic variables as well as all variables that interact with choice, branch, or limb specific dummy variables, etc. Once again, in this form, there may be different utility functions for each choice and, as described below, different utility functions defined for branches and limbs.

There is a vector of 'shallow' parameters, $[\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\delta},\boldsymbol{\theta}]$ at each level, which multiplies the attributes (at the lowest level), or, e.g., demographics, at a higher level. There are also three vectors of 'deep' parameters, which multiply the inclusive values at the middle and high levels. In principle, there is one free inclusive value parameter for each branch in the model ($J_{b|l,r}$), one for each limb ($\sigma_{l|r}$), and one for each trunk (ϕ_r). But, some may have to be restricted to equal 1.0 for identification purposes. There are some degenerate cases:

- If the model has one trunk, then the one ϕ equals 1.0.
- If the model has one limb in a trunk, the one σ also equals 1.0.
- If a limb contains a single branch, the τ for that branch equals 1.0.

The preceding describes a 'nonnormalized' model. The nested logit model also accommodates an explicit scaling factor at each level. The alternative normalizations that will reveal these scaling factors are shown in Section N28.7.

N28.3 Commands for FIML Estimation

This section will describe how to set up a nested logit model. The default estimation technique is full information maximum likelihood (FIML). That is, the entire model is estimated in a single pass. In Section N28.9, we will describe how to obtain two step, limited information maximum likelihood (LIML) estimators for a two level model. In general, LIML has no advantage when FIML is available, and is generally inferior. Moreover, as will emerge below, the LIML estimator is not able to impose many of the parametric restrictions inherent in the model.

N28.3.1 Data Setup

The arrangement of the data set for estimation of the nested logit model is exactly the same as shown in Chapter N19. There is no requirement that the choice sets be the same across individuals, but the nested logit model will require a definition of a universal choice set, so the command must contain the

; Choices = list of labels ...

specification. The nested model structure does mandate one special consideration if you are going to define utility functions for branches (*y*s), or limbs (*z*s). Since you have one line of data for each alternative, you will have more than one line of data for the variables in any branch or limb. In these cases, the values of *y* and *z* must be repeated for each alternative in the branch or limb.

The following model and setup illustrate this for a three level model: (all in trunk 1)

			*x*1	*x*2	*y*1	*y*2	*z*1	*z*2
limb 1	branch 1|1	twig 1|1,1	.6	1	3	.02	104	.9
		twig 2|1,1	.1	2	3	.02	104	.9
	branch 2|1	twig 1|2,1	.8	2	7	.15	104	.9
		twig 2|2,1	.2	3	7	.15	104	.9
limb 2	branch 1|2	twig 1|1,2	.9	6	11	.08	96	.4
		twig 2|1,2	.3	1	11	.08	96	.4
		twig 3|1,2	.4	0	11	.08	96	.4

N28.3.2 Tree Definition

The model command for estimating nested logit models is exactly as described in Chapter N19 for single level models, where the model name is now the generic **NLOGIT**;

NLOGIT ; Lhs = ... ; Choices = ... definition of choice set
; ... definition of utility functions for alternatives

All of the options described earlier are available. The nested logit model is requested by adding

; Tree = ... definition of the tree structure

to the command.

In order to specify the tree, use these conventions:

{ } specifies a trunk,
[ ] specifies a limb within a trunk,
( ) specifies a branch within a limb in a trunk.

Entries in a list are separated by commas. Names for trunks, limbs and branches are optional before the opening '{' or '[' or '('. If you elect not to provide names, the defaults chosen will be Trunk{*l*}, Lmb[*i*|*l*] and Br(*j*|*i*,*l*) respectively, where the numbering is developed reading from left to right in your tree definition. Alternative names appear inside the parentheses. Some examples are as follows:

One limb:

; Tree = travel [fly(air), ground(train,bus,car)]

One limb: (With one limb, the [ ] is optional.)

; Tree = fly(air), ground(train,bus,car)

One limb: (Branch names are optional. These would be Limb[1], Br(1|1) and Br(2|1).)

; Tree = (air), (train,bus,car)

One limb, one branch, no nesting: (This would be unnecessary and could be omitted.)

; Tree = (air,train,bus,car)

Nested logit model – two limbs, one with one branch:

```
; Tree = private [fly(air), ground(car_pas, car_drv)],
         public  [(train,bus)]
```

The fully nested 2×2×2×2 model shown in Section N28.1 could be specified with

```
; Choices = a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12,a13,a14,a15,a16
; Tree = Trunk1 {limb1 [branch1 (a1, a2),   branch2 (a3, a4)   ],
                 limb2 [branch3 (a5, a6),   branch4 (a7, a8)   ] },
         Trunk2 {limb3 [branch5 (a9, a10),  branch6 (a11, a12)],
                 limb4 [branch7 (a13, a14), branch8 (a15, a16)] }
```

N28.3.3 Utility Functions

You may define the utility functions exactly as described in Chapter N20 for one level models. You may also define utility functions for branches and limbs and trunks, but note that in order to do so, you must use the explicit form described in Section N20.4. These are specified exactly the same as those for elemental alternatives. For example, in a two level model, you might put demographic characteristics, such as income or family size, at the top level. A complete model might appear as follows:

```
NLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
           ; Tree = travel [public(bus,train), private(air,car)]
           ; Model:  U(air)      = ba +  bcost * gc + btime * ttme   /
                     U(train)    = bt +  bcost * gc + btime * ttme   /
                     U(car)      = bc +  bcost * gc + btime * ttme   /
                     U(bus)      =       bcost * gc + btime * ttme   /
                     U(public)   = ap +  apub * hinc /
                     U(private)  =       aprv  * hinc $
```

This model can be considerably collapsed;

```
           ; Model:  U(air,train,bus,car) = <ba,bc,0,bt> +
                                             bcost * gc + btime * ttme /
                     U(public,private)    = <ap,0> +
                                            <apub, aprv> * income $
```

Note that the same function specification U(...) is used for all three kinds of equations, for alternatives, branches, and limbs.

Finally, as noted earlier, you may impose equality constraints at any points in the model, just by using the same parameter name where you want the equality imposed. For example, if, for some reason, you desired to force the parameters *apub* and *bcost* to be equal, you could just change *apub* to *bcost* in the utility equation for *public*. That is, you can, if you wish, force equality of parameters at different levels of a model, once again, just by using the same parameter name in the model specification. (Given the impact of the scale parameters, this is probably inadvisable, but the program will allow you to do it nonetheless.)

The interaction of alternative specific constants, and branch and limb specific constants is complex, and it is difficult to draw generalities. As a general rule, models will usually become overdetermined, resulting in a singular Hessian, when there are more than NALT-1 constants, of all three types, in the entire model. Likewise, interactions of attributes and choice specific dummy variables can produce this effect as well. Users who encounter problems in which *NLOGIT* claims either that it is impossible to maximize the log likelihood function, or there is a singular Hessian, should examine the model for this pitfall.

N28.3.4 Setting and Constraining Inclusive Value Parameters

There is an inclusive value parameter for each limb, branch, and trunk in the model. For example, in the tree

; Choices = air,train,bus,car
; Tree = travel [public(bus,train), private(air,car)]

with the other parameters, we estimate $\tau_{public|travel}$, $\tau_{private|travel}$, σ_{travel}. Since there is only one limb, *travel*, σ_{travel} = 1.0. The other two parameters are free and unrestricted. You can modify the specification of these parameters in two ways:

- You may specify that they are equal to each other.
- You may specify that they are fixed values instead of free parameters to estimate.

To use these features, add the specification

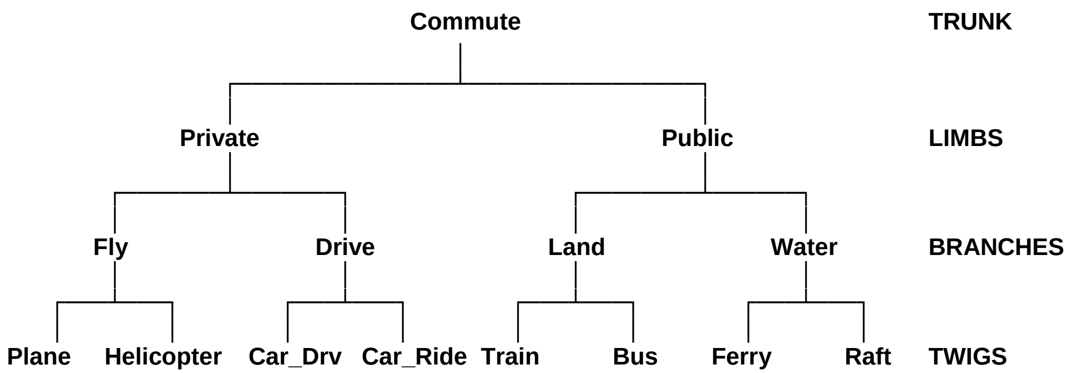
; Ivset: ... specification.

Note, once again, the presence of a colon in this specification. For purposes of this specification, τs, σs, and ϕs are treated the same. To force parameters to be equal, put the names of the branches and/or limbs together in parentheses in the **; Ivset:** specification.

For the example given above, to force the two τs to be equal in the estimated model, use

; Ivset: (public,private).

For a second example, consider this larger tree:



We would define this with

; Tree = private [fly(plane,helicptr), drive(car_ride,car_drv)],
public [land(train,bus), water(ferry,raft)].

There are six IV parameters, $\tau_{i|l}$ for each of *fly*, *drive*, *land*, and *water*, and σ_l for *private* and *public*. If it were desired to force $\sigma_{private} = \sigma_{public}$, $\tau_{fly|private} = \tau_{land|public}$, and $\tau_{water|public}$ (for some reason) to equal σ_{public}, you could use

; Ivset: (private,public,water) / (fly,land).

Note, once again, separate specifications are separated by slashes. Also, there is no problem using this device to force IV parameters at one level to equal those at another. Thus, '(**private,public,water**)' forces σ_{public} to equal $\tau_{water|public}$ and $\sigma_{private}$.

In addition to the preceding, you may fix inclusive value parameters. The setup is the same as above with the additional specification of the value in square brackets. I.e.,

; Ivset: (...) = [the value].

The list in parentheses may contain a single name, so as to fix a particular coefficient at a given value. You might have

; Ivset: (private,public) / (fly,ground) = [.75] / (land) = [.95] $

You will see a diagnostic message if you attempt to modify an inclusive value parameter that is fixed at 1.0 for identification purposes. For example, this specification of a two level model:

; Tree = travel [public(bus,train), private(air,car)]
; Ivset: (travel) = [.75]

generates an error message, since $\sigma_{\text{travel}} = 1.0$ (one limb). Note, also, that fixed IV parameters are off limits to equality constraints, as well. Thus, for this example, the specification

; Ivset: (travel,public)

also generates an error.

```
Error:  1093: You have given a spec for an IV parm that is fixed at 1.
```

You may not change the specification of ϕ_{travel}.

In the output of the estimation procedure, inclusive value parameters are denoted by the name of the branch or limb to which they are attached (or the default names given earlier).

N28.3.5 Starting Values

The preceding section shows how to specify that certain IV parameters are to be fixed at specified values. If you wish, instead, to provide starting values for the iterations, just remove the square brackets. Thus, for our earlier example:

; Ivset: (private,public) / (fly,land) = .75 / (water) = .95

makes $\sigma_{private} = \sigma_{public}$ in the model. The starting value for this one parameter is 1.0 (since none is provided). $\tau_{fly|private} = \tau_{land|public}$ in estimation, and the starting value is .75. $\tau_{water|public}$ starts at .95. Since $\tau_{drive|private}$ is not specified, it is a free parameter, and the starting value is 1.0.

NOTE: The default starting value for all IV parameters is 1.0.

The simple nonnested multinomial logit estimator is used to obtain the starting values. The model is fit as such by treating each level of the model as a simple, nonnested discrete choice model. Models are constructed as discrete choices among the choices at each level. Consider, for instance, the three level model in the example above. *NLOGIT* would compute three sets of estimates

β for the model of choice among the eight elemental choices,

α for the model of choice among the four branches,

δ for the model of choice between the two limbs.

The first of these is a consistent, albeit inefficient estimator of the elements of β. This is reported with the model results. However, the second and third are inconsistent because they omit the inclusive values from the parameters. The purpose is to provide a starting value that may be better than 0.0 (which is also inconsistent). The log likelihood function for the nested logit model is nonconvex, and in a complicated model, there may be some benefit to providing a good starting value. (These latter two sets of estimates are not reported. They are kept internally.)

You can use the output of this step to test the hypothesis of the nested logit model versus a nonnested model. An easy way to do that is to use a likelihood ratio test. The preliminary results are equivalent to a model in which all the IV parameters equal one. The later results will allow these parameters to be unrestricted. Twice the difference in the log likelihoods produces a chi squared test statistic with degrees of freedom equal to the number of free IV parameters. After each model is estimated, the scalar, *logl* will contain the log likelihood function that you will need to set up the test statistic. An example below shows these results. (Most of the model output is omitted.) The first box is produced by the initial estimator while the second is produced by the FIML estimator. Twice the difference in the two log likelihoods is about 18.4, which is larger than the critical value for two degrees of freedom of 5.99, so the hypothesis of the MNL is rejected.

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function       -199.97662  ←
Estimation based on N =    210, K =   5
Inf.Cr.AIC  =    410.0 AIC/N =    1.952
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                  MODE
Log likelihood function       -190.75302  ←
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
--------+--------------------------------------------------------------------
```

N28.3.6 Command Builder

The command builders can be used to specify the nested logit models. Select Model:Discrete Choice/Nested Logit to access the command builder. The choice variable is defined on the Main page and the rest of the model may be specified on the Options page. See Figures N28.1 and N28.2.

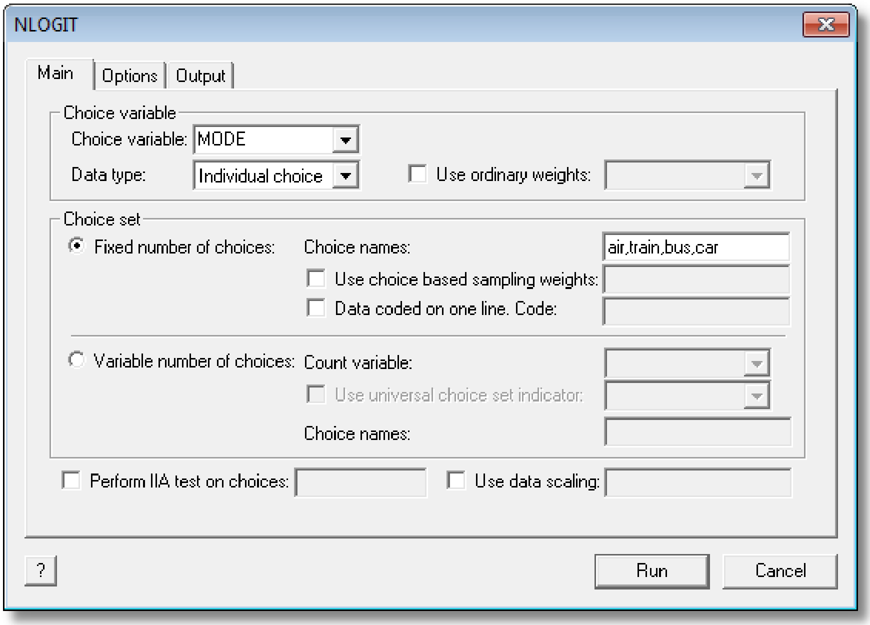


Figure N28.1 Main Page of Command Builder for Nested Logit Models

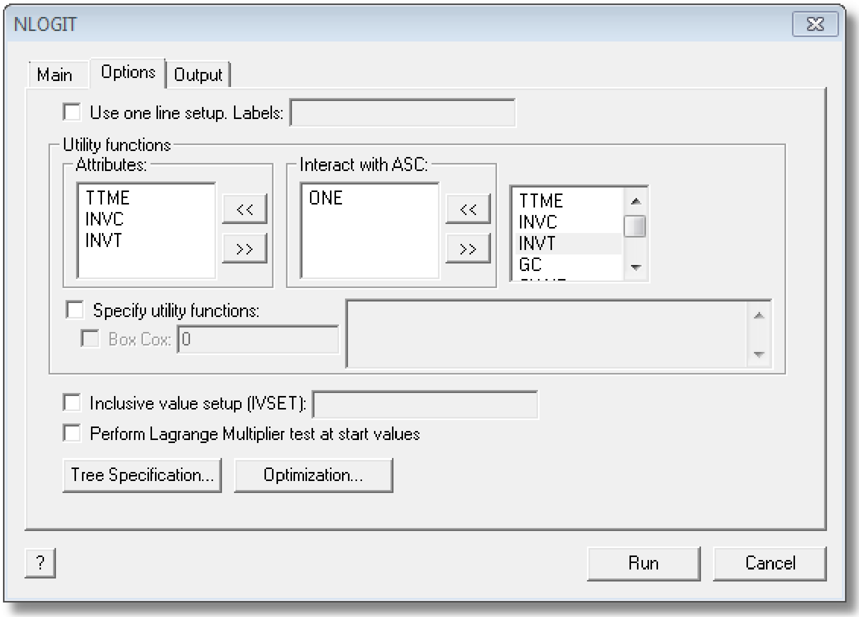


Figure N28.2 Options Page of Command Builder for Nested Logit Models

The tree is specified in a subsidiary dialog box by selecting Tree Specification at the bottom of the Options page. The dialog box, shown in Figure N28.3, allows you to define the tree graphically. Note in the dialog shown, *public* and *private* are siblings while *bus* is a child node of *public*.

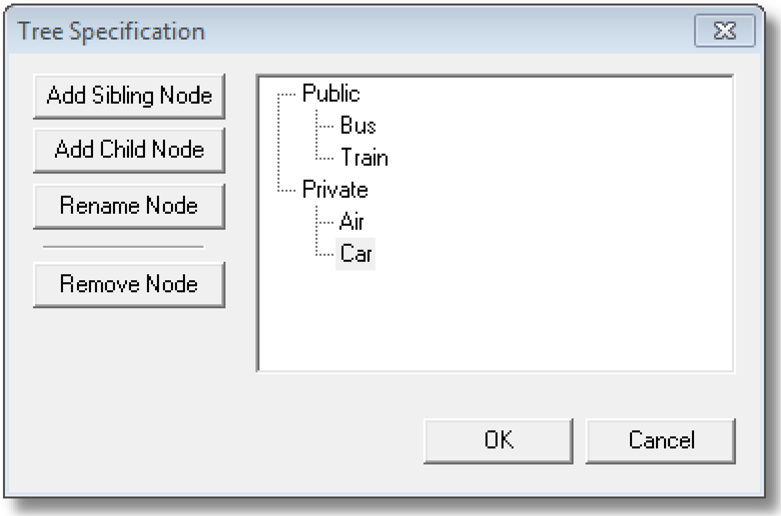


Figure N28.3 Tree Specification Dialog Box for Defining the Tree Structure

The remaining options for output and results to be saved are defined in the Output page as shown in Figure N28.4.

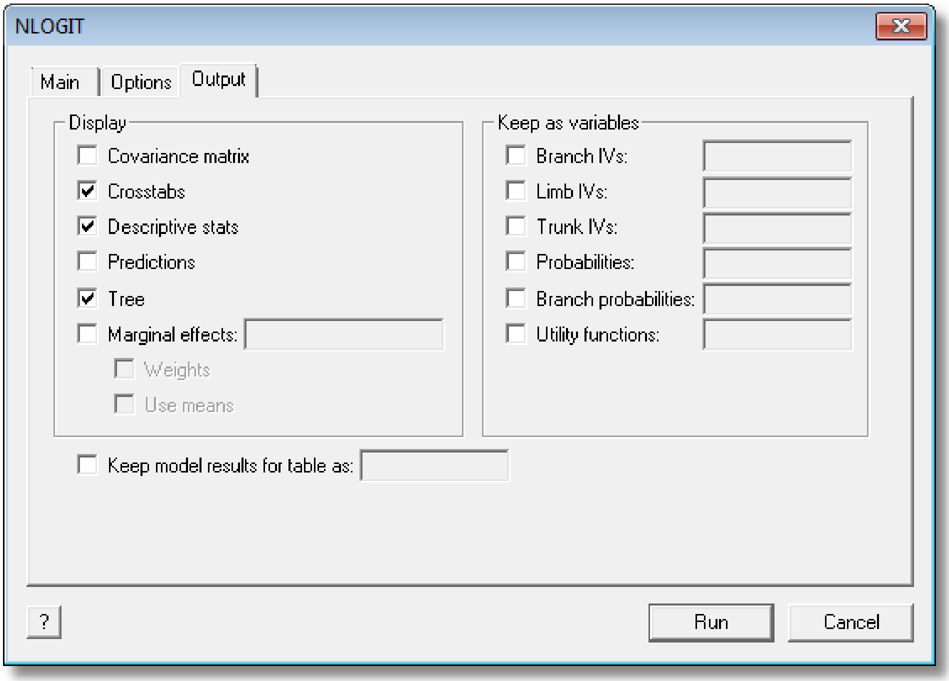


Figure N28.4 Output Page of Command Builder for Nested Logit Models

N28.4 Partial Effects and Elasticities

In the nested logit model with $P(j,b,l,r) = P(j|b,l,r) \times P(b|l,r) \times P(l|r) \times P(r)$, the marginal effect of a change in attribute k in the utility function for alternative J in branch B of limb L of trunk R on the probability of choice j in branch b of limb l of trunk r is computed using the following result: Lower case letters indicate the twig, branch, limb and trunk of the outcome upon which the effect is being exerted. Upper case letters indicate the twig, branch, limb and trunk which contain the outcome whose attribute is being changed:

$$\frac{\partial \log P(alt = j, limb = l, branch = b, trunk = r)}{\partial x(k) \mid alt = J, limb = L, branch = B, trunk = R)} = D(k \mid J, B, L, R) = \Delta(k) \times F \ ,$$

where $\Delta(k)$ = coefficient on $x(k)$ in $U(J|B,L,R)$

and

$$\begin{aligned} F = \ & \mathbf{1}(r=R) \times \mathbf{1}(l=L) \times \mathbf{1}(b=B) \times [\mathbf{1}(j=J) - P(J|BLR)] && \text{(trunk effect)},\\ & \mathbf{1}(r=R) \times \mathbf{1}(l=L) \times [\mathbf{1}(b=B) - P(B|LR)] \times P(J|BLR) \times \tau_{B|LR} && \text{(limb effect)},\\ & \mathbf{1}(r=R) \times [\mathbf{1}(l=L) - P(L|R)] \times P(B|LR) \times P(J|BLR) \times \tau_{B|LR} \times \sigma_{L|R} && \text{(branch effect)},\\ & [\mathbf{1}(r=R) - P(R)] \times P(L|R) \times P(B|LR) \times P(J|BLR) \times \tau_{B|LR} \times \sigma_{L|R} \times \phi_R && \text{(twig effect)}. \end{aligned}$$

(Note, in this expression, J, B, L and R are being used generically to indicate a particular choice, branch, limb and trunk, not the total numbers of twigs, branches, limbs and trunks.) The marginal effect is

$$\partial P(j,b,l,r)/\partial x(k)|J,B,L,R = P(j,b,l,r)\, \Delta(k)\, F.$$

A marginal effect has four components, an effect on the probability of the particular trunk, one on the probability for the limb, one for the branch, and one for the probability for the twig. (Note that with one trunk, $P(l) = P(1) = 1$, and likewise for limbs and branches.) For continuous variables, such as cost, you might be interested, instead, in the

$$Elasticity = x(k)|J,B,L,R \times \Delta(k|J,B,L,R) \times F.$$

NLOGIT will provide either. As in the case of nonnested models, marginal effects are requested with

; Effects: attribute [list of outcomes] / ...

or **; Effects: attribute (list) / ... for elasticities**

This generates a table of results for each of the outcomes listed. For example,

```
NLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Tree = travel [public(bus,train), private(air,car)]
          ; Model: U(air)   = ba + bcost * gc + btime * ttme /
                   U(train) = bc + bcost * gc + btime * ttme /
                   U(bus)   =      bcost * gc + btime * ttme /
                   U(car)   = bc + bcost * gc
          ; Effects: gc(car) ; Full $
```

This lists the effects on all four probabilities of changes in attribute generalized cost (*gc*) of choice *car*.

```
+--------------------------------------------------------------+
| Partial effects = average over observations                  |
|                                                              |
| dlnP[alt=j,br=b,lmb=l,tr=r]                                  |
| ---------------------------- = D(k:J,B,L,R) = delta(k)*F     |
| dx(k):alt=J,br=B,lmb=L,tr=R]                                 |
|                                                              |
| delta(k) = coefficient on x(k) in U(J|B,L,R)                 |
| F = (r=R)  (l=L) (b=B) [(j=J)-P(J|BLR)]                      |
|  +  (r=R)  (l=L) [(b=B) -P(B|LR)]P(J|BLR)t(B|LR)             |
|  +  (r=R) [(l=L)-P(L|R)] P(B|LR) P(J|BLR)t(B|LR)s(L|R)       |
|  + [(r=R) -P(R)] P(L|R)  P(B|IR) P(J|BIR)t(B|LR)s(L|R)f(R)   |
|                                                              |
| P(J|BLR)=Prob[choice=J |branch=B,limb=L,trunk=R]             |
| P(B|LR), P(L|R), P(R) defined likewise.                      |
| (n=N) = 1 if n=N, 0 else, for n=j,b,l,r and N=J,B,L,R.       |
| Elasticity = x(k) * D(j|B,L,R)                               |
| Marginal effect = P(JBLR)*D = P(J|BLR)P(B|LR)P(L|R)P(R)D     |
| F is decomposed into the 4 parts in the tables.              |
+--------------------------------------------------------------+
+-------------------------------------------------------------------------+
| Elasticity             averaged over observations.                      |
| Effects on probabilities of all choices in the model:                   |
| * indicates direct Elasticity effect of the attribute.                  |
+-------------------------------------------------------------------------+
+-------------------------------------------------------------------------+
| Attribute is GC       in choice CAR                                     |
|                        Decomposition of Effect if Nest    Total Effect|
|                        Trunk   Limb   Branch   Choice     Mean  St.Dev|
| Trunk=Trunk{1}                                                          |
| Limb=TRAVEL                                                             |
|    Branch=PUBLIC                                                        |
|       Choice=BUS         .000   .000    .857    .000        .857   .037 |
|       Choice=TRAIN       .000   .000    .857    .000        .857   .037 |
|    Branch=PRIVATE                                                       |
|       Choice=AIR         .000   .000  -1.015    .571       -.444   .051 |
| *     Choice=CAR         .000   .000  -1.015   -.338     -1.353   .073 |
+-------------------------------------------------------------------------+

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     BUS    TRAIN      AIR      CAR
--------+----------------------------------
     CAR|   .8570    .8570   -.4441  -1.3530
```

Note that across a row, the effects sum to the total effect given. The default method of computing the elasticities is to average the observation specific results. The results show the mean and the sample standard deviations. If you use the **; Means** specification, then the elasticities are computed once, and the results reflect the change, as shown below. (The differences are noticeably large.)

```
+------------------------------------------------------------------------+
| Elasticity             computed at sample means.                       |
| Effects on probabilities of all choices in the model:                  |
| * indicates direct Elasticity effect of the attribute.                 |
+------------------------------------------------------------------------+
+------------------------------------------------------------------------+
| Attribute is GC       in choice CAR                                    |
|                        Decomposition of Effect if Nest     Total Effect|
|                        Trunk   Limb   Branch   Choice      Mean  St.Dev|
| Trunk=Trunk{1}                                                         |
| Limb=TRAVEL                                                            |
|    Branch=PUBLIC                                                       |
|       Choice=BUS        .000   .000    .584    .000        .584   .000 |
|       Choice=TRAIN      .000   .000    .584    .000        .584   .000 |
|    Branch=PRIVATE                                                      |
|       Choice=AIR        .000   .000   -.411    .303       -.107   .000 |
| *     Choice=CAR        .000   .000   -.411   -.605     -1.016   .000 |
+------------------------------------------------------------------------+

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     BUS     TRAIN      AIR      CAR
--------+----------------------------------
     CAR|   .5843    .5843   -.1070  -1.0159
```

N28.5 Inclusive Values, Utilities, and Probabilities

You can request a listing of the actual outcomes and predicted probabilities with

; List

For large nested logit models, the listing would be extremely cumbersome, so a list can only be produced for models with seven or fewer elemental alternatives. You can also keep as variables the fitted probabilities and the branch, limb, and trunk inclusive values. The predicted probabilities are $P(j,b,l,r)$. The inclusive values for the branches are repeated for each choice (row of data) within the branches. The inclusive values for the limbs are, likewise, repeated for every alternative in the limb and similarly for trunks. An example appears in Section N21.3. The command specifications are:

; Prob = name to retain predicted probabilities as a variable
; Ivb = name to retain the branch level inclusive values as a variable
; Ivl = name to retain the limb level inclusive values as a variable
; Ivt = name to retain the trunk level inclusive values as a variable

Normally, in this setting, the unconditional probability, $P(j,b,l,r)$, is the one of interest. However, for some purpose, you might want, instead, the conditional probabilities at the twig level, $P(j,b,l,r)$. You can request to have this retained as a variable with

; Cprob = name to retain estimated conditional probabilities.

Lastly, the utility values at the twig level of the tree are

$$U(j|b,l,r) = \boldsymbol{\beta}'\mathbf{x}_{j|b,l,r}.$$

These are the values that you define in your **; Model: ...** specification. You may request to retain these for later use with

; Utility = name of the variable.

If you have not defined a utility function for an alternative, the value returned for that row of data is 0.0, not missing (-999). Utility values may be further processed like any other variable. You may find them useful, for example, for computing inclusive values in another model. An example of the use of these features is shown in the next section.

N28.6 Application of a Nested Logit Model

The following estimates a two level model. The tree has a 'degenerate' branch; the *air* branch has only a single alternative, *fly*. It also uses most of the optional features mentioned above.

```
NLOGIT    ; Lhs = mode
          ; Start = logit
          ; Choices = air,train,bus,car
          ; Tree = travel[fly(air), ground(train,bus,car)]
          ; Model: U(air,train,bus,car)  = bt*tasc+bb*basc+bg*gc+at*ttme /
                   U(fly,ground)         = aa*aasc+ah*hinca
          ; Describe
          ; Effects: gc(car) ; Pwt ; Full
          ; List
          ; Ivb = branchiv
          ; Ivl = limbiv
          ; Utility = u_choice
          ; Prob = pkji
          ; Cprob = pk_ji $
```

Starting values for the iterations are obtained by a one level multinomial logit model. The MNL also reports results of estimation of the branch choice model. These are the (inconsistent) estimates of α in the branch choice model. The MNL estimates are followed by the nested logit estimates.

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                 Choice
Log likelihood function        -378.59201
Estimation based on N =    210, K =   6
Inf.Cr.AIC  =    769.2 AIC/N =    3.663
Log-L for Choice   model =    -260.1975
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .0830 .0712
Log-L for Branch   model =    -118.3945
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
--------+--------------------------------------------------------------------
        |Model for Choice Among Alternatives
      BT|     .77779***      .20793     3.74  .0002      .37025   1.18532
      BB|    -.13076         .22872     -.57  .5675     -.57905    .31753
      BG|    -.01774***      .00405    -4.37  .0000     -.02569   -.00979
      AT|    -.01340***      .00318    -4.22  .0000     -.01963   -.00717
        |Model for Choice Among Branches
      AA|   -1.92254***      .35420    -5.43  .0000    -2.61677  -1.22832
      AH|     .02612***      .00817     3.20  .0014      .01010    .04214
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Normal exit:  27 iterations. Status=0, F=    193.6561
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                  MODE
Log likelihood function       -193.65615
Restricted log likelihood     -312.54998
Chi squared [   8 d.f.]        237.78765
Significance level                .00000
McFadden Pseudo R-squared       .3803994
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    403.3 AIC/N =    1.921
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -312.5500  .3804 .3724
Constants only   -283.7588  .3175 .3088
At start values  -287.6816  .3268 .3182
Response data are given as ind. choices
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Coefs. for branch level begin with AA
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      BT|    5.06460***      .66202     7.65  .0000     3.76706   6.36214
      BB|    4.09631***      .61516     6.66  .0000     2.89063   5.30200
      BG|    -.03159***      .00816    -3.87  .0001     -.04757   -.01560
      AT|    -.11262***      .01413    -7.97  .0000     -.14031   -.08492
        |Attributes of Branch Choice Equations (alpha)
      AA|    3.54087***     1.20813     2.93  .0034     1.17298   5.90875
      AH|     .01533         .00938     1.63  .1022     -.00306    .03372
        |IV parameters, tau(b|l,r),sigma(l|r),phi(r)
     FLY|     .58601***      .14062     4.17  .0000      .31040    .86162
  GROUND|     .38896***      .12367     3.15  .0017      .14658    .63134
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
+----------------------------------------------------------------------------+
|              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. |
| -------------------  --------    | -------------------+------------------- |
| BT             5.0646  TASC      |      .000      .000|      .000      .000 |
| BB             4.0963  BASC      |      .000      .000|      .000      .000 |
| BG             -.0316  GC        |   102.648    30.575|   113.552    33.198 |
| AT             -.1126  TTME      |    61.010    15.719|    46.534    24.389 |
+----------------------------------------------------------------------------+
+----------------------------------------------------------------------------+
|              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. |
| -------------------  --------    | -------------------+------------------- |
| BT             5.0646  TASC      |     1.000      .000|     1.000      .000 |
| BB             4.0963  BASC      |      .000      .000|      .000      .000 |
| BG             -.0316  GC        |   130.200    58.235|   106.619    49.601 |
| AT             -.1126  TTME      |    35.690    12.279|    28.524    19.354 |
+----------------------------------------------------------------------------+
PREDICTED PROBABILITIES (* marks actual, + marks prediction.)
Indiv    AIR       TRAIN     BUS       CAR
    1   .1515     .3518     .1232     .3734*+
    2   .2676     .1949     .0260     .5114*+
    3   .1563     .1040     .1509     .5888*+
    4   .3998     .1180     .0153     .4669*+
    5   .3418     .3510 +   .0469     .2603*
    6   .1323     .3423*+   .2212     .3043
    7   .4186*+   .0815     .1182     .3817
    8   .0955     .4956 +   .1848     .2241*
    9   .1685     .3915 +   .1371     .3030*
   10   .2484     .3203 +   .1122     .3191*
   11   .1965     .2143     .0269     .5623*+
   12   .2371     .1536     .0205     .5888*+
   13   .3324     .1552     .0201     .4922*+
   14   .2979     .2169     .0290     .4562*+
   15   .4731 +   .1921     .0583     .2765*
   16   .0814     .8298*+   .0340     .0548
   17   .0809     .8357*+   .0313     .0521
   18   .0573     .8456*+   .0446     .0524
   19   .1389     .3430*+   .2750     .2431
   20   .1771     .7935*+   .0022     .0273
   21   .0643     .8232*+   .0509     .0617
   22   .2078     .2684*    .0485     .4754 +
```

(Observations 11 - 210 are omitted.)

```
+--------------------------------------------------------------+
| Partial effects = prob. weighted avg.                         |
|                                                               |
| dlnP[alt=j,br=b,lmb=l,tr=r]                                   |
| ---------------------------- = D(k:J,B,L,R) = delta(k)*F      |
| dx(k):alt=J,br=B,lmb=L,tr=R]                                  |
|                                                               |
| delta(k) = coefficient on x(k) in U(J|B,L,R)                  |
| F = (r=R)  (l=L) (b=B) [(j=J)-P(J|BLR)]                       |
|  +  (r=R)  (l=L) [(b=B) -P(B|LR)]P(J|BLR)t(B|LR)              |
|  +  (r=R) [(l=L)-P(L|R)] P(B|LR) P(J|BLR)t(B|LR)s(L|R)        |
|  + [(r=R) -P(R)] P(L|R)  P(B|IR) P(J|BIR)t(B|LR)s(L|R)f(R)    |
|                                                               |
| P(J|BLR)=Prob[choice=J |branch=B,limb=L,trunk=R]              |
| P(B|LR), P(L|R), P(R) defined likewise.                       |
| (n=N) = 1 if n=N, 0 else, for n=j,b,l,r and N=J,B,L,R.        |
| Elasticity = x(k) * D(j|B,L,R)                                |
| Marginal effect = P(JBLR)*D = P(J|BLR)P(B|LR)P(L|R)P(R)D      |
| F is decomposed into the 4 parts in the tables.               |
+--------------------------------------------------------------+
+------------------------------------------------------------------------+
| Elasticity              averaged over observations.                     |
| Effects on probabilities of all choices in the model:                   |
| * indicates direct Elasticity effect of the attribute.                  |
+------------------------------------------------------------------------+
+------------------------------------------------------------------------+
| Attribute is GC        in choice CAR                                    |
|                         Decomposition of Effect if Nest    Total Effect|
|                         Trunk   Limb   Branch   Choice     Mean  St.Dev|
| Trunk=Trunk{1}                                                          |
| Limb=TRAVEL                                                             |
|    Branch=FLY                                                           |
|       Choice=AIR         .000   .000    .336    .000        .336   .022 |
|    Branch=GROUND                                                        |
|       Choice=TRAIN       .000   .000   -.063    .646        .583   .049 |
|       Choice=BUS         .000   .000   -.074    .849        .775   .049 |
| *     Choice=CAR         .000   .000   -.226  -1.128      -1.353   .066 |
+------------------------------------------------------------------------+

Elasticity wrt change of X in row choice on Prob[column choice]
--------+----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+----------------------------------
     CAR|   .3359    .5829    .7752  -1.3532
```

N28.7 Alternative Normalizations

The formulation of the nested logit model in Section N28.2 imposes no restrictions on the inclusive value parameters. However, the assumption of utility maximization and the stochastic underpinnings of the model do imply certain restrictions. For the former, in principle, the inclusive value parameters must be between zero and one. For the latter, the restrictions are implied by the way that the random terms in the utility functions are constructed. In particular, the nesting aspect of the model is obtained by writing

$$\varepsilon_{j|b,l,r} = u_{j|b,l,r} + v_{b|l,r}.$$

That is, within a branch, the random terms are viewed as the sum of a unique component and a common component. This has certain implications for the structure of the scale parameters in the model. In particular, it is the source of the oft cited (and oft violated) constraint that the IV parameters must lie between zero and one. These are explored in Hunt (2000) and Hensher and Greene (2002). *NLOGIT* provides a method of imposing the restrictions implied by the underlying theory.

There are three possible normalizations of the inclusive value parameters which will produce the desired results. These are provided in this estimator for two and three level models only. This includes most of the received applications. We will detail these and how to estimate these here. Readers are referred to the aforementioned papers for discussion. For convenience, we label these random utility formulations RU1, RU2 and RU3.

RU1

The first form is

$$P(j|b,l) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\sum_{q|b,l}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|j,l})} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\exp(J_{b|l})},$$

where $J_{b|l}$ is the inclusive value for branch b in limb l,

$$J_{b|l} = \log \Sigma_{q|b,l} \exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l}).$$

At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l,

$$P(b|l) = \frac{\exp\left[\lambda_{b|l}(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l})\right]}{\sum_{s|l}\exp\left[\lambda_{s|l}(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l})\right]} = \frac{\exp\left[\lambda_{b|l}(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l})\right]}{\exp(I_l)},$$

where I_l is the inclusive value for limb l,

$$I_l = \log \Sigma_{s|l} \exp[\lambda_{s|l}(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l})].$$

The probability of choosing limb l is

$$P(l) = \frac{\exp[\gamma_l(\boldsymbol{\delta}'\mathbf{z}_{l|} + I_l)]}{\sum_s \exp[\gamma_s(\boldsymbol{\delta}'\mathbf{z}_s + I_s)]} = \frac{\exp[\gamma_l(\boldsymbol{\delta}'\mathbf{z}_l + I_l)]}{\exp(H)}.$$

Note that this the same as the familiar normalization used earlier; this form just makes the scaling explicit at each level. If there are no branch level utility functions, then the default model will produce results according to RU1.

RU2

The second form moves the scaling down to the twig level, rather than at the branch level. Here it is made explicit that within a branch, the scaling must be the same for alternatives.

$$P(j|b,l) = \frac{\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\sum_{q|b,l}\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{q|b,l})\right]} = \frac{\exp\left[\mu_{b|l}(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\exp(J_{b|l})}.$$

Note in the summation in the inclusive value that the scaling parameter is not varying with the summation index. It is the same for all twigs in the branch. Now, $J_{b|l}$ is the inclusive value for branch j in limb l,

$$J_{b|l} = \log \Sigma_{q|b,l} \exp[\mu_{b|l} (\boldsymbol{\beta}'\mathbf{x}_{q|b,l})].$$

At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l,

$$P(b|l) = \frac{\exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + (1/\mu_{b|l})J_{b|l}\right)\right]}{\sum_s \exp\left[\gamma_s\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + (1/\mu_{s|l})J_{s|l}\right)\right]} = \frac{\exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + (1/\mu_{b|l})J_{b|l}\right)\right]}{\exp(I_l)},$$

where I_l is the inclusive value for limb l,

$$I_l = \log \sum_{s|l} \exp\left[\gamma_l\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + (1/\mu_{s|l})J_{s|l}\right)\right].$$

Finally, the probability of choosing limb l is

$$P(l) = \frac{\exp\left[\boldsymbol{\delta}'\mathbf{z}_l + (1/\gamma_l)I_l\right]}{\sum_s \exp\left[\boldsymbol{\delta}'\mathbf{z}_s + (1/\gamma_s)I_s\right]} = \frac{\exp\left[\boldsymbol{\delta}'\mathbf{z}_l + (1/\gamma_l)I_l\right]}{\exp(H)},$$

where the log sum for the full model is

$$H = \log \sum_s \exp\left[\boldsymbol{\delta}'\mathbf{z}_s + (1/\gamma_s)I_s\right].$$

In the RU2 form, with two levels (ignore γ_l above), global utility maximization requires that $0 < 1/\mu_{b|l} < 1$. It is possible to impose this restriction on the estimated parameters. *NLOGIT* does not impose the restriction because finding that the estimates are outside this range is a helpful indicator that your specification might be inadequate. By imposing the restriction, the program would preempt this diagnostic information.

RU3

A third random utility form, suggested by Bates (1999), is actually identical to the second – it is merely a transformation of the parameters. It does, however, have some intrinsic convenience, and, in a different way, emphasizes the roles of the scaling at each level of the tree. The twig probability is

$$P(j|b,l) = \frac{\exp\left[\left(1/(\lambda_{b|l}\theta_l)\right)(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\sum_{q|b,l}\exp\left[\left(1/(\lambda_{b|l}\theta_l)\right)(\boldsymbol{\beta}'\mathbf{x}_{q|b,l})\right]} = \frac{\exp\left[\left(1/(\lambda_{b|l}\theta_l)\right)(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})\right]}{\exp(J_{b|l})}.$$

Now, $J_{b|l}$ is the inclusive value for branch b in limb l,

$$J_{b|l} = \log \Sigma_{q|b,l} \exp[(1/(\lambda_{b|l}\theta_l))(\boldsymbol{\beta}'\mathbf{x}_{q|b,l})].$$

At the next level up the tree, we define the conditional probability of choosing a particular branch in limb l,

$$P(b|l) = \frac{\exp\left[(1/\theta_l)\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l}\right)\right]}{\sum_{s|l}\exp\left[(1/\theta_l)\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l}\right)\right]} = \frac{\exp\left[(1/\theta_l)\left(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + J_{b|l}\right)\right]}{\exp(I_l)},$$

where I_l is the inclusive value for limb l,

$$I_l = \log\sum_{s|l}\exp\left[(1/\theta_l)\left(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + J_{s|l}\right)\right].$$

Finally, the probability of choosing limb l is

$$P(l) = \frac{\exp\left[\boldsymbol{\gamma}'\mathbf{z}_l + I_l\right]}{\sum_s\exp\left[\boldsymbol{\gamma}'\mathbf{z}_s + I_s\right]} = \frac{\exp\left[\boldsymbol{\gamma}'\mathbf{z}_l + I_l\right]}{\exp(H)},$$

where the log sum for the full model is

$$H = \log\sum_s\exp\left[\boldsymbol{\gamma}'\mathbf{z}_s + I_s\right].$$

A moment's inspection reveals that RU2 and RU3 are the same. Also, comparing RU3 and RU1, it can be seen that in RU3, the scaling is moved down from the highest (limb) level to the lowest (twig). However, RU1 is not the same as RU2 and RU3 in general. They are equivalent under the restriction that the IV parameters are equal, as can be seen in the examples below – the signature of the equivalence is the equality of the log likelihoods. Also, as the results below show, the RU3 form IV parameters are simply the reciprocals of their counterparts in RU2. To emphasize the point, the results for RU3 will include the RU2 equivalents.

N28.7.1 Nondegenerate Cases

The various normalizations are not equivalent unless the IV parameters are forced to equality, as can be seen in the estimates of the model below. We consider, first, the cases in which all branches have at least two alternatives – these are ‘nondegenerate cases.’ The first case is RU1 with no equality restriction on the two IV parameters.

```
NLOGIT      ; Lhs  = mode
            ; Choices = air,train,bus,car
            ; Model: U(air,train,bus,car) =
                     <aa,at,ab,0>+<bh,bh,bh,0>*hinc+bg*gc+<bt,bt,bt,0>*ttme
            ; Tree = private(air,car), public(train,bus)
            ; RU1 or RU2 or RU3 $
            ; Ivset: (private,public) is optional
```

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                    MODE
Log likelihood function         -189.25341
The model has 2 levels.
Random Utility Form 1:IVparms = LMDAb|l
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    5.35139***     .80836     6.62  .0000     3.76703   6.93575
      AT|    3.23177***     .56454     5.72  .0000     2.12530   4.33824
      AB|    2.40948***     .59755     4.03  .0001     1.23829   3.58067
      BH|    -.01496*       .00866    -1.73  .0842     -.03194    .00202
      BG|    -.01710***     .00394    -4.34  .0000     -.02482   -.00938
      BT|    -.08355***     .01168    -7.15  .0000     -.10644   -.06066
        |IV parameters, lambda(b|l),gamma(l)
 PRIVATE|    2.45644***     .49136     5.00  .0000     1.49340   3.41948
  PUBLIC|    1.45631***     .26533     5.49  .0000      .93627   1.97634
        |Underlying standard deviation = pi/(IVparm*sqr(6))
 PRIVATE|     .52212***     .10444     5.00  .0000      .31742    .72681
  PUBLIC|     .88069***     .16045     5.49  .0000      .56620   1.19517
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                     MODE
Log likelihood function          -191.57011
Restricted log likelihood        -291.12182
Chi squared [   8 d.f.]           199.10341
Significance level                   .00000
McFadden Pseudo R-squared          .3419589
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    399.1 AIC/N =    1.901
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3420 .3335
Constants only   -283.7588  .3249 .3162
At start values  -196.2454  .0238 .0113
Response data are given as ind. choices
Hessian is not PD. Using BHHH estimator
The model has 2 levels.
Random Utility Form 2:IVparms = Mb|l,Gl
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    MODE|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    7.73093***     1.30062     5.94  .0000     5.18176  10.28011
      AT|    6.55253***     1.20025     5.46  .0000     4.20008   8.90498
      AB|    5.69567***     1.06585     5.34  .0000     3.60664   7.78470
      BH|    -.03931**       .01537    -2.56  .0105     -.06943   -.00920
      BG|    -.02340***      .00631    -3.71  .0002     -.03577   -.01103
      BT|    -.10933***      .02020    -5.41  .0000     -.14891   -.06974
        |IV parameters, RU2 form = mu(b|l),gamma(l)
 PRIVATE|    2.08081***      .62713     3.32  .0009      .85166   3.30997
  PUBLIC|     .97434***      .29856     3.26  .0011      .38916   1.55952
        |Underlying standard deviation = pi/(IVparm*sqr(6))
 PRIVATE|     .61637***      .18577     3.32  .0009      .25228    .98046
  PUBLIC|    1.31633***      .40336     3.26  .0011      .52576   2.10689
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

When the IV parameters are restricted to be equal, the results for all three models are identical save for the normalizations of the IV parameters and the scaling of the utility parameters. Note that the log likelihoods are identical in these cases.

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                   MODE
Log likelihood function      -194.39015
The model has 2 levels.
Random Utility Form 1:IVparms = LMDAb|l
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    5.70390***      .83296     6.85  .0000     4.07133   7.33646
      AT|    4.13484***      .57986     7.13  .0000     2.99834   5.27134
      AB|    3.50510***      .57321     6.11  .0000     2.38163   4.62857
      BH|    -.02289***      .00835    -2.74  .0061     -.03925   -.00652
      BG|    -.01180***      .00409    -2.89  .0039     -.01981   -.00379
      BT|    -.08290***      .01147    -7.23  .0000     -.10538   -.06042
        |IV parameters, lambda(b|l),gamma(l)
 PRIVATE|    1.42231***      .25732     5.53  .0000      .91797   1.92665
  PUBLIC|    1.42231***      .25732     5.53  .0000      .91797   1.92665
        |Underlying standard deviation = pi/(IVparm*sqr(6))
 PRIVATE|     .90174***      .16314     5.53  .0000      .58199   1.22148
  PUBLIC|     .90174***      .16314     5.53  .0000      .58199   1.22148
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                   MODE
Log likelihood function      -194.39015
The model has 2 levels.
Random Utility Form 2:IVparms = Mb|l,Gl
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    8.11271***     1.27720     6.35  .0000     5.60944  10.61597
      AT|    5.88103***     1.06493     5.52  .0000     3.79380   7.96825
      AB|    4.98534***      .90735     5.49  .0000     3.20697   6.76371
      BH|    -.03255**       .01320    -2.47  .0137     -.05842   -.00668
      BG|    -.01678***      .00554    -3.03  .0024     -.02764   -.00593
      BT|    -.11791***      .01981    -5.95  .0000     -.15673   -.07909
        |IV parameters, RU2 form = mu(b|l),gamma(l)
 PRIVATE|    1.42231***      .35310     4.03  .0001      .73024   2.11438
  PUBLIC|    1.42231***      .35310     4.03  .0001      .73024   2.11438
        |Underlying standard deviation = pi/(IVparm*sqr(6))
 PRIVATE|     .90174***      .22387     4.03  .0001      .46297   1.34051
  PUBLIC|     .90174***      .22387     4.03  .0001      .46297   1.34051
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

N28.7.2 Degenerate Cases

The problematic case is the common one in which there are one or more degenerate branches (branches with only one alternative) in the model. To illustrate, we formulate the tree with

; Tree = fly(air), ground(train,bus,car) ; Ivset(fly,ground)

In this instance, Hunt (2000) argues that the model above is overparameterized. RU1 allows free parameters in both branches regardless, but, in fact, the scaling in the *fly* branch is not actually identified. The results below show the two cases, again, with and without the equality constraint imposed on the IV parameters. In the first case, a problem arises in RU2 and RU3, as *NLOGIT*, recognizing the identification issue, enforces the prior restriction that the IV parameter on a degenerate branch must be 1.0. When the restriction is released, the diagnostic does not recur, and the previous pattern emerges, with RU2 and RU3 equivalent apart from the scaling.

The RU2 form is not estimable in this fashion, as shown by the diagnostic. RU3 produces the same error message.

```
Error:  1093: You have given a spec for an IV parm that is fixed at 1.
Error:  1093: You have given a spec for an IV parm that is fixed at 1.
```

RU1 is estimable with degenerate branches:

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -192.86849
The model has 2 levels.
Random Utility Form 1:IVparms = LMDAb|l
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    7.39001***      .97196      7.60  .0000     5.48502   9.29501
      AT|    5.92704***      .79701      7.44  .0000     4.36493   7.48914
      AB|    5.05369***      .75511      6.69  .0000     3.57369   6.53368
      BH|    -.02876**       .01146     -2.51  .0121     -.05123   -.00630
      BG|    -.02466***      .00771     -3.20  .0014     -.03977   -.00955
      BT|    -.11463***      .01410     -8.13  .0000     -.14226   -.08700
        |IV parameters, lambda(b|l),gamma(l)
     FLY|     .57124***      .12946      4.41  .0000      .31750    .82497
  GROUND|     .57124***      .12946      4.41  .0000      .31750    .82497
        |Underlying standard deviation = pi/(IVparm*sqr(6))
     FLY|    2.24521***      .50883      4.41  .0000     1.24793   3.24249
  GROUND|    2.24521***      .50883      4.41  .0000     1.24793   3.24249
--------+--------------------------------------------------------------------
```

Both models are estimable when the IV parameters are unrestricted.

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -192.66566
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    7.58747***     1.02396     7.41  .0000     5.58055   9.59439
      AT|    5.86134***      .80223     7.31  .0000     4.28900   7.43368
      AB|    4.94585***      .76985     6.42  .0000     3.43696   6.45473
      BH|    -.02513**       .01238    -2.03  .0425     -.04940   -.00085
      BG|    -.02707***      .00836    -3.24  .0012     -.04345   -.01069
      BT|    -.11393***      .01409    -8.09  .0000     -.14154   -.08632
        |IV parameters, tau(b|l,r),sigma(l|r),phi(r)
     FLY|     .59492***      .13720     4.34  .0000      .32602    .86383
  GROUND|     .49562***      .15442     3.21  .0013      .19296    .79828
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -192.86849
The model has 2 levels.
Random Utility Form 2:IVparms = Mb|l,Gl
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      AA|    4.22146***      .95797     4.41  .0000     2.34386   6.09905
      AT|    3.38575***      .59926     5.65  .0000     2.21122   4.56027
      AB|    2.88686***      .55032     5.25  .0000     1.80825   3.96547
      BH|    -.01643**       .00751    -2.19  .0286     -.03114   -.00172
      BG|    -.01409***      .00364    -3.87  .0001     -.02122   -.00696
      BT|    -.06548***      .01045    -6.27  .0000     -.08596   -.04500
        |IV parameters, RU2 form = mu(b|l),gamma(l)
     FLY|       1.0    .....(Fixed Parameter).....
  GROUND|     .57124***      .11465     4.98  .0000      .34652    .79595
        |Underlying standard deviation = pi/(IVparm*sqr(6))
     FLY|    1.28255   .....(Fixed Parameter).....
  GROUND|    2.24521***      .45063     4.98  .0000     1.36199   3.12843
--------+--------------------------------------------------------------------
```

N28.8 Technical Details

This section will present the functions and gradients for a three level nested logit model. The probabilities for the four level model are shown in Section N28.2. The derivations for the four level model are essentially similar, but the amount of notation increases geometrically. The following will show the forms and patterns of the computations. In what follows, we denote the choice of alternative j in branch b of limb l by $j|b,l$. Branch b in limb l is denoted $b|l$. When it is necessary to sum terms, we denote summation over the alternatives in branch $b|l$ as $\Sigma_{q|b,l}$. That is, q will be the running index for summation over the terms in the branch. Likewise, we use $\Sigma_{s|l}$ to denote summation over the branches in limb l and Σ_l to denote summation over the limbs in the model. The probabilities in the nonnormalized nested logit model are as follows: The choice probability is the conditional probability of alternative j in branch b, limb l, and trunk r, $j|b,l,r$:

$$P(j|b,l) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\sum_{q|b,l}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b,l})} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b,l})}{\exp(J_{b|l})},$$

$$P(b|l) = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + \tau_{b|l}J_{b|l})}{\sum_{s|l,r}\exp(\boldsymbol{\alpha}'\mathbf{y}_{s|l} + \tau_{s|l}J_{s|l})} = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_{b|l} + \tau_{b|l}J_{b|l})}{\exp(I_l)},$$

$$P(l) = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_l + \sigma_l I_l)}{\sum_s \exp(\boldsymbol{\delta}'\mathbf{z}_s + \sigma_s I_s)} = \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_l + \sigma_l I_l)}{\exp(H)}.$$

The unconditional probability of the observed choice made by an individual is

$$P(j,b,l) = P(j|b,l) \times P(b|l) \times P(l) .$$

This section will list the first derivatives used in maximizing the log likelihood function and in obtaining the asymptotic covariance matrix for the estimates. The following definitions will be useful:

$$\begin{aligned}
\bar{\mathbf{x}}_{b|l} &= \sum\nolimits_{q|b,l} P(q\,|\,b,l)\mathbf{x}_{q|b,l},\\
\bar{\mathbf{x}}_{l} &= \sum\nolimits_{s|l} \tau_{s|l}\, P(s\,|\,l)\bar{\mathbf{x}}_{s|l},\\
\bar{\mathbf{x}} &= \sum\nolimits_{l} \sigma_l\, P(l)\bar{\mathbf{x}}_{l},\\
\bar{\mathbf{y}}_{l} &= \sum\nolimits_{s|l} P(s\,|\,l)\mathbf{y}_{s|l},\\
\bar{\mathbf{y}} &= \sum\nolimits_{l} \sigma_l\, P(l)\bar{\mathbf{y}}_{l},\\
\bar{\mathbf{z}} &= \sum\nolimits_{l} P(l)\mathbf{z}_{l}.
\end{aligned}$$

The contribution of an observation i to the log likelihood for the model is

$$\text{Log } L_i = \log P_i(j,b,l) = \log[P_i(j|b,l) \times P_i(b|l) \times P_i(l)]$$
$$= \log P_i(j|b,l) + \log P_i(b|l) + \log P_i(l),$$

where the subscript indicates evaluation at the data for individual i. Note that the full set of results for a one level model is obtained by examining the terms below that relate to $P_i(j|b,l)$ with $b = l = 1$, while a two level model is built up from $P_i(j|b,l)P_i(b|l)$. The parameters of the model are, in order, $[\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \tau_{...}, \sigma_{...}]$. Gradients and Hessians are obtained as the sums of the derivatives of the three parts. The definitions of deviations, $\Delta\mathbf{w}_{...}$ given with the gradients are used to produce a convenient format for the Hessians, which are built up recursively. The function, $\mathbf{1}[i=j]$, equals 1.0 if i equals j and equals 0.0 if not. For interpretation, note that in a term in a Hessian that relates, say, $b|l$ and $s|m$, $\mathbf{1}[l=m]$ means 'in the same limb,' while $\mathbf{1}[b=s]$ means 'in the same branch.' This is only possible if l equals m. For convenience in the derivations below, we will drop the observation subscript.

$$\partial \log P(j|b,l)/\partial\boldsymbol{\beta} = \mathbf{x}_{j|b,l} - \overline{\mathbf{x}}_{b|l} = \Delta\mathbf{x}_{j|b,l},$$

$$\partial \log P(j|b,l)/\partial \bullet = \mathbf{0} \text{ for } \boldsymbol{\alpha}, \tau_{sq}, \boldsymbol{\gamma}, \sigma_s,$$

$$\partial \log P(b|l)/\partial\boldsymbol{\beta} = \tau_{b|l}\overline{\mathbf{x}}_{b|l} - \overline{\mathbf{x}}_l = \Delta\overline{\mathbf{x}}_{b|l},$$

$$\partial \log P(b|l)/\partial\boldsymbol{\alpha} = \mathbf{y}_{b|l} - \overline{\mathbf{y}}_l = \Delta\mathbf{y}_{b|l},$$

$$\partial \log P(b|l)/\partial\tau_{s|q} = \mathbf{1}[l=q][\mathbf{1}(b=s) - P(s|q)]\, J_{s|q},$$

$$\partial \log P(b|l)/\partial \bullet = \mathbf{0} \text{ for } \boldsymbol{\gamma}, \sigma_s,$$

$$\partial \log P(l)/\partial\boldsymbol{\beta} = \sigma_l\overline{\mathbf{x}}_l - \overline{\mathbf{x}} = \Delta\overline{\mathbf{x}}_l,$$

$$\partial \log P(l)/\partial\boldsymbol{\alpha} = \sigma_l\overline{\mathbf{y}}_l - \overline{\mathbf{y}} = \Delta\overline{\mathbf{y}}_l,$$

$$\partial \log P(l)/\partial\boldsymbol{\gamma} = \mathbf{z}_l - \overline{\mathbf{z}} = \Delta\mathbf{z}_l,$$

$$\partial \log P(l)/\partial\tau_{s|q} = \sigma_l[\mathbf{1}(q=l) - P(q)]P(s|q)]\, J_{s|q},$$

$$\partial \log P(l)/\partial\sigma_s = [\mathbf{1}(l=s) - P(s)]I_s.$$

The analytic second derivatives are used to compute the asymptotic covariance matrix of the MLE. The log likelihood function is nonconvex because of the IV parameters, and, as such, Newton's method is a poor algorithm for optimization. We use BFGS, instead. The RU1 and RU2 forms of the model add additional nonlinearities. The preceding are the base case – these are modified to produce RU1 and RU2. RU3 is a simple reparameterization of RU2, so it is not developed separately.

N28.9 Sequential (Two Step) Estimation of Nested Logit Models

The preceding applies to full information maximum likelihood (FIML) estimation of nested logit models. In brief, the technique estimates all of the parameters simultaneously by maximizing the unconditional log likelihood,

$$\text{Log } L = \Sigma_i \log P_i(j,b,l,r) = \Sigma_i \log P_i(j|b,l,r) + \log P_i(b|l,r) + \log P_i(l|r) + \log P_i(r).$$

An alternative way to fit a special case of the model is by sequential, or two step estimation. We consider two level models, though as shown below, the technique can be extended to higher level models as well. An essential element for our purposes, however, is the restriction that at the upper level, the inclusive value parameters are constrained to be equal.

At the first step, we estimate the parameters of the conditional log likelihood,

$$\begin{aligned}\text{Log } L_c &= \Sigma_i \log P_i(j|b)\\ &= \Sigma_i \log[\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b}) / \Sigma_q \exp(\boldsymbol{\beta}'\mathbf{x}_{q|b})]\\ &= \Sigma_i \log[\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b}) / \exp(J_b)].\end{aligned}$$

(Since this is strictly for two level models, we have dropped the 'l,r' from the probabilities.) This simple discrete choice model provides estimates of $\boldsymbol{\beta}$ and, using $\boldsymbol{\beta}$ and the observed data, individual estimates of the inclusive values, J_b. The conditional model estimated at the second step is

$$P_i(b) = \exp(\boldsymbol{\alpha}'\mathbf{y}_b + \tau J_b) / \Sigma_s \exp(\boldsymbol{\alpha}'\mathbf{y}_s + \tau J_s).$$

Note that there is only a single τ parameter regardless of the number of branches. With a minor modification of the **NLOGIT** command to create interactions of the inclusive value with branch specific constants, this constraint could be relaxed. However, the subsequent computation of the appropriate asymptotic covariance matrix is considerably more complicated. (In principle, this restriction need not be imposed – see McFadden (1981). However, the extension to the case in which the restriction is relaxed is quite complex and difficult to justify given the availability of FIML.) With the individual estimates of the inclusive values in hand, this can also be interpreted as a simple discrete choice model,

$$P_i(b) = \exp(\boldsymbol{\alpha}^{*\prime}\mathbf{y}_b) / \Sigma_s \exp(\boldsymbol{\alpha}^{*\prime}\mathbf{y}_s),$$

in which the inclusive value is one of the attributes (the last). The lower level parameters are consistently, albeit inefficiently, estimated by just maximizing the conditional log likelihood function, and no special consideration need be made for the estimation of standard errors. At the second step, the estimates of $\boldsymbol{\alpha}^*$ are consistent, but the usual estimator of the standard errors (the inverse of the Hessian) needs to be adjusted to account for the fact that the parameters of the inclusive values are themselves estimates. The computations are detailed in the example below.

The computations for this estimator are automated in *NLOGIT*. To request this procedure, set up the full two level nested logit model as if you were using FIML. Then, change the normal command request as follows:

Step 1. For the first step of the estimation, add

; Ivb = name for inclusive value
; Conditional

to the **NLOGIT** command. *Do not include the inclusive value in the branch level utility functions.*

Step 2. For the second step of the estimation, use exactly the same **NLOGIT** command, except change the preceding to

; Sequential

The inclusive value that you created in Step 2 must now be added as the last attribute in the utility function(s) for the branch level.

The asymptotic covariance matrix is computed as follows. Let $\mathbf{H}_{11}$ equal the Hessian from the first step estimation. Let $\mathbf{H}_{22}$ be the Hessian from the second step estimation, including the estimate of τ. Let

$$\mathbf{H}_{21} = \Sigma_b [\mathbf{y}_b^* - \bar{\mathbf{y}}^*][J(\bar{\mathbf{x}}_b - \bar{\mathbf{x}})]' \text{ (and } \mathbf{H}_{12} = \mathbf{H}_{21}'),$$

where

$$\bar{\mathbf{x}}_b = \Sigma_{q|b} P(q\,|\,b)\mathbf{x}_{q\,|\,b},\quad \bar{\mathbf{x}} = \Sigma_b P(b)\bar{\mathbf{x}}_b,\quad \bar{\mathbf{y}}^* = \Sigma_b P(b)\mathbf{y}_b.$$

Then, the appropriate asymptotic covariance matrix for the two step estimator of α^* is

$$\mathbf{V} = [\mathbf{H}_{22} - \mathbf{H}_{21}[\mathbf{H}_{11} + \mathbf{H}_{12}\mathbf{H}_2^{-1}\mathbf{H}_{21}]^{-1}\mathbf{H}_{12}]^{-1}.$$

A simple example follows:

```
NLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Tree = fly(air), ground(train,bus,car)
          ; Model: U(air,train,bus,car) = <0,at,ab,0> + bc * gc /
                   U(fly,ground)        = ah * hinca
          ; Ivb = incvlu
          ; Conditional $
NLOGIT    ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Tree = fly(air), ground(train,bus,car)
          ; Model: U(air,train,bus,car) = <0,at,ab,0> + bc * gc /
                   U(fly,ground)        = ah * hinca + aiv * incvlu
          ; Sequential $
```

```
-----------------------------------------------------------------------------
Conditional logit model for choices only
Dependent variable              Choice
Log likelihood function      -101.63595
Estimation based on N =    210, K =   4
Inf.Cr.AIC  =    211.3 AIC/N =    1.006
Log-L for Choice   model =    -101.6360
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .6418 .6346
Log-L for Branch   model =        .0000
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
--------+--------------------------------------------------------------------
        |Model for Choice Among Alternatives
      AT|    2.38614***      .36950     6.46  .0000     1.66193   3.11035
      AB|     .76659**       .32387     2.37  .0179      .13182   1.40136
      BC|    -.07659***      .01004    -7.63  .0000     -.09627   -.05691
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Second step estimates of nested logit model
Dependent variable              Choice
Log likelihood function      -476.57959
Estimation based on N =    210, K =   2
Inf.Cr.AIC  =    957.2 AIC/N =    4.558
Log-L for Choice   model =    -340.3202
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588 -.1993-.2128
Log-L for Branch   model =    -136.2594
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
--------+--------------------------------------------------------------------
        |Model for Choice Among Alternatives
      AT|    2.38614***      .36950     6.46  .0000     1.66193   3.11035
      AB|     .76659**       .32387     2.37  .0179      .13182   1.40136
      BC|    -.07659***      .01004    -7.63  .0000     -.09627   -.05691
        |Model for Choice Among Branches
      AH|    -.01386***      .00428    -3.24  .0012     -.02225   -.00548
     AIV|     .04165         .05691      .73  .4642     -.06989    .15319
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

N28.10 Combining Data Sets and Scaling in Discrete Choice Models

An important property of the discrete choice model is the independence from irrelevant alternatives (IIA). This condition is induced by the assumed independence of the unobserved individual effects in the utility functions that define the model. Mathematically, an important part of the assumption is that the covariance matrix for $[\varepsilon_0,\varepsilon_1,...,\varepsilon_J]$ equals $\sigma^2\mathbf{I}$ - identical variances and zero covariances. The nested logit model is a device for partitioning the choice set to reduce or minimize the influence of the IIA/IID property. The model does not necessarily imply an interpretation of behavior or a particular behavioral hypothesis based on a hierarchical relationship among alternatives in a choice set.

In recent years, researchers and practitioners in transportation and marketing have examined logit models based on stated preference (SP) experiments in which individuals are given hypothetical combinations of attributes associated with each alternative in a choice set and asked to choose one. The experiment is repeated a number of times with varying attribute levels and stated responses. These methods are popular in cases in which one is 'stretching' the attribute levels beyond observed levels and in which one is evaluating the demand for a new alternative. (An early application of this approach is Beggs, Cardell, and Hausman's (1981) study of the demand for electric cars. We examined another large application in Chapter N22.) Although stated choice experiments are rich in information designed to elicit marginal rates of substitution between attributes, they are limited in their ability to represent the revealed preferences (RP) of individuals and hence to reproduce observed market shares. Revealed preference data are richer in information that can reproduce observed base market shares, but usually not so rich in the data needed to evaluate switching behavior associated with the introduction of new alternatives or changes in the levels of attributes. A combination of the two types of data can provide an attractive alternative estimation strategy.

When data from two different choice studies are derived (whether for the same individuals or for different samples), we cannot naively assume that the IIA/IID condition of equal variances holds across both data sets for the set of common alternatives. For example, we might have a revealed preference data set of four modes (drive alone, ride share, train, and bus); we might also have a stated choice experiment data set for the same four modes. If we were naively to pool the two data sets, ignoring the fact that they are not strictly independent when derived from the same sample of individuals, then we are implicitly assuming that the variances are the same across all eight alternatives – the four revealed preference models and the four stated choice experiment modes. If the variances are, indeed, the same, then the ratio of any two of them equals 1.0. This provides the basis for a test of equality. When they are not equal, setting the variance in one data set to 1.0 and estimating the variance in the other will provide the appropriate scaling parameter needed to validate pooling the two data sets.

Formally, for a common alternative in the two choice sets, let

$$\text{U}(choice_{rp}) = \alpha + \boldsymbol{\beta}'\mathbf{x}_{rp} + \boldsymbol{\gamma}'\mathbf{y} + \varepsilon_{rp},$$

$$\text{U}(choice_{sp}) = \delta + \boldsymbol{\beta}'\mathbf{x}_{sp} + \boldsymbol{\theta}'\mathbf{z} + \varepsilon_{sp},$$

$$\sigma^2 = \text{Var}[\varepsilon_{rp}]/\text{Var}[\varepsilon_{sp}]$$

= a scaling parameter such that $\text{Var}[U_{rp}] = \sigma^2\text{Var}[U_{sp}]$ so that pooling of the two data sets is valid,

where $\mathbf{x}_{rp}$, $\mathbf{x}_{sp}$ = attributes common to the RP and SP data sets,

$\mathbf{y}$, $\mathbf{z}$ = observed attributes specific to the RP or SP data sets,

$[\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\delta,\boldsymbol{\theta}]$ = the unknown parameters to be estimated,

ε_j, j = RP,SP = unobserved individual effects.

NLOGIT automates the scaling procedure for two applications – joint estimation for any tree structure (nested logit) model and sequential estimation for a single level (discrete choice) model. Although scaling sequentially a nested logit model with more than one level is feasible, *NLOGIT* currently limits the rescaling to a single optimal parameter, which may not be valid for a tree structure in which the variances can be different at each branch within the tree. We suggest that joint estimation be the preferred approach for trees up to four levels, and that sequential estimation be used for single level models and for each level in a tree structure with more than four levels. (*NLOGIT* provides FIML estimates for up to four levels.)

N28.10.1 Joint Estimation

The RP parameters to be estimated are $[\alpha,\boldsymbol{\beta},\gamma]$. The SP parameters are $[\sigma\delta,\sigma\boldsymbol{\beta},\sigma\gamma]$. The scaling has no other effect on the distributional assumptions or on the conversion of the indirect utility expressions to choice probabilities. The scaling of $\sigma\boldsymbol{\beta}$ is the essential link between the two data models. The SP model, however, is nonlinear. This estimation problem can be solved with *NLOGIT* by setting up an artificial tree structure as follows: The artificial nest is constructed to have at least twice as many alternatives as are actually observed. One subset is labeled the RP alternatives and the other is labeled the SP alternatives. The indirect utility functions in each case are defined by the U_{rp} and U_{sp} expressions shown earlier, without σ. The RP alternatives are placed just below the 'root' of the tree, whereas the SP alternatives are each placed in a single alternative branch. For the SP observations, the average indirect utility of each of the 'dummy composite' alternatives (see the figure below) uses the theoretical basis of the inclusive value concept associated with linking levels in a nested logit model (McFadden (1981)) to define

$$U^{comp} = \sigma\log\sum_{j=1}^{J_{sp}} \exp(U_j),$$

in which the summation is taken over all alternatives in the nest corresponding to the composite alternative. Because each nest contains only one SP alternative, U^{comp} reduces to σU_{sp}, the expression for a single SP alternative, with every parameter including the unobserved component associated with the SP alternative scaled by σ. We refer to the estimation of the scaling approach as an artificial nested logit model because the approach acts as if we are estimating a traditional nested logit model. It draws on the empirical content of the inclusive value which links levels in a tree structure. The scaling parameter, σ, does not have to lie in the unit interval, the condition for consistency with random utility maximization (Hensher and Johnson (1981)), because individuals are not modeled as choosing from the full set of RP+SP alternatives. The scale for SP relative to RP can be greater than one.

```
                 Root
       RP         SP
       +-----------+---+---+---+
       |           |   |   |   |
 +---+---+---+     |   |   |   |
 |   |   |   |     |   |   |   |
RP1 RP2 RP3 RP4   SP1 SP2 SP3 SP4
```

Joint estimation involves 'stacking' the data. Consider an example of commuter mode choice, where we have one revealed preference and two stated choice observations, all from the same individual. As a practical consideration, we prefer to replicate the RP observations to make equal the RP and SP sample sizes. Otherwise, the SP data tend to dominate in estimation. The data are set up as follows, assuming two attributes, time and cost:

Mode	Time	Cost	Chosen	Index
RP car	40	2	1	1
RP train	60	3	0	2
RP bus	50	2	0	3
SP1 car	50	3	0	4
SP1 train	30	3	1	5
SP1 bus	40	2	0	6
RP car	40	2	1	1
RP train	60	3	0	2
RP bus	50	2	0	3
SP2 car	40	2	0	4
SP2 train	35	4	0	5
SP2 bus	50	3	1	6

In order to use this data set, it is necessary to replicate the full set of observations once for each RP choice situation, so that in each instance, only one choice is actually made. For the first SP choice situation in the three choice model above, we would have the expanded data set *(rpcar*,rptrain, rpbus,spcar,sptrain,spbus*), (*rpcar*,rptrain,rpbus,spcar,sptrain*,spbus*), where the starred choices are the ones chosen in each combined situation. The combined and expanded RP-SP data set is analyzed as the following tree:

; Tree = mode [(rpcar,rptrain,rpbus),(spcar),(sptrain),(spbus)]
; Ivset: (spcar,sptrain,spbus)

This tree structure will produce an inclusive value for the SP branches which is set to be the same across all three branches. Note that each branch in the SP part of the tree has only one degenerate alternative. We are actually 'tricking' the program in order to obtain an inclusive value parameter because this is the only observable way of identifying the scaling parameter, which is the parameter of the inclusive value.

If the sampling is choice based, rather than random, then a weighting scheme is appropriate. But, there will be no natural weighting in the population for the SP choices, so if a choice based sampling (WESML) estimator is to be used, the weights are only to apply to the RP choices. You can do this with *NLOGIT* with a minor variation to the usual setup. Suppose the model is built up from *n* RP alternatives and *m* SP choices. The **; Choices** setup with weights would appear as

```
; Choices = rp1,  rp2, ..., rpn,  sp1, sp2, ..., spm /
            wr1,  wr2, ..., wrn,  1.0,  1.0, ..., 1.0
```

That is, the usual set of weights is supplied for the RP alternatives (note that the order in your model might be different), while a 1.0 is given for the SP alternatives. The weights for the RP alternatives will sum to 1.0. When weights are given in this form, the choice based sampling weights,

$$W(j) = wRj / (pR_j/\Sigma_{j=RP\ alts}pR_j)$$

are computed for the RP alternatives while the counterpart for the SP alternatives is 1.0. Note that in the denominator, *pRj* is the sample proportion of individuals who chose alternative *Rj* among the full set of *n+m* alternatives, and that this is normalized by the sum over the RP alternatives. This way, the denominators in the *W*(*j*)s sum to 1.0 – but note that the *W*(*j*)s themselves do not sum to 1.0 because at least some of them are greater than 1.0.

N28.10.2 Sequential Estimation

The two data specifications can also be combined in the following way:

Step 1. Use the SP data by themselves to establish robust estimates of the individual's tradeoffs of the attributes in the stated choice experiment through the vector $\boldsymbol{\beta}_{sp}$ corresponding to $\mathbf{X}_{sp}$.

Step 2. Use the RP data to 'ground' the model in reality by estimating the alternative specific constants for the alternatives which are observed in the market. This ensures that the predicted aggregate model shares equal the observed RP shares. The RP model can be estimated with choice based weights. In estimating the choice specific constants, we make them conditional on the $\boldsymbol{\beta}_{rp}$ being constrained to equal $\boldsymbol{\beta}_{sp}$, but allowing for an errors-in-variables correction to $\mathbf{X}_{rp}$ through the estimation of a multiplicative scale factor, θ to rescale $\mathbf{X}_{rp}$ into the same units as $\mathbf{X}_{sp}$. The value of θ is selected so as to maximize the log likelihood for the overall model.

NLOGIT automates the search for θ with **; Scale (list of variables) = low,high,ncrude,nfine**. (For example, **; Scale (time,cost) = 0.2,1.2,11,11.**) See Section N18.10 for further discussion. Note that in sequential or joint estimation, the only attributes which are rescaled are those common to an alternative in both data sets and all of the attributes of an alternative which appears only in the SP model. Thus, the only attributes in the RP model which are not rescaled are those which are unique to the RP model.

N28.11 A Model of Covariance Heterogeneity

This is a modification of the two level nested logit model. The base case for the model is

$$P(j\,|\,b) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{j|b})}{\sum_{q=1}^{J|b}\exp(\boldsymbol{\beta}'\mathbf{x}_{q|b})}.$$

Denote the logsum, the log of the denominator, as I_b = inclusive value for branch b = IV(b). Then,

$$P(b) = \frac{\exp(\boldsymbol{\alpha}'\mathbf{y}_b + \tau_b I_b)}{\sum_{s=1}^{B}\exp(\boldsymbol{\alpha}'\mathbf{y}_s + \tau_s I_s)}.$$

The covariance heterogeneity model allows the τ_j inclusive value parameters to be functions of a set of attributes, $\mathbf{v}_j$, in the form

$$\tau_b^* = \tau_b \times \exp[\boldsymbol{\delta}'\mathbf{v}_b\,],$$

where $\boldsymbol{\delta}$ is a new vector of parameters to be estimated. Since the inclusive parameter is a scaling parameter for a common random component in the alternatives within a branch, this is equivalent to a model of heteroscedasticity.

The attributes, $\mathbf{v}_b$ may be any attributes – they are assumed to be the same for all alternatives in the branch, *b*. Also, $\mathbf{v}_b$ must not contain a constant (*one*). To use this option, just add

; Hfn = list of variables in $\mathbf{v}_b$

to the **NLOGIT** command. Once again, this option is available only for two level models. All other options for two level models remain as before. You can also obtain elasticities and marginal effects for probabilities with respect to the elements of $\mathbf{v}_b$. Just use

; Effects: variable [alts]

as usual. *NLOGIT* will figure out which branch applies from the tree structure. A separate set of results is given for variables in $\mathbf{v}_b$. If an attribute appears both in $\mathbf{y}_b$ and $\mathbf{v}_b$, there will be a separate table for the two different appearances. (This model must be specified in a command; it is not available in the command builder.)

The following illustrates the use of this model

```
NLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
           ; Tree = public(bus,train), private(air,car)
           ; Model: U(air)   = ba + bcost * gc + btime * ttme  /
                    U(train) = bt + bcost * gc + btime * ttme  /
                    U(car)   = bc + bcost * gc + btime * ttme  /
                    U(bus)   =      bcost * gc + btime * ttme
           ; Hfn = hinc
           ; Effects: hinc(*)/gc(*) $
```

```
-----------------------------------------------------------------------------
Covariance Heterogeneity Model
Dependent variable                    MODE
Log likelihood function      -188.96833
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Variable IV parameters are denoted s_...
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      BA|    3.92427***      .72034     5.45  .0000     2.51242   5.33612
   BCOST|    -.01750***      .00435    -4.02  .0001     -.02603   -.00897
   BTIME|    -.08606***      .01173    -7.34  .0000     -.10904   -.06308
      BT|     .90908***      .33711     2.70  .0070      .24835   1.56982
      BC|   -1.02251***      .37116    -2.75  .0059    -1.74997   -.29505
        |Inclusive Value Parameters
  PUBLIC|     .94983***      .31909     2.98  .0029      .32441   1.57524
 PRIVATE|    1.65970***      .61495     2.70  .0070      .45441   2.86498
Lmb[1|1]|        1.0    .....(Fixed Parameter).....
Trunk{1}|        1.0    .....(Fixed Parameter).....
        |Covariates in Inclusive Value Parameters
  s_HINC|     .01324**       .00662     2.00  .0454      .00027    .02621
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
HINC    |
--------+-----------------------------------
     BUS|   .5592     .5592    -.2583   -.2583
   TRAIN|   .5592     .5592    -.2583   -.2583
     AIR|  -.9771    -.9771     .4513    .4513
     CAR|  -.9771    -.9771     .4513    .4513

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |     BUS    TRAIN      AIR      CAR
--------+-----------------------------------
     BUS| -1.9762     .0409     .3905    .3905
   TRAIN|  -.0793   -2.3578     .8340    .8340
     AIR|  1.4332    1.4332   -1.6629    .1335
     CAR|  1.3260    1.3260    -.2692  -1.9390
```

N28.12 The Generalized Nested Logit Model

The generalized nested logit model is an extension of the nested logit model in which alternatives may appear in more than one branch. (The behavioral assumptions underlying this model are up to the user.) Alternatives which appear in more than one branch are allocated across branches probabilistically. The model estimated includes the usual nested logit framework (only two levels are supported in this framework), as well as the matrix of allocation parameters. The only difference between this and the more basic nested logit model is the specification of the tree. The model is requested by changing the command name to **GNLOGIT**. Otherwise, the model is the same as the nested logit model. The alternative form

NLOGIT ; GNL ; ...

is also useable. All features of *NLOGIT*, including marginal effects, simulations, etc. are the same as for all other models. The difference here is that when you specify the tree, you may specify that a given alternative appears in more than one branch. (Technical details appear at the end of this section.)

A small example appears below. In this nested logit model, the choice *car* appears in both branches. The probabilities for the allocation are estimated to be .16 and .84. The base case multinomial logit model appears first.

```
GNLOGIT   ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = one,gc,ttme
          ; Tree = private(air,car), ground(car,train,bus)
          ; Effects: gc(*) $
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function      -199.97662
Estimation based on N =    210, K =   5
Inf.Cr.AIC  =    410.0 AIC/N =    1.952
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2862
Chi-squared[ 2]          =    167.56429
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
--------+--------------------------------------------------------------------
      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   7.06194
   A_CAR|    3.92300***      .44199     8.88  .0000     3.05671   4.78929
  A_1BUS|    3.21073***      .44965     7.14  .0000     2.32943   4.09204
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Generalized Nested Logit Model
Dependent variable                   MODE
Log likelihood function        -195.43541
The model has 2 levels.
GNL: Model uses random utility form RU1
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    MODE|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    -.02140**       .01030    -2.08  .0379     -.04159   -.00120
    TTME|    -.09368**       .04016    -2.33  .0197     -.17240   -.01496
   A_AIR|    5.30728**      2.67168     1.99  .0470      .07088  10.54367
   A_CAR|    4.21064**      2.00982     2.10  .0362      .27147   8.14980
  A_1BUS|    3.47823**      1.68141     2.07  .0386      .18273   6.77373
        |Dissimilarity parameters. These are mu(branch).
 PRIVATE|    1.95202      1.30315     1.50  .1342     -.60211   4.50615
  GROUND|     .80675       .56368     1.43  .1524     -.29805   1.91155
        |Structural MLOGIT Allocation Model: Constants
tAIR_PRI|        0.0    .....(Fixed Parameter).....
tTRA_GRO|        0.0    .....(Fixed Parameter).....
tBUS_GRO|        0.0    .....(Fixed Parameter).....
tCAR_PRI|   -1.62462     16.42213     -.10  .9212   -33.81141  30.56217
tCAR_GRO|        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.
-----------------------------------------------------------------------------
```

```
Generalized Nested Logit
Estimated Allocations of Choices to Branches
Estimated standard errors in parentheses for
allocation values not fixed at 1.0 or 0.0.
        |Branch
--------+-----------------
CHOICE  |PRIVATE  GROUND
--------+--------+--------
AIR      1.0000    .0000
TRAIN     .0000   1.0000
BUS       .0000   1.0000
CAR       .1646    .8354
        ( .0000) ( .0000)
Note: Allocations are multinomial logit
probabilities. Underlying parameters are not
shown in the output:
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |     AIR      CAR      CAR    TRAIN
--------+------------------------------------
     AIR| -1.2007    .6088    .6088    .2953
     CAR|   .6587  -2.5515    .9015    .7905
     CAR|   .3285    .4473  -2.6094    .3941
   TRAIN|   .2449    .7727    .7727  -1.3112
```

Aside from the expanded specification of the tree, the model is otherwise the same as the nested logit model shown earlier. The model contains an allocation matrix,

$$\boldsymbol{\alpha} = [\alpha_{k|j}],$$

which defines the probabilistic allocation of alternatives *k* to branches *j*. The columns of the matrix relate to the branches while the rows refer to the alternatives. The model construction specifies that the rows of the matrix each sum to 1.0. The matrix that was estimated for the model in the example was

```
        |Branch
--------+-----------------
CHOICE  |PRIVATE  GROUND
--------+--------+--------
AIR      1.0000    .0000
TRAIN     .0000   1.0000
BUS       .0000   1.0000
CAR       .1646    .8354
```

The locations of the nonzero entries are specified by the tree definition. In the nested logit model, each row will contain a single 1.0000 and *J*-1 0.0000s. When alternatives appear in more than one branch, then a set of allocation parameters appear in the matrix. These are parameters to be estimated. When there are free parameters to be estimated in $\boldsymbol{\alpha}$, the adding up constraint is imposed by using a multinomial logit form,

$$\alpha_{k|j} = \text{Prob(alternative } k \text{ is in branch } j) = \exp(\theta_{k|j}) \,/\, \Sigma_{k,m} \exp(\theta_{j|m}),$$

where the parameters θ are actually estimated by the program. Note the denominator summation is over branches that the alternative appears in. The probabilities sum to one. The identification rule that one of the θs for each alt modeled equals one is imposed. Thus, in the output results above, $\theta_{car,ground} = 0$ and $\theta_{car,private} = -1.625$, so that the probability allocated to the *private* branch is exp(-1.625)/[exp(0)+exp(-1.625)] = 0.1646, which can be seen in the final table of results. You may also specify that these allocations depend on an individual characteristic (not a choice attribute), such as *income*, by using

; GNL = the name of a variable

(Note that even if you use the **GNLOGIT** command, you must have the **; GNL** specification in the command.) In this instance, the multinomial logit probabilities become functions of this variable,

$$\alpha_{k|j} = \text{Prob(alternative } k \text{ is in branch } j) = \exp(\theta_{k|j}+\gamma_{k|j}) \,/\, \Sigma_{k,m} \exp(\theta_{j|m}+\gamma_{k|m}).$$

Again, to achieve identification, one of the θs and one of the γs is set equal to zero. The log likelihood function is then assembled from these parameters as follows:

$$Prob(j\,|\,b) = \frac{\left[\alpha_{j|b}\exp(V_j)\right]^{1/\mu_b}}{\sum_{q=1}^{J}\left[\alpha_{q|b}\exp(V_q)\right]^{1/\mu_b}},$$

$$Prob(b) \quad = \frac{\left\{\sum_{q=1}^{J}\left[\alpha_{q|b}\exp(V_q)\right]^{1/\mu_b}\right\}^{\mu_b}}{\sum_{s=1}^{B}\left\{\sum_{q=1}^{J}\left[\alpha_{q|s}\exp(V_q)\right]^{1/\mu_s}\right\}^{\mu_s}}.$$

Derivatives of this log likelihood function are computed numerically, using two sided finite differences. The BHHH estimator is used for the asymptotic covariance matrix.

N28.13 Box-Cox Nested Logit Model

This variant of the nested logit model allows some attributes to be transformed using the Box-Cox transformation. The model specification adds a degree of flexibility to the functional form. The model specification is the general nested logit form, with

$$U(j) = \sum_{k=1}^{B}\beta_k\left(\frac{x_{jk}^{\lambda_k}-1}{\lambda_k}\right) + \sum_{m=1}^{K}\beta_m x_{jm} + \sum_{j=1}^{J}\sum_{c=1}^{C} d_{jc} z_c + \varepsilon_j\,.$$

The utility function contains B attributes, x_{jb} that are transformed, each by an attribute specific transformation parameter, λ_b. It also contains K attributes, x_{jk} that are untransformed – this is the form we have assumed up to this point. Finally, there may be C variables, z_c that are interacted with alternative specific constants. Again, this is the form we have used up to this point. Save for the first term, this is the same model we have used before.

The command setup is

```
NLOGIT      ; Lhs = ... ; Choices = ...
            ; Tree = specification
            ; Rhs = choice varying attributes
            ; Rh2 = choice invariant characteristics and one
            ; ... any other options
            ; Bcl =  list of attributes among the Rhs variables that are
                     subject to the Box-Cox transformation $
```

The utility functions must be in the Rhs/Rh2 format for this specification. An example is

```
NLOGIT      ; Lhs = mode ; Choices = air,train,bus,car
            ; Tree = private(air,car),public(train,bus)
            ; Rhs = gc,invc,invt
            ; Rh2 = one,hinc
            ; Bcl = invc,invt
            ; Effects: gc(*) / invt(*) $
```

The results below compare the Box-Cox model to the model based on the untransformed variables.

```
-----------------------------------------------------------------------------
Box-Cox Nested Logit Model
Dependent variable                 MODE
Log likelihood function      -212.68485
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|      .01954**       .00887     2.20  .0276       .00216     .03693
    INVC|     -.06628         .04760    -1.39  .1638     -.15957     .02701
    INVT|     -.28549         .27341    -1.04  .2964     -.82136     .25038
   A_AIR|   -3.53251***      1.18141    -2.99  .0028    -5.84802   -1.21699
AIR_HIN1|      .01245         .01145     1.09  .2769     -.01000     .03490
 A_TRAIN|     -.01422         .50666     -.03  .9776    -1.00726     .97883
TRA_HIN3|     -.00582         .00761     -.76  .4446     -.02073     .00910
   A_BUS|     -.83602         .62644    -1.33  .1820    -2.06382     .39179
BUS_HIN4|      .00063         .01241      .05  .9598     -.02371     .02496
        |IV parameters, tau(b|l,r),sigma(l|r),phi(r)
 PRIVATE|     4.61679***     1.73915     2.65  .0079     1.20811    8.02547
  PUBLIC|     4.19463***     1.57932     2.66  .0079     1.09922    7.29005
        |Box-Cox Transformation Parameters
  bcINVC|      .76751***      .19128     4.01  .0001      .39261    1.14241
  bcINVT|      .41250***      .15108     2.73  .0063      .11640     .70860
--------+--------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |      AIR       CAR     TRAIN      BUS
--------+------------------------------------
     AIR|   2.8005     .7946   -2.2198   -2.2198
     CAR|   1.2956    3.1602   -2.4029   -2.4029
   TRAIN|  -2.8203   -2.8203    4.7134    2.1691
     BUS|  -1.3942   -1.3942    1.2654    3.5177

Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
INVT    |      AIR       CAR     TRAIN      BUS
--------+------------------------------------
     AIR|  -2.9548    -.8361    2.1269    2.1269
     CAR|  -2.7409   -6.5490    5.3720    5.3720
   TRAIN|   4.8023    4.8023   -7.0336   -3.1028
     BUS|   2.5239    2.5239   -2.1424   -6.1462
```

```
-----------------------------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -223.81970
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Number of obs.=   210, skipped    0 obs
```

MODE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Attributes in the Utility Functions (beta)					
GC	.00199	.00827	.24	.8099	-.01421	.01819
INVC	-.00266	.00863	-.31	.7578	-.01958	.01426
INVT	-.00325**	.00133	-2.45	.0143	-.00586	-.00065
A_AIR	-1.40526***	.35771	-3.93	.0001	-2.10635	-.70417
AIR_HIN1	.00192	.00468	.41	.6810	-.00725	.01109
A_TRAIN	.01699	.21993	.08	.9384	-.41406	.44803
TRA_HIN3	-.00813	.00582	-1.40	.1625	-.01954	.00328
A_BUS	-.97208***	.32416	-3.00	.0027	-1.60743	-.33673
BUS_HIN4	.00173	.00852	.20	.8393	-.01497	.01843
	IV parameters, tau(b|l,r),sigma(l|r),phi(r)					
PRIVATE	12.2211***	3.50815	3.48	.0005	5.3453	19.0970
PUBLIC	7.49804***	2.15617	3.48	.0005	3.27203	11.72405

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Elasticity wrt change of X in row choice on Prob[column choice]

GC	AIR	CAR	TRAIN	BUS
AIR	.7003	.4962	-.6973	-.6973
CAR	.3736	.5633	-.5596	-.5596
TRAIN	-.5419	-.5419	.8111	.5523
BUS	-.2299	-.2299	.2749	.5040

Elasticity wrt change of X in row choice on Prob[column choice]

INVT	AIR	CAR	TRAIN	BUS
AIR	-1.5410	-1.1059	1.4256	1.4256
CAR	-3.7304	-5.5957	5.4474	5.4474
TRAIN	4.3326	4.3326	-6.0594	-4.0800
BUS	2.0758	2.0758	-2.4286	-4.4770

N29: Random Parameters Logit Model

N29.1 Introduction

The random parameters logit (RPL) model, also referred to as the mixed logit model, is the most general model form in *NLOGIT* in terms of the variety of model specifications it can accommodate and in terms of the range of behavior that it can model. (On this latter point, see McFadden and Train (2000).) This chapter will develop the numerous different specifications of the model that can be accommodated.

NLOGIT offers an extensive set of specifications within the mixed logit structure. This model is gaining great popularity in applications. Capabilities provided by the estimator include

- Choosing from among a large number of analytical distributions for each random parameter
- Accounting for the non-independence between observations associated with the same respondent (a theme of importance in stated choice studies)
- Decomposing the mean and standard deviation of one or more random parameters to reveal sources of systematic taste heterogeneity
- Accounting for correlation of random parameters
- Imposing priors based on known choices in model estimation
- Imposing constraints on distributions (e.g. constraining the triangular or normal to ensure that it does not change sign over its range)
- Selecting subsets of pre-specified variables to interact with the mean and standard deviation of random parameterized attributes
- Deriving willingness to pay estimates when both the numerator and denominator are random parameter estimates

We note before beginning that this model also includes the error components model presented in Chapter N30. The error components can be simply included as part of the mixed logit model. This is described in Section N29.5. The random parameters model also includes the nonlinear random parameters model in Chapter N31, the latent class random parameters model in Chapter N32 and the generalized mixed logit model in Chapter N33.

N29.2 Random Parameters (Mixed) Logit Models

This model is somewhat similar to the random coefficients model for linear regressions. (See Bhat (1996), Jain, Vilcassim, and Chintagunta (1994), Revelt and Train (1998), and Berry, Levinsohn, and Pakes (1995).) The model formulation is a one level multinomial logit model, for individuals $i = 1,...,N$ in choice setting t. Neglecting for the moment the error components aspect of the model, we begin with the basic form of the multinomial logit model, with (optional) alternative specific constants α_{ji} and attributes $\mathbf{x}_{ji}$,

$$\text{Prob}(y_{it} = j) = \frac{\exp\left(\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{ji}\right)}{\sum_{q=1}^{J_i}\exp\left(\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qi}\right)}.$$

The RPL model emerges as the form of the individual specific parameter vector, $\boldsymbol{\beta}_i$ is developed. The most familiar, simplest version of the model specifies

$$\beta_{ki} = \beta_k + \sigma_k v_{ik},$$

and

$$\alpha_{ji} = \alpha_j + \sigma_j v_{ji},$$

where β_k is the population mean, v_{ik} is the individual specific heterogeneity, with mean zero and standard deviation one, and σ_k is the standard deviation of the distribution of β_{ik}s around β_k. The term 'mixed logit' is often used in the literature (e.g., Revelt and Train (1998)) for this model. The choice specific constants, α_{ji} and the elements of $\boldsymbol{\beta}_i$ are distributed randomly across individuals with fixed means. A refinement of the model is to allow the means of the parameter distributions to be heterogeneous with observed data, $\mathbf{z}_i$, (which does not include *one*). This would be a set of choice invariant characteristics that produce individual heterogeneity in the means of the randomly distributed coefficients so that

$$\beta_{ki} = \beta_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \sigma_k v_{ki},$$

and likewise for the constants. The model is not limited to the normal distribution. We consider several alternatives below. One important variation is the lognormal model,

$$\beta_{ki} = \exp(\rho_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \sigma_k v_{ki}).$$

The v_{jki}s are individual and choice specific, unobserved random disturbances – the source of the heterogeneity. Thus, as stated above, in the population, if the random terms are normally distributed,

$$\alpha_{ji} \text{ or } \beta_{ki} \sim \text{Normal or Lognormal } [\rho_{j\ or\ k} + \boldsymbol{\delta}_{j\ or\ k}'\mathbf{z}_i, \sigma_{j\ or\ k}^2].$$

(Other distributions may be specified.) For the full vector of K random coefficients in the model, we may write the full set of random parameters as

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i.$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix which contains σ_k on its diagonal. For convenience at this point, we will simply gather the parameters, choice specific or not, under the subscript 'k.' (The notation is a bit more cumbersome for the lognormally distributed parameters. We will return to that in the technical details.)

We can go a step further and allow the random parameters to be correlated. All that is needed to obtain this additional generality is to allow $\boldsymbol{\Gamma}$ to be a triangular matrix with nonzero elements below the main diagonal. Then, the full covariance matrix of the random coefficients is $\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Gamma}'$. The standard case of uncorrelated coefficients has $\boldsymbol{\Gamma} = \text{diag}(\sigma_1,\sigma_2,\ldots,\sigma_k)$. If the coefficients are freely correlated, $\boldsymbol{\Gamma}$ is a full, unrestricted, lower triangular matrix and $\boldsymbol{\Sigma}$ will have nonzero off diagonal elements. (It will be convenient to aggregate this one step further. We may gather the entire parameter vector for the model in this formulation simply by specifying that for the nonrandom parameters in the model, the corresponding rows in $\boldsymbol{\Delta}$and $\boldsymbol{\Gamma}$ are zero.) We will also define the data and parameter vector so that any choice specific aspects are handled by appropriate placements of zeros in the applicable parameter vector.

An additional extension of the model allows the distribution of the random parameters to be heteroscedastic. As stated above, the variance of v_{ik} is taken to be a constant. The model is made heteroscedastic by assuming, instead, that

$$\text{Var}[v_{ik}] = \sigma_{jk}^2\,[\exp(\boldsymbol{\omega}_k'\mathbf{hr}_i)]^2.$$

A convenient way to parameterize this is to write the full model as

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma\Omega}_i\mathbf{v}_i$$

where $\boldsymbol{\Omega}_i$ is the diagonal matrix of individual specific variance terms; $\omega_{ik} = \exp(\boldsymbol{\omega}_k'\mathbf{hr}_i)$.

The list of variations above produces an extremely flexible, general model. Typically, you would use only some of them, though in principle, all could appear in the model at once. We will develop them in parts in the sections to follow. A convenient form of the full random parameters logit model to begin with is

$$\text{Prob}(y_{it} = j) = \frac{\exp(\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit})}{\sum_{q=1}^{J_{it}}\exp(\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit})}.$$

Finally, an additional layer of individual heterogeneity may be added to the model in the form of the error components detailed in Chapter N30. The full model with all components is

$$\text{Prob}(y_{it} = j) = \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{jm}\theta_m\exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{qm}\theta_m\exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im}\right]},$$

where the components of the model are as follows:

Random Alternative Specific Constants and Taste Parameters:

$(\alpha_{ji},\boldsymbol{\beta}_i) = (\alpha_j,\boldsymbol{\beta}) + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma\Omega}_i\mathbf{v}_i$, $\boldsymbol{\Omega}_i =$ diag(ω_{i1}, ω_{i2}, ...) or $\boldsymbol{\Omega}_i = \text{diag}(\sigma_1,\ldots,\sigma_k)$

$\boldsymbol{\beta},\alpha_{ji}$ = constant terms in the distributions of the random taste parameters

Uncorrelated Parameters with Homogeneous Means and Variances

β_{ik} = $\beta_k + \sigma_k v_{ik}$ when $\boldsymbol{\Delta} = \mathbf{0}$, $\boldsymbol{\Gamma} = \mathbf{I}$, $\boldsymbol{\Omega}_i = \text{diag}(\sigma_1,\ldots,\sigma_k)$

$\mathbf{x}_{jit}$ = all observed choice attributes and individual characteristics

$\mathbf{v}_i$ = random unobserved taste variation, with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$

Uncorrelated Parameters with Heterogeneous Means and Variances

β_{ik} = $\beta_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \sigma_k \exp(\boldsymbol{\omega}_k'\mathbf{hr}_i)v_{ik}$ when $\boldsymbol{\Gamma} = \mathbf{I}$, $\boldsymbol{\Omega}_i = \text{diag}(\omega_{i1}, \omega_{i2}, ...)$

$\boldsymbol{\Delta}$ = parameters that enter the heterogeneous means of the distributions of the random parameters; $\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i$ = the heterogeneous means

ω_{ik} = $\exp(\boldsymbol{\omega}_k'\mathbf{hr}_i)$ = heterogeneity in the variances of the distributions of the random parameters

$\boldsymbol{\omega}_k$ = parameters in the variance heterogeneity of the random parameters

σ_{ik} = $\sigma_k\omega_{ik}$ = heterogeneous standard deviations in the distributions of the random parameters; $\sigma_{ik} = \sigma_k$ in a homoscedastic model

$\mathbf{z}_i$ = observed variables that measure the heterogeneity in the means of the random parameters

$\mathbf{hr}_i$ = observed variables that measure the heterogeneity in the variances of the random parameters

Correlated Parameters with Heterogeneous Means

β_{ik} = $\beta_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \Sigma_{s=1}^{k} \Gamma_{ks}v_{is}$ when $\boldsymbol{\Gamma} \neq \mathbf{I}$, and $\boldsymbol{\Omega}_i = \text{diag}(\sigma_1,...,\sigma_k)$

$\boldsymbol{\Gamma}$ = lower triangular matrix with ones on the diagonal that allows correlation across random parameters when $\boldsymbol{\Gamma} \neq \mathbf{I}$

Individual Error Components

E_{im} = the individual specific underlying random error components, $m = 1,...,M$, $E_{im} \sim N[0,1]$

d_{jm} = 1 if E_{im} appears in utility for alternative j and 0 otherwise

θ_m = scale factor for error component m

γ_{im} = $\exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)$ = heterogeneity in the variances of the error components

λ_{im} = $\theta_m\gamma_{im}$ = standard deviations of random error components

$\boldsymbol{\gamma}_m$ = parameters in the heteroscedastic variances of the error components

$\mathbf{he}_i$ = individual choice invariant characteristics that produce heterogeneity in the variances of the error components

The model specification will dictate which parameters are random and which are not, how the heteroscedasticity, if any, is parameterized, the distributions of the random terms, and how the error components enter the model.

The probabilities defined above are conditioned on the random terms, $\mathbf{v}_i$ and the error components, $\mathbf{E}_i$. The unconditional probabilities are obtained by integrating v_{ik} and E_{im} out of the conditional probabilities: $P_j = E_{\mathbf{v},\mathbf{E}}[P(j|\mathbf{v}_i,\mathbf{E}_i)]$. This is a multiple integral which does not exist in closed form. The integral is approximated by sampling *nrep* draws from the assumed populations and averaging. (See Bhat (1996) and Revelt and Train (1998) and Greene (2012) for discussion.) Parameters are estimated by maximizing the simulated log likelihood,

$$\log L_s = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_{ir}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im,r}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_{ir}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im,r}\right]},$$

with respect to ($\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$,$\boldsymbol{\Omega}$, $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$), where

R = the number of replications,

$\boldsymbol{\beta}_{ir}$ = $\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\boldsymbol{\Omega}_i\mathbf{v}_{ir}$ = the rth draw on $\boldsymbol{\beta}_i$,

$\mathbf{v}_{ir}$ = the rth multivariate draw for individual i,

$E_{im,r}$ = the rth univariate normal draw on the underlying effect for individual i.

(Note that the multivariate draw, $\mathbf{v}_{ir}$ is actually K independent draws. The heteroscedasticity is induced first by multiplying by $\boldsymbol{\Omega}_i$, then the correlation is induced by multiplying $\boldsymbol{\Omega}_i\mathbf{v}_{ir}$ by $\boldsymbol{\Gamma}$.) Technical details on the estimation procedure are given in Section N29.11.

The model components may be restricted and varied in several ways.

- A variety of distributions may be chosen for the random parameters, and they need not be the same for all parameters.
- The observed heterogeneity, $\boldsymbol{\Delta}\mathbf{z}_i$, is optional. You may specify that a coefficient is randomly distributed around a fixed mean. Thus, $\boldsymbol{\delta}_k$ may be set to a zero vector for some or all random coefficients.
- σ_k may be set equal to zero for some coefficients. This may change the way a coefficient enters the model. If $\sigma_k = 0$ and $\boldsymbol{\delta}_k = \mathbf{0}$, then the coefficient is a nonrandom fixed parameter. But, including it in $\boldsymbol{\beta}$ allows you to force a coefficient to be positive. This device also allows you to form a hierarchical model with nonrandom coefficients.
- Any coefficient in the model may be fixed at a specific value.
- The heteroscedasticity may apply to some or all (or none) of the random parameters.
- Different variables may be placed in the heterogeneous means ($\boldsymbol{\Delta}\mathbf{z}_i$) or the heteroscedastic variances ($\boldsymbol{\Omega}_i$) of any of the random parameters.
- The variables that enter the heteroscedasticity of the error components may be different.
- The model with both heteroscedasticity and cross parameter correlation is not estimable. (There is no way to make the covariance heterogeneous.)

A number of additional features are listed in the sections to follow.

N29.3 Command for the Random Parameters Logit Models

The command for the mixed logit model is as follows:

```
RPLOGIT    ; Lhs = ... as usual
           ; Choices = ...
           ; ... Utility function specification using
           ; Rhs = ... ; Rh2 = ...  or
           ; Model: U(...) = ... to specify utilities
           ; Fcn = specification of random parameters $
```

(The model command **NLOGIT ; RPL** is equivalent.) The last specification is used to define the random parameters. There are many variants. We begin with the simplest, and add features as we proceed. The model as specified is a random parameters, multinomial logit model based on random utility. This may be changed to random regret by using the model command

```
RPRRLOGIT ; …
```

The **; Fcn** specification takes the basic form

```
; Fcn = parameter label (type)
```

where 'parameter label' is defined either by a variable name that you use in your **; Rhs** specification or by the name you give in your **; Model:...** definitions and the 'type' is one of the distributions defined in the next section. Alternative specific constants are a special case. You will generally not want to specify the parameters that multiply Rh2 variables as random. These two cases are considered specifically below. For example, the following specifies two normally distributed random parameters:

```
RPLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
           ; Rhs = gc,ttme,invc ; Rh2 = hinc
           ; Fcn = gc(n),ttme(n) $
```

(The 'type' in the example is 'n' indicating normally distributed parameters. Several other specifications would probably be added.) Alternatively, you might use the following to specify a model with two random parameters:

```
RPLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
           ; Model:U(air) = a_air+bgc*gc+btt*ttme+binvc*invc+ghinc*hinc/
                   U(train,bus,car) = a_ground+bgc*gc
           ; Fcn = a_ground(n),btt(n) $
```

Note that the specifications of the random parameters are separated by commas, not semicolons. The next several subsections will describe the various parts of the specifications of the random parameters. The last part of this section describes the command builder for this model. Because so much of this model is custom made for the particular application, the command builder is somewhat limited compared to the command form indicated above.

N29.3.1 Distributions of Random Parameters in the Model

There are many distributions that can be (and have been) used for the random parameters. The most common will be the normal, which is used in the example above. Many alternatives are supported, however. Some of these should be viewed as experimental. Moreover, we note that as such, some of these choices may not perform well in a particular data set. The normal distribution is a natural choice for a random parameter, based on ideas of individual heterogeneity and the central limit theorem. It is difficult to motivate, e.g., the scaled beta on this basis. Some useful special cases others are described further in Section N29.3.8.) The basic distributions are specified with the following:

; Fcn = parameter name (type), ...

The types are

1	*c*	nonstochastic	$\beta_i = \beta$
2	*n*	normal	$\beta_i = \beta + \sigma v_i, v_i \sim$ N[0,1]
3	*s*	skew normal	$\beta_i = \beta + \sigma v_i + \lambda\|w_i\|, v_i, w_i \sim$ N[0,1]
4	*l*	lognormal	$\beta_i = \exp(\beta + \sigma v_i), v_i \sim$ N[0.1]
5	*z*	truncated normal	$\beta_i = \beta + \sigma v_i, v_i \sim$ truncated normal (-1.96 to 1.96)
6	*u*	uniform	$\beta_i = \beta + \sigma v_i, v_i \sim$ U[-1,1]
7	*f*	one sided uniform	$\beta_i = \beta + \beta v_i, v_i \sim$ uniform[-1,1]
8	*t*	triangular	$\beta_i = \beta + \sigma v_i, v_i \sim$ triangle[-1,1]
9	*o*	one sided triangular	$\beta_i = \beta + \beta v_i, v_i \sim$ triangle[-1,1]
10	*d*	beta, dome	$\beta_i = \beta + \sigma v_i, v_i \sim 2\times$beta(2,2) - 1
11	*b*	beta, scaled	$\beta_i = \beta v_i, v_i \sim$ beta(3,3)
12	*e*	Erlang	$\beta_i = \beta + \sigma v_i, v_i \sim$ gamma(1,4) - 4
13	*g*	gamma	$\beta_i = \exp(\beta + \sigma v_i), v_i = \log(-\log(u1*u2*u3*u4))$
14	*w*	Weibull	$\beta_i = \beta + \sigma v_i, v_i = 2(-\log u_i)^{\sqrt{.5}}, u_i \sim$ U[0,1]
15	*r*	Rayleigh	$\beta_i = \exp(\beta_i$ (Weibull))
16	*p*	exponential	$\beta_i = \beta + \sigma v_i, v_i \sim$ exponential - 1
17	*q*	exponential, scaled	$\beta_i = \beta v_i, v_i \sim$ exponential
18	*x*	censored (left)	$\beta_i = \max(0, \beta_i$ (normal))
19	*m*	censored (right)	$\beta_i = \min(0, \beta_i$ (normal))
20	*v*	exp(triangle)	$\beta_i = \exp(\beta_i$ (triangular))
21	*i*	type I extreme value	$\beta_i = \beta + \sigma v_i, v_i \sim$ standard Gumbel

In the list above, we have denoted the constant in the distribution as 'β.' However, the parameter definition may involve heterogeneity in the mean – see Section N29.3.4 – so, what appears there may be of the form $\theta_i = \beta + \boldsymbol{\delta}'\mathbf{z}_i$. We have also written the scaling parameter in each form as 'σ,' however, you may also specify heterogeneity in the variances – see Section N29.4 – so what appears there may be of the form $\sigma_i = \sigma\exp(\boldsymbol{\omega}'\mathbf{h}_i)$. The list above suggests the variety of different distributions that may be used. Numerous modifications and restrictions are shown in Section N29.3.8.

Any distribution may be used for any parameter. The normal distribution will be the usual choice. However, you may wish to restrict a particular coefficient in the model to be positive. The lognormal distribution is the obvious choice, though there are several other possibilities. The normal, lognormal, skew normal, exponential, Erlang, Rayleigh and Weibull distributions all have infinite ranges. If you wish to restrict the range of variation of a parameter, then the triangular, dome or uniform can be used. The lognormal distribution has an infinite tail in the positive direction and is anchored at zero while the exponential, Erlang and Weibull models as specified have infinite range from $\beta - \sigma E[v_i]$ to $+\infty$. Section N29.3.8 shows how to restrict these distributions so that they, like the lognormal, are anchored at zero. As shown there, however, these models will differ in that the support of the distributions may be the negative or the positive half line.

It is important to note that the means and variances of the distributions are not always simple functions when the parameters are not linear functions of the underlying random variables. For many of the distributions shown above, the mean of v_i is zero, which centers the distributions at β. For the lognormal, skew normal, Weibull and several other models, the mean depends on the parameters. This is also true of the modified distributions shown below. This means that one must be careful in interpreting the estimated coefficients, even in simple cases in which there is no heterogeneity in the means or variances. It is possible to learn about these empirically, as described in Section N29.8, however, it is often not possible to state a priori what the population means are for most of the distributions. The problem becomes yet more complicated as additional features such as heterogeneity in the means and heteroscedasticity are added to the model.

Some practical aspects of the specifications are as follows:

- If you will be mixing distributions, the specification of correlated parameters, while allowable, produces ambiguous results. The nature of the correlation is difficult to define. However, the program will have no unusual difficulty estimating a model in which correlated parameters have different distributions. One particular case worth noting is a mixture of normal and lognormal parameters. In such a model, the reported correlation will be between the normally distributed parameter and the log of the lognormally distributed parameter. This is probably not a useful result.

- Researchers often find that the long, thick tail of the lognormal distribution produces an implausible distribution of parameters. The restricted triangular distribution as well as several alternatives described in Section N29.3.8 may be preferable. The skew normal distribution appears to be a very promising alternative.

- Type '*c*' is the same as not including the parameter in the Fcn list, which is how this usually should be done. But sometimes, for convenience, this might be preferred. Variable name(*c*) specifies a free mean and zero variance of the parameter.

Model results for these distributions will display the structural parameters, not necessarily the means and variances of the parameter distributions. Note, for example, that the means of the lognormal and the Weibull distributions are not equal to β; for the lognormal it is $\exp(\beta+\sigma^2/2)$ while for the Weibull it is $\beta+2\sigma\Gamma(1+1/\sqrt{2})$. Consider an example. The following estimates a model with two random parameters. We will use the normal, Weibull and exponentiated Weibull (our 'Rayleigh') distributions. Since the exponentiated Weibull estimator forces the coefficient to be positive, and the coefficients on the two variables would naturally be negative, we reverse the signs on the data before estimation.

The commands are:

```
CREATE      ; mgc = -gc ; mttme = -ttme $
RPLOGIT     ; Lhs = mode  ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme
            ; Rh2 = one
            ; Fcn = mgc(n),mttme(n) ? Normally distributed parameters
            ; Maxit = 50 ; Pts = 25 ; Halton ; Pds = 3 $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme
            ; Rh2 = one
            ; Fcn = mgc(w),mttme(w) ? Weibull distributed parameters
            ; Maxit = 50 ; Pts = 25 ; Halton ; Pds = 3 $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme
            ; Rh2 = one
            ; Fcn = mgc(r),mttme(r) ? Modified Weibull distributed parameters
            ; Maxit = 50 ; Pts = 25 ; Halton ; Pds = 3 $
```

These are the reported random parameter estimates. (The nonrandom alternative specific constants are not shown.) The values for the random parameters are β and σ. For the normally distributed variables, these are the means and standard deviations. For the other distributions, they are only the structural parameters. To see the similarity, however, note for the coefficient on *mgc* in the Rayleigh model, exp(-3.35979) is about 0.034, which resembles the value for the normal distribution. Accounting for σ would likely bring them yet closer. Section N29.8 considers methods of examining these effects empirically.

```
--------+Multinomial logit with nonrandom parameters
     MGC|     .01578***      .00438     3.60  .0003      .00719    .02437
   MTTME|     .09709***      .01044     9.30  .0000      .07664    .11754
--------+Normal Random parameters in utility functions
        |Random parameters in utility functions
     MGC|     .02167***      .00676     3.20  .0014      .00842    .03493
   MTTME|     .14113***      .01952     7.23  .0000      .10287    .17938
        |Distns. of RPs. Std.Devs or limits of triangular
   NsMGC|     .00762         .01342      .57  .5702     -.01869    .03393
 NsMTTME|     .07420***      .01494     4.97  .0000      .04492    .10347
--------+Weibull Random parameters in utility functions
        |Random parameters in utility functions
     MGC|     .03194         .01957     1.63  .1027     -.00642    .07030
   MTTME|     .23823***      .03315     7.19  .0000      .17327    .30320
        |Distns. of RPs. Std.Devs or limits of triangular
   WsMGC|     .00507         .00887      .57  .5673     -.01231    .02246
 WsMTTME|     .05594***      .01258     4.45  .0000      .03129    .08059
--------+Rayleigh  Random parameters in utility functions
     MGC|   -3.35979**      1.38032    -2.43  .0149    -6.06516   -.65442
   MTTME|   -1.26343***      .21593    -5.85  .0000    -1.68664   -.84021
        |Distns. of RPs. Std.Devs or limits of triangular
   RsMGC|     .32940         .90086      .37  .7146    -1.43626   2.09507
 RsMTTME|     .47275***      .10965     4.31  .0000      .25784    .68765
```

N29.3.2 Spreads, Scaling Parameters and Standard Deviations

As evident in Section N29.2, with all its different components, the RPL model is complicated. It is also necessary to note that the interpretation of the parameters is partly a function of the specification chosen. What are described earlier as the 'means' and 'variances' are actually only those parameters in the simplest cases. The reported parameters may need to be interpreted, and manipulated further to obtain the expected results. We consider several examples. In a model with a normally distributed parameter,

$$\beta_i = \beta + \delta z_i + \sigma v_i,\ v_i \sim N[0,1],$$

$(\beta + \delta z_i)$ is, indeed, the conditional mean and σ is the standard deviation. The model results might appear as follows, in which the parameter on variable *mgc* is specified to have a normal distribution with a mean that is a function of *hinc*, which has a mean of about 35. The specification is

RPLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = mgc,ttme,one
; RPL = hinc ; Pts = 15 ; Maxit = 10 ; Pds = 3 ; Halton
; Fcn = mgc(n) $

```
        |Random parameters in utility functions
     MGC|     .01123         .01082      1.04  .2995     -.00999    .03245
        |Nonrandom parameters in utility functions
    TTME|    -.09941***      .01086     -9.15  .0000     -.12069   -.07813
   A_AIR|    5.98884***      .69676      8.60  .0000     4.62321   7.35447
 A_TRAIN|    4.08360***      .47295      8.63  .0000     3.15663   5.01057
   A_BUS|    3.38479***      .48263      7.01  .0000     2.43886   4.33072
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|     .00024         .00024       .99  .3241     -.00024    .00071
        |Distns. of RPs. Std.Devs or limits of triangular
   NsMGC|     .01924**       .00895      2.15  .0316      .00170    .03677
```

According to these results, the population mean of parameters on *mgc* computed at the mean income, or an estimate of $E[\beta_i|E[z_i]] \approx E_z E[\beta_i|z]]$ is roughly .01123 + 35(.00024) = .01963 and the population standard deviation is about .01924. Suppose in the same model, we change the distribution to lognormal with **; Fcn = mgc(l)**. The results change to

```
        |Random parameters in utility functions
     MGC|   -4.68371***      .81153     -5.77  .0000    -6.27428  -3.09313
        |Nonrandom parameters in utility functions
    TTME|    -.09838***      .01033     -9.52  .0000     -.11863   -.07812
   A_AIR|    5.90948***      .70945      8.33  .0000     4.51898   7.29998
 A_TRAIN|    4.03754***      .49729      8.12  .0000     3.06287   5.01221
   A_BUS|    3.32542***      .53657      6.20  .0000     2.27377   4.37707
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|     .01198         .01477       .81  .4172     -.01696    .04092
        |Distns. of RPs. Std.Devs or limits of triangular
   LsMGC|     .77048         .65552      1.18  .2398     -.51431   2.05527
```

But, the reported parameters are those of the underlying normal distribution. In this model,

$$\beta_i = \exp(\beta + \delta z_i + \sigma v_i),\ v_i \sim N[0,1].$$

The conditional (population) mean of the distribution will be

$$E[\beta_i|z_i] = \exp(\beta + \delta z_i + \sigma^2/2).$$

Inserting the estimated parameters and the mean of 35 for income, we obtain an estimate of the overall population mean of 0.01892, which is quite similar to the .01963 for the normal distribution. The variance for the lognormal is obtained as

$$\text{Var}[\beta_i|z_i] = \{E[\beta_i|z_i]\}^2 [\exp(\sigma^2) - 1].$$

Inserting our estimates and taking the square root produces an estimate of the population standard deviation of 0.017035. The result for the normal distribution is .01925. (We emphasize, we are implicitly averaging over incomes in these computations – the results are close to, but not exactly equal to the analytical results.)

The results for the lognormal distribution, correctly interpreted, are quite similar to those for the normal distribution. The structural parameters, however, are quite different. A similar characterization applies to the other distributions that are obtained as transformations of the underlying random terms. In most cases, it is not possible to obtain closed form results for the overall means and variances – the lognormal distribution is a convenient special case. The program will report its estimates of the structural parameters, but it is not generally possible to disentangle the reduced form to report the actual 'mean' and 'standard deviation' in spite of the labeling of the estimates in the program output.

Random parameter distributions that depend on the uniform distribution present another ambiguity in the interpretation of the results. For the uniform distribution, we estimate the spread of the distribution, not the standard deviation or the variance. Suppose we now change the earlier model to **; Fcn = mgc(u)**. By this construction,

$$\beta_i = \beta + \delta z_i + \sigma v_i,\ v_i \sim U[-1,1],$$

the values of β_i are distributed uniformly between $(\beta + \delta z_i - \sigma)$ and $(\beta + \delta z_i + \sigma)$. The mean is $\beta + \delta z_i$, but the variance is $4\sigma^2/12$, with a standard deviation of $\sigma/\sqrt{3}$. The estimated parameters are as follows:

```
        |Random parameters in utility functions
     MGC|     .01081         .01051      1.03  .3037     -.00979     .03142
        |Nonrandom parameters in utility functions
    TTME|    -.09888***      .01077     -9.18  .0000     -.11999    -.07776
   A_AIR|    5.95871***      .69349      8.59  .0000     4.59950    7.31792
 A_TRAIN|    4.06604***      .47177      8.62  .0000     3.14139    4.99070
   A_BUS|    3.36060***      .48065      6.99  .0000     2.41854    4.30266
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|     .00024         .00024      1.00  .3161     -.00023     .00070
        |Distns. of RPs. Std.Devs or limits of triangular
   UsMGC|     .02859*        .01476      1.94  .0529     -.00035     .05753
```

Based on these results, the overall mean is about .01081 + 35(.00024) = .01921, again comparable, and the standard deviation is .016506. What is reported is a scale factor, or spread parameter, not the standard deviation of the distribution. The standard deviation would be .02859/√3.

The triangular distribution presents the same ambiguity. In this model,

$$\beta_i = \beta + \delta z_i + \sigma v_i,\ v_i \sim \text{Triangular}[-1,1],$$

The distribution has the shape shown in Figure N29.2 in Section N29.3.8. The mean is $\beta + \delta z_i$, but the variance is $\sigma^2/6$, which is one half the variance of the uniform distribution with the same spread (and mean). Repeating the previous estimation, now with **; Fcn = mgc(t)**, we obtain the results below.

```
        |Random parameters in utility functions
     MGC|     .01083         .01061     1.02  .3077     -.00998    .03163
        |Nonrandom parameters in utility functions
    TTME|    -.09906***      .01081    -9.17  .0000     -.12024   -.07788
   A_AIR|    5.96646***      .69391     8.60  .0000     4.60642   7.32651
 A_TRAIN|    4.06893***      .47113     8.64  .0000     3.14554   4.99233
   A_BUS|    3.36673***      .48073     7.00  .0000     2.42451   4.30895
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|     .00024         .00024      .99  .3209     -.00023    .00071
        |Distns. of RPs. Std.Devs or limits of triangular
   TsMGC|     .04296**       .02159     1.99  .0466      .00064    .08529
```

Now, the mean is .01923 and the standard deviation is .04296/√6 = .17538,

The preceding serves to emphasize the need to interpret the estimated model parameters on a case by case basis. Each distribution has different characteristics. Worse yet, in some of those cases, we do not even have the convenient formulas given above to use to convert the parameters to population moments. Consider the Rayleigh distribution, which we obtain with **; Fcn = mgc(r)**. For this model,

$$\exp(\beta + \delta z_i + \sigma v_i),\ v_i = (-2\log u_i)^{\sqrt{.5}},\ u_i \sim U[0,1].$$

The estimated parameters of the model are as follows:

```
        |Random parameters in utility functions
     MGC|   -3.23112***      .94955    -3.40  .0007    -5.09220  -1.37004
        |Nonrandom parameters in utility functions
    TTME|    -.09851***      .01046    -9.42  .0000     -.11900   -.07802
   A_AIR|    5.93604***      .71733     8.28  .0000     4.53009   7.34199
 A_TRAIN|    4.05857***      .50264     8.07  .0000     3.07341   5.04373
   A_BUS|    3.34994***      .53989     6.20  .0000     2.29177   4.40811
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|     .01252         .01541      .81  .4164     -.01767    .04271
        |Distns. of RPs. Std.Devs or limits of triangular
   RsMGC|     .90592         .75815     1.19  .2321     -.58004   2.39187
```

There is no obvious way to translate these back to a mean and variance. But, there is an indirect method that is developed further in Section N29.8.

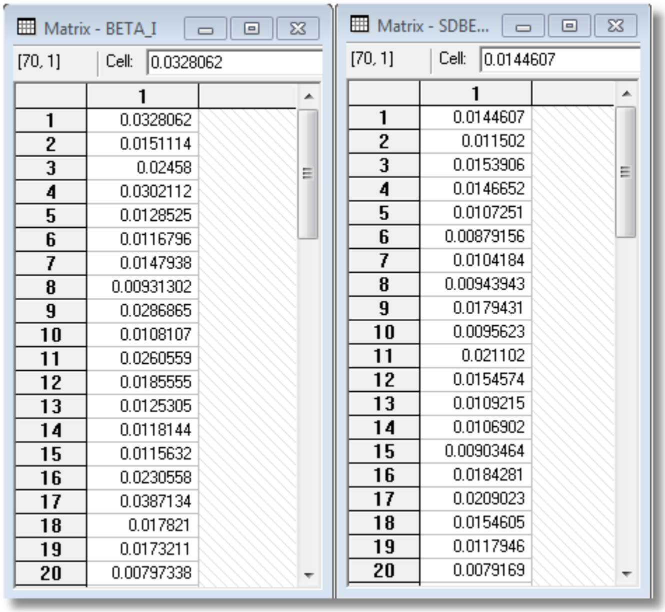
If you add

; Parameters

to your **RPLOGIT** command, then *NLOGIT* creates two matrices from the model results. The matrix *beta_i* contains for each random parameter (column) and each individual (row), an estimate of

$$\hat{\beta}_{ik} = \hat{E}[\beta_{ik} \mid \text{all information about individual } i].$$

(The method of computation is discussed in Section N29.8) The information about individual i includes their choices, so this is not quite the same as the estimator that we are using above, $E[\boldsymbol{\beta}_i|\mathbf{z}_i]$. But, since the average of conditional means gives the unconditional mean, the average of the estimates contained in *beta_i* provides an estimator of the conditional population mean that we are estimating above. A second matrix named *sdbeta_i* reports the estimated standard deviations of this distribution. Figure N29.1 below shows the first 20 rows of this 70×1 matrix as created by the model command that generated the Weibull results above.



Matrix - BETA_I [70, 1] Cell: 0.0328062

	1
1	0.0328062
2	0.0151114
3	0.02458
4	0.0302112
5	0.0128525
6	0.0116796
7	0.0147938
8	0.00931302
9	0.0286865
10	0.0108107
11	0.0260559
12	0.0185555
13	0.0125305
14	0.0118144
15	0.0115632
16	0.0230558
17	0.0387134
18	0.017821
19	0.0173211
20	0.00797338

Matrix - SDBE... [70, 1] Cell: 0.0144607

	1
1	0.0144607
2	0.011502
3	0.0153906
4	0.0146652
5	0.0107251
6	0.00879156
7	0.0104184
8	0.00943943
9	0.0179431
10	0.0095623
11	0.021102
12	0.0154574
13	0.0109215
14	0.0106902
15	0.00903464
16	0.0184281
17	0.0209023
18	0.0154605
19	0.0117946
20	0.0079169

Figure N29.1 Estimated Conditional Means and Standard Deviations

We can estimate the overall mean by averaging the elements in *beta_i*. This produces

MATRIX ; List ; ebi = 1/70*beta_i'1 $

```
     EBI|          1
--------+-------------
       1|     .0197955
```

which is the now familiar result. Estimating the population variance is a bit more complicated because the population variance is not the average of the conditional variances. Rather, the variance we seek equals the average of the conditional variances (squares of the elements in *sdbeta_i*) plus the variance of the conditional means. This is pursued in greater detail in Section N29.8.

The computation can be done (a bit inelegantly) with

MATRIX ; vi = Dirp(sdbeta_i,sdbeta_i) $
MATRIX ; evi = 1/70*vi'1 ; vei = 1/70*beta_i'beta_i - ebi*ebi $
MATRIX ; v = evi + vei ; Peek ; sd = Sqrt(v) $

```
Display of all internal digits of matrix SD
SD      [0001] =    .16969722289433440D-01
```

The result of this computation is 0.01696972. Recall, the counterpart for the normal distribution that we examined at the outset was .01924.

N29.3.3 Alternative Specific Constants

If you have used the **; Rhs = list** specification with choices specific constants, then the constants will be labeled *a_name*. For example, if you have used

; Choices = bus,train,car
; Rhs = one,cost

then to specify the model for random ASCs, you might use

; Fcn = a_bus(n),a_train(n)

As long as you are using **; Rhs = one** or **; Rh2 = one…**, you can simplify this a bit further by letting the program find the names. Use

; Fcn = choice name (n)

For example,

; Fcn = car(n), train(n), bus(n).

If you are using the **; Model:** form, then you will have supplied your own names for the ASCs.

Random choice specific constants in the random utility model with cross section data produce a random term that is a convolution of the original extreme value random variable and the one specified in your model command. Suppose, for example, that you specify a normally distributed random constant for '*car*.' Then, the utility function for *car* will be

$$U(car) = \alpha_{car} + \text{(the rest of the utility function)} + \sigma_{car} v_{car} + \varepsilon_{car}$$
$$= \alpha_{car} + \text{(the rest of the utility function)} + u_{car}.$$

The random term in this equation is the sum of a normally distributed variable and one with an extreme value distribution. This produces a different stochastic model, but probably not a useful extension of the model in general. For this reason, unless you are using panel data – see Section N29.10 – it is generally not useful to specify random constant terms in the random parameters logit model. That said, however, there is an exception which might prove useful. Random constant terms that are correlated will produce correlation across the alternatives, which is one of the oft cited virtues of the multinomial probit model. In addition, the error components logit specification produces a useful extension that serves much the same function as a random constant term.

N29.3.4 Heterogeneity in the Means of the Random Parameters

The **RPLOGIT** command requests the random parameters model generally, with the parameters specified in the **; Fcn** list varying around a mean that is the same for all individuals. The variables in $\mathbf{z}_i$ provide the variation of the mean across individuals. To specify the variables in $\mathbf{z}_i$, use

; RPL = list of variables in $\mathbf{z}_i$

If you desire to specify that $\mathbf{z}_i$ enter the means of some of the coefficients but not all, you can change the specification of the random coefficients in the **; Fcn** specification as follows:

name (type) implies $\mathbf{z}_i$ enters the mean
name [type] implies that $\mathbf{z}_i$ does not enter the mean

The difference here is the parentheses in the first as opposed to the brackets in the second. The second of these forces the applicable row of $\mathbf{\Delta}$ to contain zeros instead of free parameters. There are also some variations on this specification that allow some flexibility in the construction of $\mathbf{\Delta}$. First, an alternative, equivalent form of **name [type]** is

name (type | #)

This requests that if there are RPL variables (**; RPL = list**), these not appear in the mean for this parameter. This puts a row of zeros in the $\mathbf{\Delta}$ matrix. For example,

```
; RPL = income
; Fcn  = gc(n),ttme(n|#)
```

specifies that *income* does not appear in the mean of the *ttme* parameter. This form may be extended to exclude and include specific variables from the RPL list in the mean of a particular parameter. The specification is

name (type | # pattern)

where the pattern consists of ones and zeros which indicate which variables in the list are included (ones) and excluded (zeros). There must be the same number of items in the pattern as there are in the list. For example, the specification

```
; RPL =  age,sex,income
; Fcn  =  gc(n),
          ttme  (n|#101)
          invt  (n|#011)
          invc  (n|#000)
```

includes all three variables in the mean of *gc*, excludes *sex* from the mean of *ttme*, excludes *age* from the men of *invt*, and excludes all three variables from the mean of *invc*. All parameters may be specified independently, and there is no restriction on how this feature is used. Do note, however, if you exclude an RPL variable from all parameters, the model becomes inestimable.

N29.3.5 Fixed Coefficients

You may use

; Fix = variable [value],...

or **; Fix = name [value]**

to fix the coefficient on the specified variable at the value given in the **; Rhs = list** form and **label [value]** in the utility specification. This will override this entire specification for the indicated coefficient, in that **; Fix** specifies not only that $\mathbf{z}_i$ not enter the mean of the coefficient, but that the variance be zero as well.

N29.3.6 Correlated Parameters

The model specified thus far assumes that the random parameters are uncorrelated. Use

; Correlation

to allow free correlation among the parameters. In this case, estimates of the below diagonal elements of $\mathbf{\Gamma}$ will be obtained with the other parameters of the model. After these are presented, the elements of $\mathbf{\Sigma} = \mathbf{\Gamma\Gamma}'$ are given. An example appears below. Some ambiguity in the results will be unavoidable when this feature is used with other modifications of the model, such as mixed distributions and heteroscedasticity. The most favorable case for use of this feature would be a sparse model,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{\Gamma}\mathbf{v}_i.$$

We would note, many, perhaps most of the received applications of the mixed logit model are of this form – it is much less restrictive than its bare appearance would suggest.

In the model developed thus far, the covariance matrix for the random components for the simple distributions (normal, uniform, triangle) is

$$\text{Var}[\boldsymbol{\beta}_i|\mathbf{x}_i,\mathbf{z}_i] = \mathbf{\Sigma} = \mathbf{\Gamma\Gamma}'.$$

In the uncorrelated case, $\mathbf{\Gamma}$ is a diagonal matrix, and the variance of β_{ik} is simply σ_k^2. When the parameters are correlated, then the diagonal element of $\mathbf{\Sigma}$ is $\boldsymbol{\gamma}_k'\boldsymbol{\gamma}_k$ where $\boldsymbol{\gamma}_k$ is the kth row of $\mathbf{\Gamma}$. The model results will show the elements of $\mathbf{\Gamma}$ and the implied standard deviations. The following demonstrates the computations. The command below specifies two correlated random parameters.

```
RPLOGIT    ; Lhs  = mode ; Choices = air,train,bus,car
           ; Rhs  = gc,ttme
           ; Rh2 = one
           ; Fcn  = gc(n),ttme(n)
           ; Correlated
           ; Maxit = 50 ; Pts = 25 ; Halton ; Output = 3 ; Pds = 3 $
```

The relevant results from estimation are as follows. The coefficients reported are, first, $\boldsymbol{\beta}$ from the random parameter distributions, then the nonstochastic $\boldsymbol{\beta}$ from the distributions of the nonrandom alternative specific constants. The next results display the elements of the 2×2 lower triangular matrix, Γ. The diagonal elements appear first, then the below diagonal element(s). The 'Standard deviations of parameter distributions' are derived from $\boldsymbol{\Gamma}$. The first is $(.00973^2)^{1/2}$ = .00973. The second is $((-.07128)^2 + .03616^2)^{1/2}$ = .07993. The standard errors for these estimators are computed using the delta method. Hensher, Rose and Greene (2015) discuss the Cholesky decomposition in detail.

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                 MODE
Log likelihood function      -169.41265
Restricted log likelihood    -291.12182
Chi squared [   8 d.f.]       243.41833
Significance level               .00000
McFadden Pseudo R-squared      .4180695
Estimation based on N =    210, K =   8
Inf.Cr.AIC  =    354.8 AIC/N =    1.690
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4181 .4106
Constants only   -283.7588  .4030 .3953
At start values  -199.9766  .1528 .1419
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.02240***       .00644    -3.48  .0005     -.03502   -.00977
    TTME|    -.14423***       .02184    -6.61  .0000     -.18703   -.10143
        |Nonrandom parameters in utility functions
   A_AIR|    8.61917***      1.07974     7.98  .0000     6.50292  10.73542
 A_TRAIN|    6.87634***       .91972     7.48  .0000     5.07372   8.67896
   A_BUS|    6.03178***       .90733     6.65  .0000     4.25345   7.81012
        |Diagonal values in Cholesky matrix, L.
    NsGC|     .00973          .00762     1.28  .2019     -.00521    .02466
  NsTTME|     .03616          .03176     1.14  .2549     -.02610    .09842
        |Below diagonal values in L matrix. V = L*Lt
 TTME:GC|    -.07128***       .02311    -3.08  .0020     -.11657   -.02599
        |Standard deviations of parameter distributions
    sdGC|     .00973          .00762     1.28  .2019     -.00521    .02466
  sdTTME|     .07993***       .01792     4.46  .0000      .04480    .11506
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Correlation Matrix for Random Parameters
--------+--------------------------
Cor.Mat.|           GC          TTME
--------+--------------------------
      GC|      1.00000      -.891811
    TTME|     -.891811       1.00000
--------+--------------------------
```

We emphasize, these results apply to the linear functions of the underlying random variables, not necessarily to the implied distributions of the random parameters themselves. In most of the specifications, the parameters involve nonlinear transformations of these variables. A method of examining the results empirically is suggested in Section N29.8.

You may impose some restrictions on the correlation matrix by using

; Cor = pattern list

where the pattern list defines where zero and nonzero entries appear in Γ. The entire matrix must be specified. For example,

; Cor = 1, 1,1, 0,0,1, 0,0,0,1, 0,0,0,1,1

specifies a matrix in which parameter 3 is uncorrelated with all the others, and several other restrictions. Some cautions: A zero on the diagonal will prevent convergence. This is a somewhat volatile feature; some patterns will produce an inestimable model. This is data dependent, so it is not possible to enumerate the situations. The following uses this device to make the parameters on *gc* and *ttme* uncorrelated in this model.

```
RPLOGIT    ; Lhs  = mode ; Choices = air,train,bus,car
           ; Rhs  = gc,ttme,invc
           ; Rh2 = one
           ; Fcn  = gc(n),ttme(n),invc(n)
           ; Cor=1, 0,1, 1,1,1
           ; Maxit = 50 ; Pts = 25 ; Halton ; Output = 3 ; Pds = 3 $
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.02610**       .01026    -2.54  .0110     -.04622   -.00598
    TTME|    -.07707***      .01090    -7.07  .0000     -.09843   -.05571
    INVC|     .01304         .01099     1.19  .2354     -.00850    .03458
        |Nonrandom parameters in utility functions
   A_AIR|    5.35798***     1.18878     4.51  .0000     3.02802   7.68794
 A_TRAIN|    3.82199***      .55031     6.95  .0000     2.74340   4.90058
   A_BUS|    3.17271***      .53329     5.95  .0000     2.12748   4.21794
        |Diagonal values in Cholesky matrix, L.
    NsGC|     .01683         .01028     1.64  .1017     -.00333    .03699
  NsTTME|     .01281         .02760      .46  .6425     -.04129    .06692
  NsINVC|     .01533         .01049     1.46  .1442     -.00524    .03589
        |Below diagonal values in L matrix. V = L*Lt
 TTME:GC|        0.0    .....(Fixed Parameter).....
 INVC:GC|    -.00796         .01005     -.79  .4283     -.02766    .01174
INVC:TTM|    1.00010***      .07133    14.02  .0000      .86030   1.13990
        |Standard deviations of parameter distributions
    sdGC|     .01683         .01028     1.64  .1017     -.00333    .03699
  sdTTME|     .01281         .02760      .46  .6425     -.04129    .06692
  sdINVC|    1.00025***      .07133    14.02  .0000      .86044   1.14005
--------+--------------------------------------------------------------------
```

```
Correlation Matrix for Random Parameters
--------+---------------------------------------
Cor.Mat.|          GC          TTME          INVC
--------+---------------------------------------
      GC|     1.00000       .000000    -.00796080
    TTME|     .000000       1.00000       .999851
    INVC|  -.00796080       .999851       1.00000
--------+---------------------------------------
```

N29.3.7 Restricted Standard Deviations and Hierarchical Logit Models

The unconditional standard deviations of the random parameters (before any consideration of heteroscedasticity), σ_k are placed on the diagonal of $\mathbf{\Gamma}$ for purpose of estimation. You may restrict the diagonal elements of $\mathbf{\Gamma}$ by specifying that they be either free parameters or be fixed at specific values. The device is

; SDV = list of specifications

The list of specifications is one symbol for each random parameter, in the order in which they are given in your **; Fcn** specification. Use any alphabetic symbol for a free parameter, or the desired fixed value, including 0.0 if desired, for the fixed parameters. For example, suppose your specification were

; Fcn = gc(n),ttme(n),invt(n)

(*invt* is in vehicle time). You could specify

; SDV = 0,stt,sit

This makes the coefficient on *gc* (generalized cost) nonrandom, as its standard deviation is zero. As stated, with no other specifications, this is an ambiguous specification. The same effect could be achieved just by putting *gc* among the nonrandom parameters. But, you can use this device to create a 'hierarchical' model. Consider the specification

; Choices = air,train,bus,car
; Rhs = gc,ttme,invt
; Rh2 = one
; RPL = age,income
; Fcn = gc(n),ttme[n],invt[n]
; SDV = 0,stt,sit

This produces the model

$$\begin{aligned}
U(air) &= \alpha_{air} + (\beta + \delta_1 age + \delta_2 income) \times gc + (\beta_{ttme} + v_{ttme}) \times ttme + (\beta_{invt} + v_{invt}) \times invt + \varepsilon_a \\
U(train) &= \alpha_{train} + (\beta + \delta_1 age + \delta_2 income) \times gc + (\beta_{ttme} + v_{ttme}) \times ttme + (\beta_{invt} + v_{invt}) \times invt + \varepsilon_t \\
U(bus) &= \alpha_{bus} + (\beta + \delta_1 age + \delta_2 income) \times gc + (\beta_{ttme} + v_{ttme}) \times ttme + (\beta_{invt} + v_{invt}) \times invt + \varepsilon_b \\
U(car) &= (\beta + \delta_1 age + \delta_2 income) \times gc + (\beta_{ttme} + v_{ttme}) \times ttme + (\beta_{invt} + v_{invt}) \times invt + \varepsilon_c
\end{aligned}$$

NOTE: Using '**name(c)**' in the **; Fcn** specification is the same as setting a standard deviation to zero with **; SDV**.

You can take this a bit further and use this device to specify an entirely nonrandom, hierarchical parameter vector. The simplest way to do so is to use

```
; RPL = the list of variables
; Fcn = name(c),name(c), ...
```

This specifies that all parameters are to be nonrandom, and to have means that are functions of the variables in the RPL list. For example,

```
; Choices = air,train,bus,car
; Rhs = gc,ttme
; Rh2 = one
; RPL = age,income
; Fcn = gc(c),ttme(c)
```

This produces the model

$$U(air) = \alpha_{air} + (\beta_{gc} + \delta_{1g}age + \delta_{2g}income) \times gc + (\beta_{tt} + \delta_{1t}age + \delta_{2t}income) \times ttme + \varepsilon_a$$
$$U(train) = \alpha_{train} + (\beta_{gc} + \delta_{1g}age + \delta_{2g}income) \times gc + (\beta_{tt} + \delta_{1t}age + \delta_{2t}income) \times ttme + \varepsilon_t$$
$$U(bus) = \alpha_{bus} + (\beta_{gc} + \delta_{1g}age + \delta_{2g}income) \times gc + (\beta_{tt} + \delta_{1t}age + \delta_{2t}income) \times ttme + \varepsilon_b$$
$$U(car) = (\beta_{gc} + \delta_{1g}age + \delta_{2g}income) \times gc + (\beta_{tt} + \delta_{1t}age + \delta_{2t}income) \times ttme + \varepsilon_c$$

This is a convenient way to create interactions between attributes (such as *gc*) and characteristics (such as *age* and *income*).

This method of formulating the model can produce large numbers of parameters and produce instability in the estimator. One possibility in this event is to create interaction terms and specify them with random parameters. For example,

```
CREATE     ; gc_age = gc*age $
RPLOGIT    ; Choices = air,train,bus,car
           ; Rhs = gc,ttme,gc_age
           ; Rh2 = one
           ; RPL
           ; Fcn = gc(c),ttme(c),gc_age(n) $
```

corresponds to the model

$$U(air) = \alpha_{air} + \beta_{gc} \times gc + \beta_{tt} \times ttme + (\beta_{gc\_age} + v) \times gc \times age + \varepsilon_a$$
$$U(train) = \alpha_{train} + \beta_{gc} \times gc + \beta_{tt} \times ttme + (\beta_{gc\_age} + v) \times gc \times age + \varepsilon_t$$
$$U(bus) = \alpha_{bus} + \beta_{gc} \times gc + \beta_{tt} \times ttme + (\beta_{gc\_age} + v) \times gc \times age + \varepsilon_b$$
$$U(car) = \beta_{gc} \times gc + \beta_{tt} \times ttme + (\beta_{gc\_age} + v) \times gc \times age + \varepsilon_c$$

N29.3.8 Special Forms of Random Parameter Specifications

Several particular forms of random parameter specifications are provided for particular model aspects.

Restricting the Sign of a Parameter

There are many applications in which it is believed a priori that the sign of a coefficient must always be positive (or negative). Several of the available distributions allow you to force the sign of a coefficient to be positive. These include the following types

o	one sided triangular	$\beta_i = \beta + \beta v_i$, $v_i \sim$ triangular $(-1,1)$ $(\sigma = \beta)$
l	lognormal	$\beta_i = \exp(\beta + \sigma v_i)$, $v_i \sim N[0.1]$
x	maximum	$\beta_i = \text{Max}(0, \beta + \sigma v_i)$ $v_i \sim N[0.1]$
r	Rayleigh	$\beta_i = \exp(\beta + \sigma v_i)$, $v_i = 2(-\log u_i)^{\sqrt{.5}}$, $u_i \sim U[0,1]$
b	beta, scaled	$\beta_i = \beta v_i$, $v_i \sim$ beta(3,3)
q	exponential, scaled	$\beta_i = \beta v_i$, $v_i \sim$ exponential(1)
v	exp(triangle)	$\beta_i = \exp(\beta_i$ (triangular))

If you need to force a coefficient to be negative, rather than positive, you can use these distributions anyway – just multiply the variable by -1 before estimation. (Note, what we have labeled the 'Rayleigh' variable is not actually a Rayleigh variable, though it does resemble one. (We are using up the available symbols, however, so we have borrowed this one.) It has a shape similar to the lognormal, however, its tail is thinner, so it may be a more plausible model. Do note, however, if you specify these distributions for a coefficient which would be negative if unrestricted, the estimator will fail to converge, and issue a diagnostic that it could not locate an optimum of the function (log likelihood). Note, as well, the maximum and minimum specifications are not continuous in the parameters, and will often not be estimable.

Restricting the Range of a Parameter

Researchers often find that the infinite range of the normal distribution is unsatisfactory for the parameter in question. The fact that it allows coefficients, such as a price coefficient to take either sign is also implausible. The distributions noted above can be used to restrict the sign of a coefficient. You can also restrict the range of a coefficient. The following tighten the restrictions on the parameter distribution. Some distributions construct the range of variation to be $\beta \pm \sigma$. What we have labeled the 'dome' distribution is constructed from the beta(2,2) which has a smooth, symmetric, dome shaped distribution in (0,1). These two distributions specifically limit the range of a coefficient.

u	uniform	$\beta_i = \beta + \sigma v_i$, $v_i \sim U[-1.1]$
t	triangular	$\beta_i = \beta + \sigma v_i$, $v_i \sim$ triangle$[-1.1]$
d	dome	$\beta_i = \beta + \sigma v_i$, $v_i \sim 2$beta$(2,2) - 1$

Anchoring a Distribution at Zero

Seven alternative specifications allow you to force the entire parameter distribution to lie on one side of zero. These are

g	gamma	$\beta_i = \beta v_i$, v_i ~ gamma(1,4),
q	exponential, scaled	$\beta_i = \beta v_i$, v_i ~ exponential,
a	Rayleigh	$\beta_i = \exp(\beta + \sigma v_i)$,$v_i$~ Weibull,
b	beta, scaled	$\beta_i = \beta v_i$, v_i~ beta(2,2),
t	triangle	$\beta_i = \beta + \beta v_i$, v_i ~ triangle[-1,1],
u	uniform	$\beta_i = \beta + \beta v_i$, v_i ~ U[0,1],
l	lognormal	$\beta_i = \exp(\beta + \sigma v_i)$, v_i ~ N(0,1).

The effect is achieved in three ways in the preceding list. The lognormal variable naturally ranges from 0.0 to +∞. For the gamma, exponential-A, Weibull-A and beta cases, the estimated parameter 'mean' now acts as a scale factor against the underlying random variable, which is positive. These four specifications anchor the distribution at zero at one end. The direction of the variation is determined by β. This is not restricted. Note that no σ parameter is specified. If you use this model, σ is constrained to equal zero, and any variance heterogeneity specified is not applied to this parameter. Also, if parameters are assumed to be correlated, that feature is disabled for these parameters as well. For the gamma distribution, the mean of the underlying variable is 4, so the mean of the parameter distribution is 4β. For the beta distribution, it is β/2, while for the Rayleigh, the form we have chosen has a mean of $2\Gamma(1+0.5^{0.5}) = 2(.910005) = 1.82001$. (See http://mathworld.wolfram.com/WeibullDistribution.html.) Hence, the mean of the scaled Rayleigh distribution is β×1.82001. The exponential random variable has a mean of one, so the mean of the parameter distribution in this case is β. Note that in all four cases, we are restricting the shape of the distribution as well as the mean and variance. The first three of these are likely to be attractive alternatives to the lognormal distribution. Finally, the triangle and uniform distributions are constructed so that the spread parameter equals the mean parameter. This construction is described in the next section. The beta model is likely to be an attractive alternative to the triangle and uniform models because of the smoothness of the distribution.

Restricting the Sign and Range of a Triangular Parameter

A common device used to fix the sign of a parameter is to specify that it have a lognormal distribution. However, the lognormal distribution has a long, thick tail, which can imply an implausible empirical distribution of parameter values. An alternative is to use a random parameter with a finite range of variation. You may do this with the triangular, uniform or beta distribution, using

; Fcn = name(o) for triangular, or **; Fcn = name(f) for uniform or (h) for beta**

This specifies that the mean of the distribution is a free parameter, β, but the two endpoints of the distribution are fixed at zero and 2β, so there is no free variance (scaling) parameter. The parameter can be positive or negative. Figure N29.2 shows the result of this specification for these three distributions with β = 1.375.

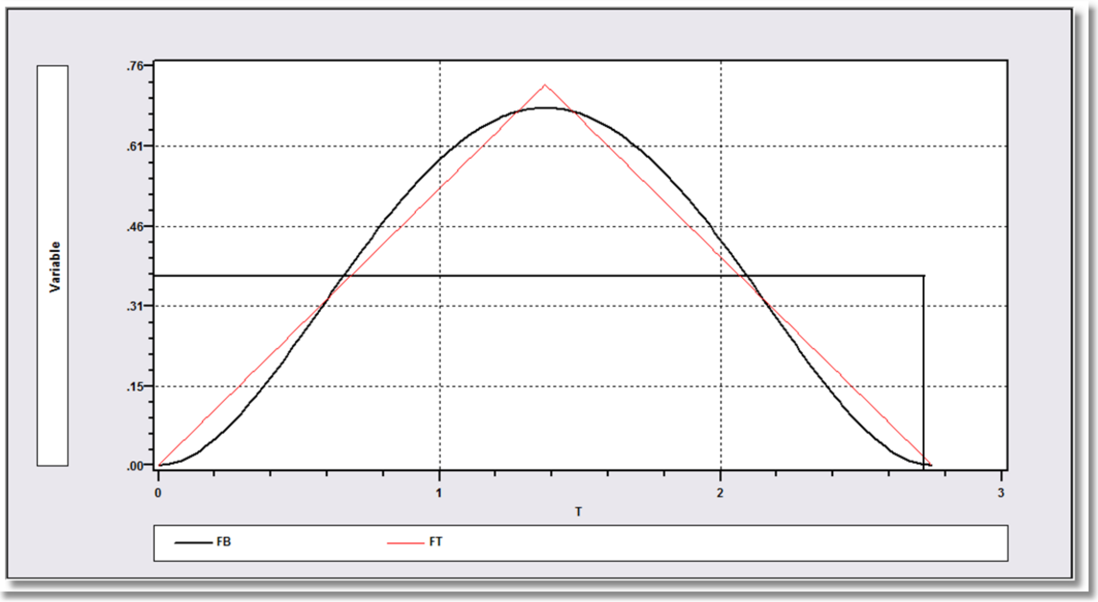


Figure N29.2 Estimated Constrained Triangular Distribution

The lognormal distribution is often used to constrain the sign of a parameter. If you use

; Fcn = name (l)

then the coefficient will be of the form exp(β+σw), which is positive. If the coefficient must be negative, for example if it is a price coefficient, then a common trick is to multiply the variable by -1 and then allow the coefficient to be positive. Alternative, you may put the sign in the command, using

; Fcn = - name(l).

Fixing the Mean of a Parameter

Use

; Fcn = name(type|value)

to fix the parameter at the specified value (with zero variance). The type is actually irrelevant, but something must be there as a placeholder. For example,

; Fcn = gc(c | -.02)

fixes the parameter at -.02. If you use this feature in a model with a heterogeneous mean, then the parameters in the heterogeneity component are fixed at zero. We do note a caution. If you attempt to fix a parameter at a value that is far from the unrestricted value, you may cause instability in the estimator. Nonsense values of parameters will produce nonsense results. The indicator that this happens will sometimes be instant convergence of the iterations at implausible estimates of the model parameters.

Fixing the Scaling Parameter

The specification

; Fcn = name (type, value)

specifies that the scaling parameter is equal to the absolute value of the mean of the distribution times the value given. The value given may equal one. For example,

; RPL = income
; Fcn = invt(n,1)

says the $\sigma_{invt} = 1 * |\beta_{invt}|$ The parameter that enters the absolute value function is the constant term in the parameter mean.

In the preceding example, we would have

$$\beta_{i,invt} = \beta + \delta income + \sigma_{invt} v_{i,invt},$$

$$\sigma_{invt} = 1 \times |\beta|.$$

(Note that when you have a heterogeneous mean, this construction becomes somewhat ambiguous. For the specification above, for example, if the uniform distribution were specified, the range of variation of the parameter, for a given value of income is from $\delta income$ to $\delta income + 2\beta$.) The uniform and triangular distributions with *value* = 1 are special cases, as this device allows you to anchor the distribution at zero for this case.

Constraining Both Mean and Scaling Parameter

The specification

; Fcn = name (type,#,value)

places a zero row in $\Delta$and constrains the corresponding σ to equal *value* * $|\beta|$. This specifies the same as **(type,value)** except in addition, if there are variables in the **;RPL = list**, these variables do not enter the mean of this parameter. This combines **(type,value)** and **(type|#)**. When specifying a fixed coefficient, you can use **name(type,#,1)**.

Fixing the Mean at a Value

The specification

; Fcn = name (type,*,value)

specifies that the mean of the parameter distribution is fixed at this value and the variance is free. This also makes sure that any **; RPL = list** variables do not enter the mean of this parameter. *This may not be used with the triangular or uniform distribution.* Note: this allows a type of 'random effects' model by fixing a parameter at zero but allowing its variance to be free. (The error components logit model of Chapter N30 and Section N29.5 is another, more direct approach for this same application.)

This specification must be used carefully. Fixing parameters in MNL models at values far from the MLEs can produce numerical instability in the estimator. The following shows a small application of this specification. This is a random effects model with two common effects, one shared by the private modes, *air* and *car*, and the other shared by the public modes, *bus* and *train*. The commands are:

```
CREATE     ; apriv = aasc + casc ; apub = tasc + basc $
RPLOGIT    ; Lhs = mode
           ; Choices = air,train,bus,car
           ; Rhs = gc,ttme,apriv,apub ; Rh2 = one
           ; Fcn = apriv(n,*,0),apub(n,*,0)
           ; Maxit = 50 ; Pts = 25
           ; Halton ; Output = 3 ; Pds = 3 $
```

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                   MODE
Log likelihood function      -196.32280
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
   APRIV|        0.0    .....(Fixed Parameter).....
    APUB|        0.0    .....(Fixed Parameter).....
        |Nonrandom parameters in utility functions
      GC|    -.01587***      .00480    -3.30  .0010     -.02528   -.00646
    TTME|    -.10009***      .01143    -8.75  .0000     -.12249   -.07768
   A_AIR|    6.00286***      .72222     8.31  .0000     4.58733   7.41840
 A_TRAIN|    4.04405***      .54052     7.48  .0000     2.98464   5.10345
   A_BUS|    3.34499***      .54667     6.12  .0000     2.27353   4.41645
        |Distns. of RPs. Std.Devs or limits of triangular
 NsAPRIV|     .17603       3.19219      .06  .9560    -6.08055   6.43261
  NsAPUB|    1.38597**       .61866     2.24  .0251      .17343   2.59852
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

N29.3.9 Other Optional Specifications

Elasticities, marginal effects, etc. are requested as usual, as are **; Prob = name**, **; Ivb = name**, and **; Utility = name** for keeping estimated probabilities, inclusive values, and estimated utilities. The inclusive value is the IV for the entire model, since this is a one level model. IVs are sometimes used for computing consumer surplus measures. Other standard output and optimization options are also used as in other models. (See Chapter N19.) The parameters used in computing the probabilities, elasticities, utilities, simulations (see Chapters N21 and N22), and so on, are the individual specific estimates described in Section N29.8. Elasticities and partial effects reported by this model account for all the aspects of the model, and include multiple effects if a variable appears in more than one place in the model.

The following options are *not* available for this model:

- Choice based sampling
- Scaling of the data and searching for a scale factor
- Nesting – this is a one level model
- Conditional probabilities – probabilities in this model are unconditional

Also, though there are several ways for you to set the starting values for the estimator, unless there is some compelling reason to do so, it is best to let the program choose its own values.

The model may be fit with ranks data. However, in order to set up that model properly, you must fit the model first without ranks data, using the first ranked choice in the choice model. (This would be a natural step in any event.)

N29.4 Heteroscedasticity and Heterogeneity in the Variances

The random parameters model allows heterogeneity in the variances as well as in the means in the distributions of the random parameters. The model is expanded to

$$\sigma_{ik} = \sigma_k \exp[\boldsymbol{\omega}_k'\mathbf{hr}_i],$$

If $\boldsymbol{\gamma}$ equals $\mathbf{0}$, this returns the homoscedastic model. The implied form of the RPL model is

$$\begin{aligned}\beta_{ik} &= \beta + \boldsymbol{\delta}_k'\mathbf{z}_i + \sigma_{ik}v_{ik}.\\ &= \beta + \boldsymbol{\delta}_k'\mathbf{z}_i + \sigma_k \exp(\boldsymbol{\omega}_k'\mathbf{hr}_i)v_{ik}.\end{aligned}$$

Request the heteroscedasticity model with

; Hfr = list of variables in hr*i*

The variables in $\mathbf{hr}_i$ may be any variables, but they must be choice invariant. Only the last value in *J* rows for choice situation *it* is used. This specification will produce the same form of heteroscedasticity in each parameter distribution – note that each parameter has its own parameter vector, $\boldsymbol{\gamma}_k$.

Section N29.3.4 describes a method of modifying the specification of the heterogeneous means of the parameters so that some RPL variables in $\mathbf{z}_i$ may appear in the means of some parameters and not others. A similar construction may be used for the variances. The general form of the specification is as follows: For any parameter specification,

; Fcn = name (type ...)

(it may contain more information beyond just the distribution type), the specification may end with an exclamation point, '**!**' to indicate that the particular parameter is to be homoscedastic even if others are heteroscedastic. For example, the following produces a model with heterogeneous means, and one heteroscedastic variance:

; RPL = age,sex
; Hfr = income
; Fcn = gc(n),ttme(n | # 01 !)

The parameter on *gc* has both heterogeneous mean and heteroscedastic variance. The parameter on *ttme* has heterogeneous mean, but *age* is excluded, and homogeneous variance. Note that there are no commas before or after the **!**. As in the case of the means, when there is more than one Hfr variable, you may add a pattern to the specification to include and exclude them from the model. To continue the previous example, consider

; RPL = age,sex
; Hfr = income,family,urban
; Fcn = gc(n),ttme(n | # 01 ! 101)

Now, the variance for *gc* includes all three variables, but the variance for *ttme* excludes *family*.

NOTE: The model with both correlated parameters (**; Correlated**) and heteroscedastic random parameters is not estimable. If your model command contains both **; Correlated** and **; Hfr = list**, the heteroscedasticity takes precedence, and the **; Correlated** is ignored.

N29.5 Random Effects and Error Components

A pure random effects form of the multinomial logit model would appear as

$$U_{ijt} = \alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \varepsilon_{ijt}.$$

This model can be specified as a mixed logit with random alternative specific constants. The full specification can be simplified with the model command

REMLOGIT ; Lhs = …; Choices = … etc.
; Pds = …
; Rh2 = one $

The effects may be assumed to be correlated across the utility functions with **; Correlated**. Other controls of the mixed logit specification, such as Halton draws and the number of draws are as usual. For example,

```
LABELS     ; modes = air,train,bus,car $
REMLOGIT   ; Lhs = mode ; Choices = modes
           ; Rhs = gc,invc,invt,ttme
           ; Pts = 50 ; Halton
           ; Pds = 3 ; Rh2 = one ; Correlated $
```

produces the following results:

```
Iterative procedure has converged
Normal exit:  37 iterations. Status=0, F=    .1485990D+03

-----------------------------------------------------------------------------
Random Effects Multinomial Logit Model
Dependent variable                    MODE
Log likelihood function         -148.59904
Restricted log likelihood       -291.12182
Chi squared [ 13](P= .000)       285.04556
Significance level                  .00000
McFadden Pseudo R-squared         .4895641
Estimation based on N =    210, K =  13
Inf.Cr.AIC  =    323.2 AIC/N =    1.539
---------------------------------------
            Log likelihood R-sqrd R2Adj
No coefficients  -291.1218  .4896 .4788
Constants only   -283.7588  .4763 .4653
At start values  -184.5067  .1946 .1776
Note: R-sqrd = 1 - logL/Logl(constants)
---------------------------------------
Response data are given as ind. choices
Replications for simulated probs. =  50
Used Halton sequences in simulations.
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random effects in utility functions..................................
   reAIR|     .54284        1.93579      .28  .7792    -3.25124   4.33693
 reTRAIN|    5.79915***      .98518     5.89  .0000     3.86823   7.73008
   reBUS|    5.63935***     1.00100     5.63  .0000     3.67743   7.60128
        |Nonrandom parameters in utility functions............................
      GC|     .20774***      .06155     3.38  .0007      .08711    .32837
    INVC|    -.23098***      .06513    -3.55  .0004     -.35863   -.10334
    INVT|    -.04007***      .00964    -4.16  .0000     -.05896   -.02119
    TTME|    -.13069***      .02107    -6.20  .0000     -.17199   -.08939
```

```
        |Diagonal values in Cholesky matrix, L...............................
    sAIR|    7.36930***     1.74081     4.23  .0000     3.95737  10.78123
  sTRAIN|    2.72350**      1.32473     2.06  .0398      .12708   5.31991
    sBUS|     .01245         .57954      .02  .9829    -1.12342   1.14832
        |Below diagonal values in L matrix. V = L*Lt.........................
reTR:reA|   -1.04145        1.05298     -.99  .3226    -3.10524   1.02235
reBU:reA|   -1.01324         .85377    -1.19  .2353    -2.68661    .66012
reBU:reT|    -.78824         .75053    -1.05  .2936    -2.25925    .68277
        |Standard deviations of parameter distributions.....................
 sdreAIR|    7.36930***     1.74081     4.23  .0000     3.95737  10.78123
sdreTRAI|    2.91583**      1.43955     2.03  .0428      .09436   5.73730
 sdreBUS|    1.28380         .99937     1.28  .1989     -.67494   3.24254
        |Covariances of Random Parameters....................................
reTR:reA|   -7.67472        7.82593     -.98  .3268   -23.01326   7.66383
reBU:reA|   -7.46689        6.10422    -1.22  .2212   -19.43095   4.49716
reBU:reT|   -1.09153        2.75254     -.40  .6917    -6.48642   4.30336
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

--------+-------------------------
Cor.Mat.|   reAIR  reTRAIN    reBUS
--------+-------------------------
   reAIR| 1.00000  -.35717  -.78925
 reTRAIN| -.35717  1.00000  -.29159
   reBUS| -.78925  -.29159  1.00000
```

In the model thus far, unobserved heterogeneity is introduced into the model through the random parameters. The probability for alternative *j* by individual *i* in choice situation *t* is

$$\text{Prob}(y_{it}=j) \;=\; \frac{\exp\left[\alpha_{ji}+\boldsymbol{\beta}_i'\mathbf{x}_{jit}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi}+\boldsymbol{\beta}_i'\mathbf{x}_{qit}\right]},$$

Chapter N30 introduces an alternative model in which the unobserved heterogeneity is brought into the model in the form of individual specific random effects that are associated with the choices, not the parameters. The probability for alternative *j* by individual *i* in choice situation *t* in that model is

$$\text{Prob}(y_{it}=j) \;=\; \frac{\exp\left[\alpha_{j}+\boldsymbol{\beta}'\mathbf{x}_{jit}+\boldsymbol{\Sigma}_{m=1}^{M}d_{jm}\theta_m E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{q}+\boldsymbol{\beta}'\mathbf{x}_{qit}+\boldsymbol{\Sigma}_{m=1}^{M}d_{qm}\theta_m E_{im}\right]}.$$

Note that the taste parameters in this model, $\boldsymbol{\beta}$, and the alternative specific constants, α_j are fixed (nonrandom). The random parameters model described in this chapter and the error components model described in Chapter N30 may be combined simply by adding the error components specification to the random parameters model already described. The new specification is

; ECM = the specification of the error components

The specification is described in detail in Section N30.2. With this specification, the random parameters model is expanded to

$$\text{Prob}(y_{it} = j) = \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{jm}\theta_m E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{qm}\theta_m E_{im}\right]},$$

Nothing in the random parameters model is changed. This feature is simply layered on top of it. All of the features of the error components model are supported as well. This includes heterogeneity in the variances (heteroscedasticity) of the error components. The model now becomes the most general form of the random parameters model,

$$\text{Prob}(y_{it} = j) = \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i) E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit} + \boldsymbol{\Sigma}_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i) E_{im}\right]},$$

This model is specified with

; Hfe = the list of variables in he$_i$

(Note, **; Hfr** specifies the heteroscedasticity in the random parameters and **; Hfe** specifies the heteroscedasticity in the random error components.) The full specification of this model appears in Section N29.3.

The following shows a small example. The model contains two correlated random parameters:

```
CREATE     ; mgc = -gc ; mttme = -ttme $
RPLOGIT    ; Lhs = mode
           ; Choices = air,train,bus,car
           ; Rhs = mgc,mttme ; Rh2 = one
           ; Fcn = mgc(n),mttme(n)
           ; Correlated
           ; ECM  = (air,car),(train,bus)
           ; Maxit = 50 ; Pts = 25
           ; Halton ; Pds  = 3 $
```

The full set of results for this model is shown below.

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function       -199.97662
Estimation based on N =    210, K =   5
Inf.Cr.AIC  =    410.0 AIC/N =    1.952
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2839
Chi-squared[ 2]            =    167.56429
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
--------+--------------------------------------------------------------------
     MGC|     .01578***      .00438     3.60  .0003      .00719    .02437
   MTTME|     .09709***      .01044     9.30  .0000      .07664    .11754
   A_AIR|    5.77636***      .65592     8.81  .0000     4.49078   7.06194
 A_TRAIN|    3.92300***      .44199     8.88  .0000     3.05671   4.78929
   A_BUS|    3.21073***      .44965     7.14  .0000     2.32943   4.09204
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   29 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                 MODE
Log likelihood function      -162.36216
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
     MGC|     .03217***      .00751     4.29  .0000      .01746    .04688
   MTTME|     .16313***      .02901     5.62  .0000      .10628    .21998
        |Nonrandom parameters in utility functions
   A_AIR|    10.2395***     1.73855     5.89  .0000      6.8320   13.6470
 A_TRAIN|    8.57301***     1.68226     5.10  .0000     5.27585  11.87018
   A_BUS|    7.56924***     1.84504     4.10  .0000     3.95303  11.18546
        |Diagonal values in Cholesky matrix, L.
   NsMGC|     .01267         .01142     1.11  .2669     -.00970    .03505
 NsMTTME| .14029D-04         .03499      .00  .9997 -.68561D-01  .68589D-01
        |Below diagonal values in L matrix. V = L*Lt
MTTM:MGC|     .08814***      .02594     3.40  .0007      .03730    .13897
        |Standard deviations of latent random effects
SigmaE01|    2.16127**       .87386     2.47  .0134      .44852   3.87401
SigmaE02|     .69870        1.37520      .51  .6114    -1.99665   3.39405
        |Standard deviations of parameter distributions
   sdMGC|     .01267         .01142     1.11  .2669     -.00970    .03505
 sdMTTME|     .08814***      .02594     3.40  .0007      .03730    .13897
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
 Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative   E01 E02
+-------------+---+---+
|  AIR        | * |   |
+-------------+---+---+
|  TRAIN      |   | * |
+-------------+---+---+
|  BUS        |   | * |
+-------------+---+---+
|  CAR        | * |   |
+-------------+---+---+

Covariance Matrix for Random Parameters
Matrix COV.MAT. has  2 rows and  2 columns.
         MGC           MTTME
        +---------------------------
MGC     |     .00016       .00114
MTTME   |     .00114       .00788
```

N29.6 Controlling the Simulations

There are two parameters of the simulations that you can change, the number of draws used in the replications and the type of sequence used to effect the integration.

N29.6.1 Number and Initiation of the Random Draws

R is the number of points (replications) in the simulation. Authors differ in the appropriate value. Generally, the more complex the model is, and the greater the number of random parameters in it, the larger will be the number of draws required to stabilize the estimates. Train (2009) recommends several hundred. Bhat (2001) suggests 1,000 is an appropriate value. The program default is 100. You can choose the value with

; Pts = number of draws, R

The RPL model is fairly time consuming to estimate. For exploratory work while you develop a final model specification, you will find that setting *R* to a small value such as 10 or 20 (as we do in the examples in this chapter) will be a useful time saver. Once a specification is finalized, a larger value will be appropriate.

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) $

We generally use **CALC ; Ran(12345) $** before each of our examples, precisely for this reason. The specific value you use for the seed is not of consequence; any odd number will do.

N29.6.2 Halton Draws and Random Draws for Simulations

The standard approach to simulation estimation is to use random draws from the specified distribution. As suggested immediately above, good performance in this connection usually requires fairly 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. A currently emerging literature has documented dramatic speed gains with no degradation in simulation performance through the use of a smaller number of Halton draws instead of a large number of random draws. Some authors (e.g., Bhat (2001) have found that a Halton sequence with a far small number of replications (as low as a tenth for a single parameter) is often as effective as a far larger number of random draws. To use this approach, add

; Halton

to your model command. Halton draws and this approach to estimation are described in the technical details in Section N29.11.3. Train et al. (2004) and others have examined a refinement of the method of Halton sequences that involves assembling the pool of draws, which are a deterministic Markov chain, and shuffling them before using them in estimation. The authors document improvements in the performance of estimators using this technique. You can use this method by changing **; Halton** to **; Shuffled** in the command. We note, this seems to speed the estimation up very slightly, but also appears to make very little difference in the estimation results.

N29.7 Model Estimates

Because of the numerous components of the model, the results for a random parameters model are somewhat more involved than for other specifications. For an example, we use the command below, which specifies a fairly involved, heterogeneous RPL model with two error components.

```
RPLOGIT    ; Lhs = mode ; Choices = air,train,bus,car
           ; Rhs = gc,ttme,one
           ; Effects: gc(air) ; RPL = hinc ; Pts = 25
           ; Maxit = 100 ; Halton ; Fcn = gc(n),ttme(n)
           ; Correlated ; ECM = (air,car),(train,bus) $
```

The initial display options for the model requested with **; Show** are the same as in other cases. The **; Describe** and **; Crosstab** are as well. These were not requested below. As usual, the estimates for the MNL model are given first. These are used as starting values for the estimates. Other parameters of the distributions of the random components are started at zeros.

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function       -199.97662
Estimation based on N =    210, K =   5
Inf.Cr.AIC  =    410.0 AIC/N =    1.952
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2816
Chi-squared[ 2]            =   167.56429
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
--------+--------------------------------------------------------------------
      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   7.06194
 A_TRAIN|    3.92300***      .44199     8.88  .0000     3.05671   4.78929
   A_BUS|    3.21073***      .44965     7.14  .0000     2.32943   4.09204
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Results from the random parameters logit model display the standard pattern, an initial box containing diagnostic statistics, followed by an indication of the size (R) and type (random or Halton) of the simulation, then the output for the model. In this model, there are likely to be many different components of the probability function, such as in the earlier example. As shown in the sample output below, the results will contain the lowest level structural parameters, first the constant terms in the random parameters in the utility functions, then the nonrandom parameters, and, finally, the parameters of the underlying distribution. The final parameters shown are the scale factors for the underlying random terms in the parameters. The leading character matches your specification in the **; Fcn** part of your command. The 's' to follow indicates this is a diagonal element of $\mathbf{\Gamma}$. Finally, up to five characters of the original name are appended.

```
-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                  MODE
Log likelihood function       -178.27968
Restricted log likelihood     -291.12182
Chi squared [  12 d.f.]        225.68428
Significance level                .00000
McFadden Pseudo R-squared       .3876114
Estimation based on N =    210, K =  12
Inf.Cr.AIC  =    380.6 AIC/N =    1.812
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3876 .3757
Constants only   -283.7588  .3717 .3595
At start values  -199.9766  .1085 .0912
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.03364         .02517    -1.34  .1813     -.08296    .01568
    TTME|    -.23249***      .08747    -2.66  .0079     -.40393   -.06105
        |Nonrandom parameters in utility functions
   A_AIR|    15.3078***     5.04275     3.04  .0024      5.4242   25.1914
 A_TRAIN|    12.8244***     4.57845     2.80  .0051      3.8508   21.7980
   A_BUS|    11.5665**      4.52366     2.56  .0106      2.7003   20.4327
        |Heterogeneity in mean, Parameter:Variable
  GC:HIN|    -.00049         .00053     -.93  .3534     -.00153    .00055
TTME:HIN|    -.00099         .00095    -1.04  .3006     -.00286    .00088
        |Diagonal values in Cholesky matrix, L.
    NsGC|     .01906         .02543      .75  .4534     -.03077    .06890
  NsTTME|     .04670         .04973      .94  .3476     -.05076    .14416
        |Below diagonal values in L matrix. V = L*Lt
 TTME:GC|     .15033**       .06722     2.24  .0253      .01859    .28208
        |Standard deviations of latent random effects
SigmaE01|    1.52524        1.42523     1.07  .2845    -1.26815   4.31863
SigmaE02|    1.66106        1.70779      .97  .3307    -1.68614   5.00826
        |Standard deviations of parameter distributions
    sdGC|     .01906         .02543      .75  .4534     -.03077    .06890
  sdTTME|     .15742**       .06301     2.50  .0125      .03392    .28092
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
 Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative   E01 E02
+-------------+---+---+
|  AIR        | * |   |
+-------------+---+---+
|  TRAIN      |   | * |
+-------------+---+---+
|  BUS        |   | * |
+-------------+---+---+
|  CAR        | * |   |
+-------------+---+---+
Parameter Matrix for Heterogeneity in Means.
--------+-------------
Delta   |         HINC
--------+-------------
      GC|  -.491237E-03
    TTME|  -.987818E-03
Correlation Matrix for Random Parameters
--------+----------------------------
Cor.Mat.|            GC          TTME
--------+----------------------------
      GC|       1.00000       .954981
    TTME|       .954981       1.00000
--------+----------------------------
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.7753    .8887    .9471    .6433
```

Note two important points about the estimated covariance matrix of the distribution of the random parameters:

- If $\boldsymbol{\Gamma}$ is diagonal, then the diagonal elements are used to scale the random elements in the parameters. However, these scale parameters are only the standard deviations of the random terms when these variables are normally distributed. Otherwise, there is some specific scale parameter that must be added to the calculation.

- If $\boldsymbol{\Gamma}$ is not diagonal, then $\boldsymbol{\Gamma}$ is not the covariance matrix of the random terms, and the diagonal elements of $\boldsymbol{\Gamma}$ are not the standard deviations even in the normal case. In this instance, $\boldsymbol{\Gamma}$is the Cholesky decomposition of the covariance matrix, which must be recovered from the estimates. The results given will include this decomposition, as shown below for this application.

Partial effects for the RPL model are computed in the same fashion as for other models, with one important exception. As in other cases, the elasticities are computed by individual, and averaged to obtain the estimate. However, in the RPL model, the individual specific estimates of the parameters described in the next section, not the population averages, are used to compute the estimates.

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.7753    .8887    .9471    .6433
```

Results saved automatically by this estimator are the same as the other estimators in *NLOGIT*, i.e.,

Matrices: *b* and *varb*

Scalars: *logl*, *kreg*, *nreg*
(Note that *nreg* is the number of individuals, not the number of rows of data in the sample.)

Last Model: See Chapter N19 for discussion of how to recover previous results.

You can also save the probabilities and utilities as follows:

; Prob = saves unconditional probabilities, based on individual parameters,
; Utility = saves values of utility functions, based on individual parameters.

This estimator will also save various matrices. These are discussed in the next section.

N29.8 Individual Specific Estimates

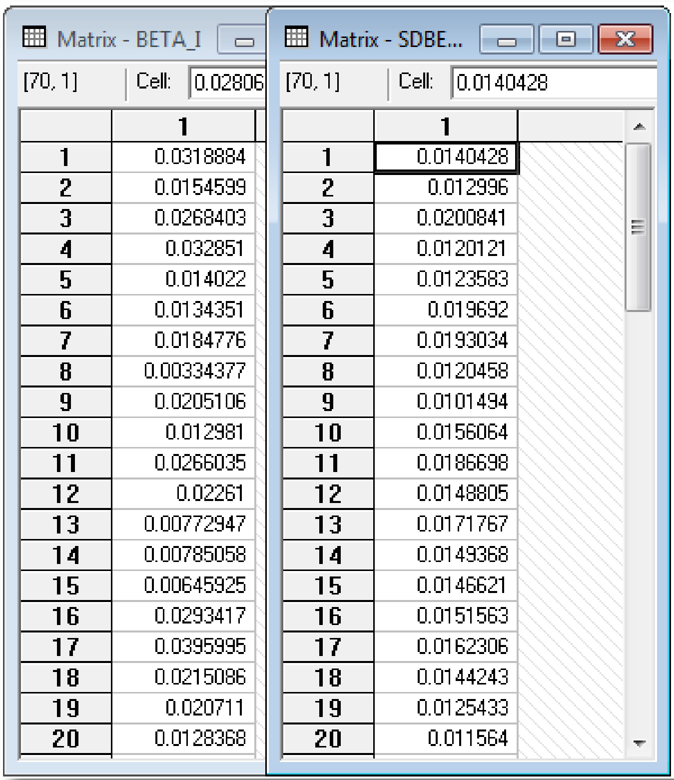
If you include

; Parameters

in your **RPLOGIT** command, *NLOGIT* will create an $n \times K$ matrix named *beta_i* that contains in a row for each individual an estimate of the random parameters in $E[\beta_i|\text{all data for individual } i]$. The model command,

```
RPLOGIT   ; Lhs = mode ; Choices = air,train,bus,car
          ; Rhs = mgc,ttme,one
          ; RPL = hinc ; Pts = 15 ; Maxit = 10 ; Pds = 3 ; Parameters
          ; Fcn = mgc(n) $
```

specifies one random parameter. The sample in use has 210/3 = 70 individuals. The matrix shown below contains the conditional estimates of the mean of the parameter on *mgc*. (The additional matrix *sdbeta_i*, is explained below.)



	BETA_I 1	SDBETA_I 1
1	0.0318884	0.0140428
2	0.0154599	0.012996
3	0.0268403	0.0200841
4	0.032851	0.0120121
5	0.014022	0.0123583
6	0.0134351	0.019692
7	0.0184776	0.0193034
8	0.00334377	0.0120458
9	0.0205106	0.0101494
10	0.012981	0.0156064
11	0.0266035	0.0186698
12	0.02261	0.0148805
13	0.00772947	0.0171767
14	0.00785058	0.0149368
15	0.00645925	0.0146621
16	0.0293417	0.0151563
17	0.0395995	0.0162306
18	0.0215086	0.0144243
19	0.020711	0.0125433
20	0.0128368	0.011564

Figure N29.3 Estimated Conditional Means and Standard Deviations

The next section will describe how these matrices are computed.

Note that **; Par** saves only the random part of the parameter vector in *beta_i*. If you wish to access the parameters in the data area, for example to compute probabilities, some manipulation will be required. The following example illustrates: Note, first, that we can expand *beta_i* into the data as a set of variables by using **; Par = namelist**. We use **; Par = brp** in the example below.

```
NAMELIST    ; (new) ; brp = brpgc,brpinvc $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car ; Pts = 10 ; Pds = 3
            ; Par = brp
            ; Fcn = gc(n),invc(n) ; Halton
            ; Rhs = gc,invc,invt,ttme $
NAMELIST    ; xp = gc,invc $
CREATE      ; urp = brpgc*gc + brpinvt*invt + b(3)*invt + b(4)*ttme $
CREATE      ; mnlrp = Mnl_probs(urp,Set=4) $
```

There are two other ways you might compute the probabilities. First, just adding **; Prob = mnlrp** to the **RPLOGIT** command would work. Second, you could access *beta_i* as follows:

```
CREATE      ; i = Trn(12,0) $
CREATE      ; urp = Mbx(beta_i,i,xp) $
CREATE      ; urp = urp + b(3)*invt + b(4)*ttme $
CREATE      ; mnlrp = mnl_probs(urp,Set=4) $
```

N29.8.1 Computing the Individual Specific Parameter Estimates

The random parameters model and the simulation based estimator used to estimate it allow the analyst to derive more information from the data than is usually available from models with fixed parameters. In particular, the model specifies that

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\boldsymbol{\Omega}_i\mathbf{v}_i,$$

where, for simplicity, if there are any, we include the alternative specific constants in $\boldsymbol{\beta}_i$, and where, if there are nonrandom parameters in the model, these are accommodated simply by having rows and columns of zeros in the appropriate places in $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}_i$. There may also be rows of zeros in $\boldsymbol{\Delta}$ for parameters that have homogeneous means. We are interested in learning as much as possible about $\boldsymbol{\beta}_i$ and functions of $\boldsymbol{\beta}_i$ from the data. The unconditional mean of $\boldsymbol{\beta}_i$ is

$$E[\boldsymbol{\beta}_i \mid \mathbf{z}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i.$$

Absent any other information, this provides the template that one would use to form their best estimate of $\boldsymbol{\beta}_i$. However, there is other information about individual i in the sample, namely the choices they made, $\mathbf{y}_i$ and other information about their heterogeneity, $\mathbf{hr}_i$. Moreover, we may also have information about individual specific error components, E_{im}, specifically in the form of $\mathbf{he}_i$, the observed heterogeneity in the variation of the error components. The following details a method of forming a conditional estimator, $E[\boldsymbol{\beta}_i|$ all data on individual $i]$.

By using Bayes Theorem, we can form the joint distribution of $\boldsymbol{\beta}_i$ and $\mathbf{y}_i = (y_{i1},y_{i2},...,y_{it})$ as follows: Denote the unconditional (marginal) distribution of $\boldsymbol{\beta}_i|\mathbf{z}_i,\mathbf{hr}_i$ as $p(\boldsymbol{\beta}_i|\mathbf{z}_i,\mathbf{hr}_i)$. This distribution is implied by whatever is assumed about $\mathbf{v}_i$ in the general model,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\boldsymbol{\Omega}_i\mathbf{v}_i$$

where, if there is heteroscedasticity, $\omega_{ik} = \sigma_k\exp[\boldsymbol{\omega}_k'\mathbf{hr}_i]$. (Elements of $\boldsymbol{\beta}_i$ might also be functions of the exponent of this expression for the lognormal and Weibull distributions.) We can also form the conditional distribution of $(\mathbf{y}_i|\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)$ based on the assumptions about $\mathbf{v}_i$ and $\mathbf{E}_i = (E_{i1},E_{i2},...,E_{iM})$ in the conditional multinomial logit model,

$$\text{Prob}(y_{it} = j_{it}, t=1,...,T_i) = \prod_{t=1}^{T_i} \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{he}_i)E_{im}\right]}.$$

(The conditional distribution is defined by the multinomial logit probabilities for the outcomes that have been assumed throughout.) We are looking ahead a bit here and treating the panel data case here rather than developing it separately later. Note as well that $\mathbf{x}_i$ denotes the collection of data on attributes and characteristics that appear in the utility functions for all the choices and in all periods or choice situations. Denote this implied conditional distribution as $p(\mathbf{y}_i|\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)$ where $\boldsymbol{\alpha}_i$ is the set of ASCs. With these in hand, we will form $p(\boldsymbol{\beta}_i|\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{hr}_i,\mathbf{he}_i,\mathbf{E}_i)$ as follows:

First, we will have to eliminate $\mathbf{E}_i$ from the conditional distribution of $\mathbf{y}_i$. The unconditional distribution is

$$p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i) = \int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,d\mathbf{E}_i\,.$$

Note that the marginal distribution is actually known – it is the M-variate standard normal distribution. Nonetheless, it will be more convenient to carry it through in generic form below. We now obtain the conditional density of $\boldsymbol{\beta}_i$ using Bayes theorem:

$$p(\boldsymbol{\beta}_i\,|\,\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i) = \frac{\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,d\mathbf{E}_i\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)}{\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,d\mathbf{E}_i}$$

$$= \frac{\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,d\mathbf{E}_i\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)}{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,d\mathbf{E}_i\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)\,d\boldsymbol{\beta}_i}.$$

Note that it is the joint density, $p(\boldsymbol{\beta}_i,\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{hr}_i,\mathbf{hf}_i)$ that appears in the fraction, the product of the conditional density times the marginal density.

Proceeding, we are interested in forming the conditional expectation, $E(\boldsymbol{\beta}_i|\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{hr}_i,\mathbf{hf}_i)$. Since the preceding gives the conditional density, the conditional expectation is formed in the usual manner,

$$E(\boldsymbol{\beta}_i\,|\,\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i) = \frac{\int_{\boldsymbol{\beta}_i}\boldsymbol{\beta}_i\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)d\mathbf{E}_i\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)d\mathbf{E}_i\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\boldsymbol{\beta}_i}$$

$$= \frac{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i}\boldsymbol{\beta}_i\,p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\mathbf{E}_i d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\mathbf{E}_i)\,p(\boldsymbol{\beta}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\mathbf{E}_i d\boldsymbol{\beta}_i}.$$

The reordering of terms to obtain the second expression is permissible because $\mathbf{E}_i$ and $\boldsymbol{\beta}_i$ are independent. Moreover, since they are independent, their joint distribution equals the product of the marginal distributions, so we may rewrite the preceding in a more useful form as

$$E(\boldsymbol{\beta}_i\,|\,\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i) = \frac{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i}\boldsymbol{\beta}_i\,p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\boldsymbol{\beta}_i,\mathbf{E}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\mathbf{E}_i d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)\,p(\boldsymbol{\beta}_i,\mathbf{E}_i\,|\,\mathbf{z}_i,\mathbf{hr}_i)d\mathbf{E}_i d\boldsymbol{\beta}_i}.$$

This would provide the basis of the conditional estimator. Note that it is precisely the form of the posterior mean if this were a Bayesian application.

The integrals in the conditional mean for $\boldsymbol{\beta}_i$ will not exist in closed form, so some other method must be used to do the integration. Note, first, that in the expression above, the term $p(\mathbf{y}_i\,|\,\boldsymbol{\beta}_i,\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_i)$ is the contribution to the conditional likelihood function (not its log) of individual i, L(**parameters** | $\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i$), and the integral is the unconditional likelihood. Second, integration over the range of $(\boldsymbol{\beta}_i,\mathbf{E}_i)$ with weighting function equal to the joint marginal density of $\boldsymbol{\beta}_i$ and $\mathbf{E}_i$ can be done by simulation. The implication is that the preceding integrals can be approximated using the simulation method used to maximize the simulated likelihood. Combining our results, we have the simulation based conditional estimator

$$\hat{E}(\boldsymbol{\beta}_i\,|\,\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\mathbf{he}_i,\mathbf{hr}_i) = \frac{\frac{1}{R}\sum_{r=1}^{R}\hat{\boldsymbol{\beta}}_{ir}\,p(\mathbf{y}_i\,|\,\hat{\boldsymbol{\beta}}_{ir},\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_{ir})}{\frac{1}{R}\sum_{r=1}^{R} p(\mathbf{y}_i\,|\,\hat{\boldsymbol{\beta}}_{ir},\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_{ir})},$$

where

$$\hat{\boldsymbol{\beta}}_{ir} = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Omega}}_i\mathbf{v}_{ir},$$

$$\hat{\boldsymbol{\Omega}}_i = diag[\exp(\hat{\boldsymbol{\omega}}_k'\mathbf{hr}_i)],$$

$$p(\mathbf{y}_i\,|\,\hat{\boldsymbol{\beta}}_{ir},\mathbf{x}_i,\mathbf{he}_i,\mathbf{E}_{ir}) = \prod_{t=1}^{T_i}\frac{\exp\left[\hat{\alpha}_{jir} + \hat{\boldsymbol{\beta}}_{ir}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\hat{\theta}_m\exp(\hat{\boldsymbol{\gamma}}_m'\mathbf{he}_i)E_{im,r}\right]}{\sum_{q=1}^{J_i}\exp\left[\hat{\alpha}_{qir} + \hat{\boldsymbol{\beta}}_{ir}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\hat{\theta}_m\exp(\hat{\boldsymbol{\gamma}}_m'\mathbf{he}_i)E_{im,r}\right]}.$$

The simulation over ($\boldsymbol{\beta}_i$,$\mathbf{E}_i$) is actually a simulation over the structural random components, $\mathbf{v}_i$ and $\mathbf{E}_i$. The preceding shows how to do the simulation once the maximum likelihood estimates of the structural parameters, [$\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma},\boldsymbol{\Omega},\boldsymbol{\theta},\boldsymbol{\gamma}$], are in hand. A final representation of the results is useful;

$$\hat{E}(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{he}_i, \mathbf{hr}_i) = \sum\nolimits_{r=1}^{R} \hat{w}_{ir}\hat{\boldsymbol{\beta}}_{ir}$$

where
$$\hat{w}_{ir} = \frac{L(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}}_{ir}, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_{ir}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})}{\Sigma_{r=1}^{R} L(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}}_{ir}, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_{ir}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})}$$

and $L(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}}_{ir}, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_{ir}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})$ is the likelihood function for individual *i* computed at the maximum simulated likelihood estimates of all the parameters, the individual's own data, and the *r*th simulated draw on ($\mathbf{v}_i$,$\mathbf{E}_i$)

The preceding shows how *NLOGIT* simulates 'estimates' of $\boldsymbol{\beta}_i$. These form the inputs for the computation of elasticities and partial effects. There is a parameter vector computed for each individual in the sample. If you include **; Parameters** in the **RPLOGIT** command, *NLOGIT* creates the matrix named *beta_i* that contains these estimates. In the preceding, any nonrandom parameter is simply identically reproduced. As such, *beta_i* contains only the conditional means for the random parameters in the model.

Whether this estimator, $\hat{E}(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{he}_i, \mathbf{hr}_i) = \sum\nolimits_{r=1}^{R} \hat{w}_{ir}\hat{\boldsymbol{\beta}}_{ir}$ is an *estimator* of $\boldsymbol{\beta}_i$ is subject to interpretation. The vector $\boldsymbol{\beta}_i$ is a draw from a distribution that has an unconditional mean,

$$E[\boldsymbol{\beta}_i|\mathbf{z}_i,\mathbf{hr}_i] = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i$$

and a conditional mean

$$E(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{he}_i, \mathbf{hr}_i) = \frac{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} \boldsymbol{\beta}_i\, p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i)\, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i)\, d\mathbf{E}_i\, d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i}\int_{\mathbf{E}_i} p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i)\, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i)\, d\mathbf{E}_i\, d\boldsymbol{\beta}_i}.$$

What we are computing here are estimates of the means of these distributions. In principle, these are conditioned on the particular data sets associated with individual *i*, not individual *i* themselves as such. To underscore the point, note that the computations would produce the same predictions for two individuals, say *i* and *i′*, if they have the same measured data, even though they would have different draws from the underlying population, ($\mathbf{v}_i$,$\mathbf{E}_i$) and ($\mathbf{v}_{i'}$,$\mathbf{E}_{i'}$). So, the mean computed here is an estimate of the center of this distribution, not a formal estimator of $\boldsymbol{\beta}_i$ as such.

We can take this a step further and examine the unconditional and conditional distributions. The variance of the unconditional distribution is

$$Var[\boldsymbol{\beta}_i|\mathbf{z}_i,\mathbf{hr}_i] = \boldsymbol{\Gamma}\boldsymbol{\Omega}_i^2\boldsymbol{\Gamma}'$$

for a particular element of $\boldsymbol{\beta}_i$, the variance is

$$Var[\beta_{ik}] = [\exp(\hat{\boldsymbol{\omega}}_k'\mathbf{hr}_i)]^2 \times \boldsymbol{\Sigma}_{s=1}^{k}\boldsymbol{\Gamma}_{sk}^2 .$$

For the conditional distribution, no such expression exists. For a particular element of β_i,

$$Var(\beta_{ik} \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{he}_i, \mathbf{hr}_i) = \frac{\int_{\boldsymbol{\beta}_i} \int_{\mathbf{E}_i} \beta_{ik}^2 \, p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i) \, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i) d\mathbf{E}_i d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i} \int_{\mathbf{E}_i} p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i) \, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i) d\mathbf{E}_i d\boldsymbol{\beta}_i} - \left[\frac{\int_{\boldsymbol{\beta}_i} \int_{\mathbf{E}_i} \beta_{ik} \, p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i) \, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i) d\mathbf{E}_i d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i} \int_{\mathbf{E}_i} p(\mathbf{y}_i \mid \boldsymbol{\beta}_i, \mathbf{x}_i, \mathbf{he}_i, \mathbf{E}_i) \, p(\boldsymbol{\beta}_i, \mathbf{E}_i \mid \mathbf{z}_i, \mathbf{hr}_i) d\mathbf{E}_i d\boldsymbol{\beta}_i} \right]^2.$$

The second term is the square of the mean that was estimated earlier. The first is the expected square, which can, like the mean, be estimated by simulation. Combining the results already obtained, then, we have an estimator of the conditional variance,

$$V\hat{a}r(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{he}_i, \mathbf{hr}_i) = \sum_{r=1}^{R} \hat{w}_{ir} (\hat{\beta}_{ir,k})^2 - \left[\sum_{r=1}^{R} \hat{w}_{ir} \hat{\beta}_{ir,k} \right]^2.$$

The square root of this quantity provides an estimate, for individual *i*, for each random parameter, an estimate of the conditional standard deviation. These diagonal elements appear in the matrix *sdbeta_i*.

We illustrate this with a model that includes most of the features described above:

```
RPLOGIT   ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,ttme ; Rh2 = one
          ; ECM = (air,car),(train,bus)
          ; RPL = hinc
          ; Fcn = gc(n),ttme(n) ; Correlated
          ; Parameters ; Halton
          ; Pds = 3 ; Pts = 200 $
```

```
-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                  MODE
Log likelihood function       -164.04264
Replications for simulated probs. = 200
Halton sequences used for simulations
RPL model with panel has        70 groups
Fixed number of obsrvs./group=         3
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
-----------------------------------------------------------------------------
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.03160         .02066    -1.53  .1263     -.07210    .00891
    TTME|    -.13631***      .02899    -4.70  .0000     -.19313   -.07950
        |Nonrandom parameters in utility functions
   A_AIR|    10.1329***     1.89857     5.34  .0000      6.4118   13.8541
 A_TRAIN|    8.19227***     1.76395     4.64  .0000     4.73498  11.64956
   A_BUS|    7.18526***     1.94752     3.69  .0002     3.36819  11.00232
        |Heterogeneity in mean, Parameter:Variable
  GC:HIN|-.41147D-05        .00047     -.01  .9930 -.92263D-03  .91440D-03
TTME:HIN|    -.00077        .00056    -1.37  .1720     -.00187    .00033
        |Diagonal values in Cholesky matrix, L.
    NsGC|     .01120        .01935      .58  .5627     -.02673    .04913
  NsTTME|     .06701        .07481      .90  .3704     -.07961    .21362
        |Below diagonal values in L matrix. V = L*Lt
 TTME:GC|    -.05562        .08696     -.64  .5224     -.22605    .11481
        |Standard deviations of latent random effects
SigmaE01|    1.40438       3.86563      .36  .7164    -6.17212   8.98089
SigmaE02|    1.72038       3.00199      .57  .5666    -4.16342   7.60418
        |Standard deviations of parameter distributions
    sdGC|     .01120        .01935      .58  .5627     -.02673    .04913
  sdTTME|     .08708***     .02846     3.06  .0022      .03130    .14287
--------+--------------------------------------------------------------------
```

```
Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative   E01 E02
+-------------+---+---+
|  AIR        | * |   |
+-------------+---+---+
|  TRAIN      |   | * |
+-------------+---+---+
|  BUS        |   | * |
+-------------+---+---+
|  CAR        | * |   |
+-------------+---+---+
Parameter Matrix for Heterogeneity in Means.
Correlation Matrix for Random Parameters
--------+----------------------------
Cor.Mat.|            GC          TTME
--------+----------------------------
      GC|       1.00000      -.638719
    TTME|      -.638719       1.00000
--------+----------------------------
```

The elements in the matrices are shown in Figure N29.4. As shown there, there is a considerable amount of variation in the estimated conditional means.

Matrix - BETA_I

[70, 2] Cell: -0.0395798

	1	2
1	-0.0395798	-0.25413
2	-0.0355292	-0.201576
3	-0.0340885	-0.200423
4	-0.039173	-0.248876
5	-0.0389149	-0.258568
6	-0.0332549	-0.161305
7	-0.0307606	-0.123531
8	-0.0179085	-0.0131606
9	-0.0236096	-0.0628686
10	-0.0286683	-0.10347
11	-0.0295245	-0.159875
12	-0.0334801	-0.18405
13	-0.0296582	-0.120611
14	-0.0210955	-0.0726241
15	-0.0281587	-0.101363
16	-0.027097	-0.108297
17	-0.0230777	-0.0564106
18	-0.0275766	-0.102563

Matrix - SDBETA_I

[70, 2] Cell: 0.0102271

	1	2
1	0.00918462	0.0486303
2	0.00887192	0.040517
3	0.00914014	0.0319871
4	0.00931433	0.0480623
5	0.00970439	0.042022
6	0.00998961	0.0672001
7	0.00965884	0.0556294
8	0.00748776	0.0413817
9	0.00992802	0.0493418
10	0.00889882	0.0438791
11	0.00943099	0.0479573
12	0.0100912	0.0689408
13	0.00969669	0.0479281
14	0.00926333	0.0494225
15	0.00974455	0.0622301
16	0.0096749	0.0568169
17	0.00993824	0.0463192
18	0.00885961	0.0317932

Figure N29.4 Conditional Means and Standard Deviations

N29.8.2 Examining the Distribution of the Parameters

As shown in Section N29.3.2 with several examples, the structural parameters often give a misleading picture of the parameters in a model. Consider the following modification of the model estimated in the previous section: We are going to fit the model as above, but change the distribution of the random parameters from normal to Weibull. The Weibull model forces parameters to be positive, so we also reverse the signs on the two attributes in the model.

```
CREATE      ; mgc = -gc ; mttme = -ttme $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme ; Rh2 = one
            ; ECM = (air,car),(train,bus)
            ; RPL = hinc
            ; Parameters ; Halton ; Pds = 3 ; Pts = 200
            ; Fcn = mgc(n),mttme(n) ; Correlated $
MATRIX      ; bn = beta_i ; sn = sdbeta_i $
```

The estimation and analysis is repeated with the Weibull distribution. Replace the last two lines with:

```
            ; Fcn = mgc(w),ttme(w) ; Correlated $
MATRIX      ; bw = beta_i ; sw = sdbeta_i $
```

The unconditional values in the first column of the matrix in Figure N29.4 and the nonstochastic estimates for the MNL model should suggest the likely values of the two random parameters. However, it would be difficult to deduce this from the estimated structural parameters for the Weibull model, which are completely different. The Weibull distribution, which involves the exponent of $\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma\Omega}_i\mathbf{v}_i$, looks quite different from the normal.

These are the basic MNL estimates, with both parameters fixed.

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
     MGC|     .01578***      .00438     3.60  .0003       .00719    .02437
   MTTME|     .09709***      .01044     9.30  .0000       .07664    .11754
   A_AIR|    5.77636***      .65592     8.81  .0000      4.49078   7.06194
 A_TRAIN|    3.92300***      .44199     8.88  .0000      3.05671   4.78929
   A_BUS|    3.21073***      .44965     7.14  .0000      2.32943   4.09204
--------+--------------------------------------------------------------------
```

This is the same model, with two correlated normally distributed random parameters with heterogeneous means. There are also two random error components in the model.

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
     MGC|     .03160         .02066     1.53  .1263      -.00891    .07210
   MTTME|     .13631***      .02899     4.70  .0000       .07950    .19313
        |Nonrandom parameters in utility functions
   A_AIR|    10.1329***     1.89857     5.34  .0000       6.4118   13.8541
 A_TRAIN|    8.19227***     1.76395     4.64  .0000      4.73498  11.64956
   A_BUS|    7.18526***     1.94752     3.69  .0002      3.36819  11.00232
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN| .41147D-05         .00047      .01  .9930 -.91440D-03  .92263D-03
MTTM:HIN|     .00077         .00056     1.37  .1720      -.00033    .00187
        |Diagonal values in Cholesky matrix, L.
   NsMGC|     .01120         .01935      .58  .5627      -.02673    .04913
 NsMTTME|     .06701         .07481      .90  .3704      -.07961    .21362
        |Below diagonal values in L matrix. V = L*Lt
MTTM:MGC|     .05562         .08696      .64  .5224      -.11481    .22605
        |Standard deviations of latent random effects
SigmaE01|    1.40438        3.86563      .36  .7164     -6.17212   8.98089
SigmaE02|    1.72038        3.00199      .57  .5666     -4.16342   7.60418
        |Standard deviations of parameter distributions
   sdMGC|     .01120         .01935      .58  .5627      -.02673    .04913
 sdMTTME|     .08708***      .02846     3.06  .0022       .03130    .14287
--------+--------------------------------------------------------------------
```

This is the same model once again, now with Weibull distributed parameters.

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
     MGC|     .01855         .04792      .39  .6987      -.07537    .11247
   MTTME|     .24966***      .09109     2.74  .0061       .07112    .42820
        |Nonrandom parameters in utility functions
   A_AIR|    10.0151***     1.72490     5.81  .0000       6.6344   13.3959
 A_TRAIN|    7.89123***     1.63492     4.83  .0000      4.68684  11.09562
   A_BUS|    6.88616***     1.80398     3.82  .0001      3.35042  10.42190
```

```
        |Heterogeneity in mean, Parameter:Variable
 MGC:HIN|-.34931D-04         .00050     -.07  .9448 -.10240D-02  .95418D-03
MTTM:HIN|     .00072         .00058     1.24  .2137     -.00042    .00186
        |Diagonal values in Cholesky matrix, L.
   WsMGC|     .00741         .02697      .27  .7835     -.04546    .06028
 WsMTTME|     .06388***      .02259     2.83  .0047      .01960    .10816
        |Below diagonal values in L matrix. V = L*Lt
MTTM:MGC|    -.00033         .04326     -.01  .9940     -.08511    .08445
        |Standard deviations of latent random effects
SigmaE01|    1.52875        7.43234      .21  .8370   -13.03837  16.09587
SigmaE02|    1.53098        7.21667      .21  .8320   -12.61344  15.67539
        |Standard deviations of parameter distributions
   sdMGC|     .00741         .02697      .27  .7835     -.04546    .06028
 sdMTTME|     .06388***      .02261     2.83  .0047      .01957    .10818
--------+-------------------------------------------------------------------
```

The ASCs in the three models resemble one another, but the coefficients on the attributes are vastly different, and would seem to suggest very different models. In fact, that is not the case, as we now examine. In order to compare these sets of estimates, we propose to examine the estimated conditional means. We will use two devices. A direct approach is to examine the distribution of estimates of $E[\boldsymbol{\beta}_i|*]$ across the observations in the sample. The averages of the conditional means will estimate the population mean (averaged across $\mathbf{z}_i$ as well). The variances require a bit of manipulation, since as noted, the variance of the conditional means underestimates the overall variance (by the mean of the conditional variances). We will also examine the distribution of conditional means in the sample with a kernel density estimator.

First estimate the models. The parameter estimates are shown above.

```
SAMPLE      ; All $
CREATE      ; mgc = -gc ; mttme = -ttme $
CLOGIT      ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme ; Rh2 = one $
CALC        ; bgmnl = b(1) ; btmnl = b(2) $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme ; Rh2 = one
            ; ECM = (air,car),(train,bus) ; RPL = hinc
            ; Parameters ; Halton ; Pds = 3 ; Pts = 200
            ; Fcn = mgc(n),mttme(n) ; Correlated $
MATRIX      ; bn = beta_i ; sn = sdbeta_i $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme ; Rh2 = one
            ; ECM = (air,car),(train,bus) ; RPL = hinc
            ; Parameters ; Halton ; Pds = 3 ; Pts = 200
            ; Fcn = mgc(w),mttme(w) ; Correlated $
MATRIX      ; bw = beta_i ; sw = sdbeta_i $
```

Now, move the matrices to the data area so we can examine them.

```
SAMPLE      ; 1-70 $
CREATE      ; bgn = 0 ; btn  = 0 ; bgw = 0 ; btw= 0 $
CREATE      ; sgn = 0 ; stn = 0 ; sgw = 0 ; stw = 0 $
NAMELIST    ; betan = bgn,btn ; betaw = bgw,btw $
NAMELIST    ; sbetan = sgn,stn ; sbetaw = sgw,stw $
CREATE      ; betan = bn $
CREATE      ; betaw = bw $
CREATE      ; sbetan = sn $
CREATE      ; sbetaw = sw $
```

Now compare the different estimates. The results below show that the normal and Weibull coefficients are much more similar than the raw parameter estimates would suggest. We first estimate the population means by averaging the conditional means.

```
CALC        ; List ; bgmnl ; Xbr(bgn) ; Xbr(bgw) $
CALC        ; List ; btmnl ; Xbr(btn) ; Xbr(btw) $
```

These are the three estimates of $E[\beta_{gc}]$

```
[CALC] BGMNL    =        .0157837
[CALC] *Result*=         .0318215  ←   (Normally distributed)
[CALC] *Result*=         .0306660      (Weibull distributed)
```

These are the three estimates of $E[\beta_{ttme}]$

```
[CALC] BTMNL    =        .0970905
[CALC] *Result*=         .1661441  ←   (Normally distributed)
[CALC] *Result*=         .1575502      (Weibull distributed)
```

Are the correlations the same? Note these are the correlations of the conditional means, not the correlations of the coefficients.

```
CALC        ; List ; Cor(bgn,btn) ; Cor(bgw,btw) $
```

```
[CALC] *Result*=         .9596877      (Two normally distributed parameters)
[CALC] *Result*=         .1786886      (Two Weibull distributed parameters)
```

The following estimate the standard deviations of the population marginal distribution of the two parameters. Once again, the similarity is striking given the quite large differences in the estimates of the structural parameters.

```
CREATE      ; vbgn = sgn^2 ; vbtn = stn^2 ; vbgw = sgw^2 ; vbtw = stw^2 $
CALC        ; List  ; sdbgn  = Sqr(xbr(vbgn) + Var(bgn))
                    ; sdbgw  = Sqr(xbr(vbgw)+ Var(bgw))
                    ; sdbtn  = Sqr(xbr(vbtn)  + Var(btn))
                    ; sdbtw  = Sqr(xbr(vbtw) + Var(btw)) $
```

```
[CALC] SDBGN    =        .0113592
[CALC] SDBGW    =        .0098213
[CALC] SDBTN    =        .0884111
[CALC] SDBTW    =        .0858662
```

A final comparison is based on the kernel density estimators for the distributions of the conditional means. Only the two for β_{gc} are shown.

```
KERNEL     ; Rhs = bgn,bgw
           ; Title = Kernel Density for E[b_gc|*,normal,Weibull]
           ; Endpoints = .01,.05 $
KERNEL     ; Rhs = btn,btw
           ; Title = Kernel Density for E[b_ttme|*,normal,Weibull] $
```

Based on the results obtained thus far, it seems that the impact of the Weibull specification is to increase the variance of the empirical distribution.

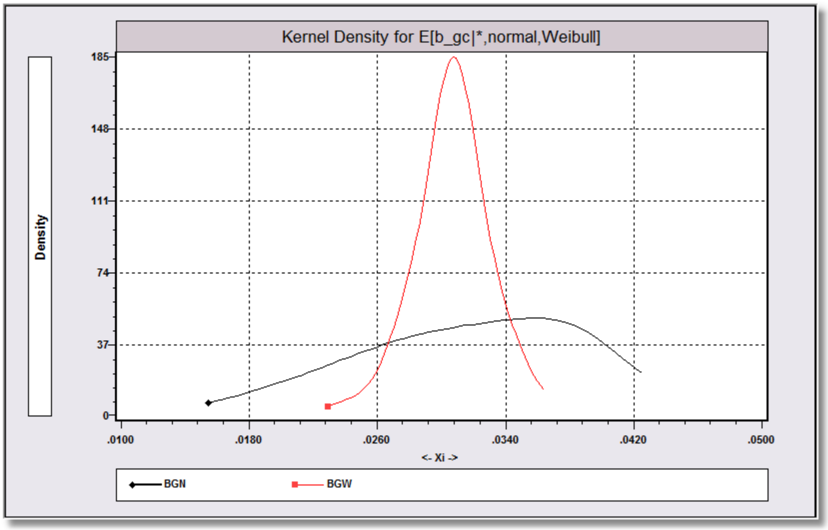


Figure N29.5 Kernel Densities for Parameter Distributions

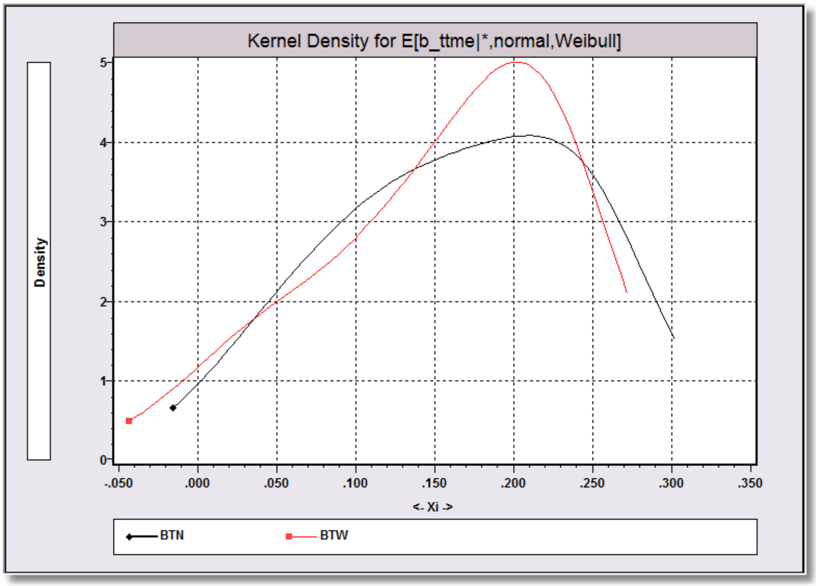


Figure N29.6 Kernel Densities for Conditional Means for βttme

N29.8.3 Conditional Confidence Intervals for Parameters

Finally, we consider an alternative approach to examining the distribution of parameters across individuals. We have for each individual, an estimate of the mean of the conditional distribution of parameters from which their specific vector is drawn. This is the estimate of $E[\boldsymbol{\beta}_i|i]$ that is in row i of *beta_i*. We also have an estimate of the standard deviation of this conditional distribution. As a general result, an interval in a distribution for a continuous random variable defined by the mean plus and minus two standard deviations will encompass 95% or more of the mass of the distribution. This enables us to form a sort of confidence interval for $\boldsymbol{\beta}_i$ itself, conditioned on all the information known about the individual. To roughly this level of confidence, the interval

$$E[\beta_{ik}|\text{all information on individual } i] \pm 2\times\text{SD}[\beta_{ik}|\text{all information on individual } i]$$

will contain the actual draw for individual i. (The probability is somewhat reduced because we are using estimates of the structural parameters, not the true values.) The centipede plot feature of **PLOT** allows us to produce this figure, as follows: We plot the figure for β_{gc} for the Weibull model:

```
CREATE    ; lowerbgc = bgw - 2*sgw ; upperbgc = bgw + 2*sgw $
CREATE    ; person = Trn(1,1) $
CALC      ; meanbgw = Xbr(bgw) $
CALC      ; highbgw = meanbgw + 2*sdbgw $
CALC      ; lowbgw = meanbgw - 2*sdbgw $
PLOT      ; Lhs = person ; Rhs = lowerbgc,upperbgc
          ; Centipede
          ; Title = Confidence Limits for b_gc for Weibull Model
          ; Bars = meanbgw,highbgw,lowbgw
          ; Endpoints = 0,75 $
```

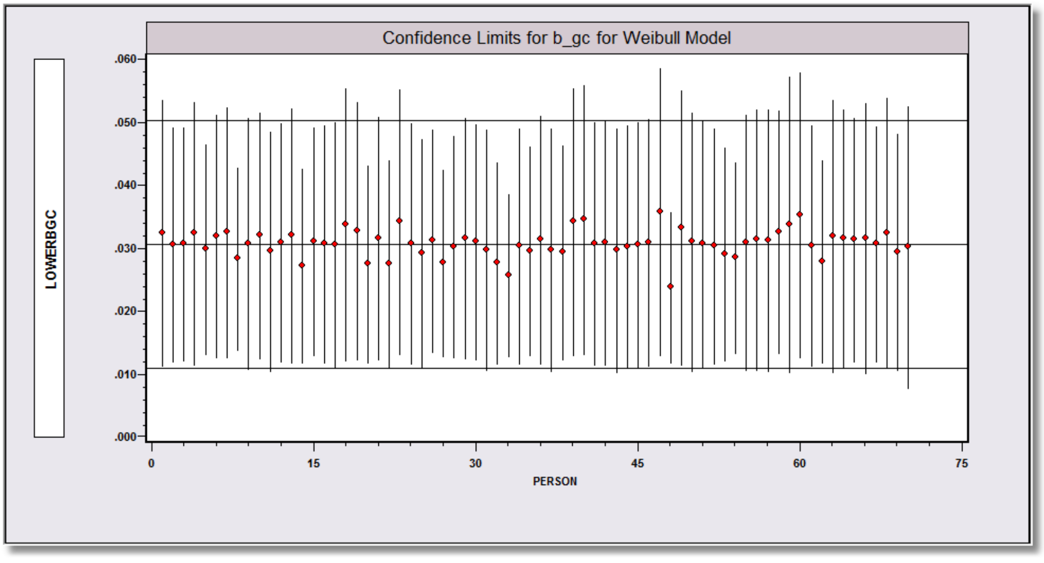


Figure N29.7 Conditional and Unconditional Distributions of Parameters

In the figure, each vertical 'leg' of the centipede plot shows the conditional confidence interval for β_{gc} for that person. The dot is the midpoint of the interval, which is the point estimate. The center horizontal bar in the figure shows the mean of the conditional means, which estimates the population mean. This was reported earlier as 0.031688. The upper and lower horizontal bars show the overall mean plus and minus twice the estimated population standard deviation – this was reported earlier as 0.009629. Thus, the unconditional population range of variation is estimated to be about .01 to .05. Note that this is the range of variation in the kernel density estimates given in Figure N29.5. Figure N29.7 demonstrates clearly how the additional information for each individual is used to reduce the 'uncertainty' about the individual specific estimates.

N29.8.4 Willingness to Pay Estimates

The previous section showed how to estimate a function of the random (or nonrandom) parameters using the simulation method. We estimated the conditional variance using a simulation based estimator of $E[\boldsymbol{\beta}_i^2|\text{all information on individual } i]$. Another useful function of the parameters in the model is the 'willingness to pay function.' This is typically measured using

WTP = attribute coefficient / income or price coefficient

The random parameters logit model will compute and retain person specific WTP measures. Use

; WTP = name/name

where names are either variable names if **; Rhs** is used or parameter names if utility functions are specified directly. In general, the WTP calculation will have an attribute level coefficient in the numerator and a cost or income measure in the denominator. Parameters can be random or nonrandom. This will create two matrices, *wtp_i* and *sdwtp_i*. These are computed the same way that *beta_i* and *sdbeta_i* are computed, where *wtp_i* contains estimates of the conditional expectation of WTP and *sdwtp_i* contains estimates of the conditional standard deviation. These matrices can be examined and analyzed in precisely the same way that *beta_i* was used earlier. You may compute more than one WTP variable by adding additional ratios in the command separated by commas. For example,

; WTP = time/income, space/price

To illustrate, we use the Weibull model once again, with a small modification:

```
SAMPLE      ; All $
RPLOGIT     ; Lhs = mode ; Choices = air,train,bus,car
            ; Rhs = mgc,mttme,hinca ; Rh2 = one
            ; ECM = (air,car),(train,bus)
            ; WTP = mttme/hinca
            ; Fcn = mgc(w),mttme(w) ; Correlated
            ; Parameters ; Halton ; Pds = 3 ; Pts = 200 $
```

The willingness to pay is computed as the ratio of the terminal time in minutes to the income variable, *hinca* – this equals income for the air alternative and zero otherwise. The basic coefficient estimates are

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
     MGC|     .04241**      .01863     2.28  .0228       .00590    .07893
   MTTME|     .24850        .22299     1.11  .2651      -.18856    .68556
        |Nonrandom parameters in utility functions
   HINCA|     .02870        .02293     1.25  .2106      -.01624    .07364
   A_AIR|    8.53653***    1.74215     4.90  .0000      5.12199  11.95108
 A_TRAIN|    7.60548***    1.54234     4.93  .0000      4.58255  10.62842
   A_BUS|    6.66168***    1.70845     3.90  .0001      3.31319  10.01017
        |Diagonal values in Cholesky matrix, L.
   WsMGC|     .00889        .00931      .95  .3396      -.00936    .02714
 WsMTTME|     .00945        .10374      .09  .9274      -.19388    .21278
        |Below diagonal values in L matrix. V = L*Lt
MTTM:MGC|    -.06409**      .02727    -2.35  .0188      -.11754   -.01063
        |Standard deviations of latent random effects
SigmaE01|     .41678       5.32188      .08  .9376    -10.01390  10.84747
SigmaE02|    1.57765       1.50521     1.05  .2946     -1.37251   4.52781
        |Standard deviations of parameter distributions
   sdMGC|     .00889        .00931      .95  .3396      -.00936    .02714
 sdMTTME|     .06478*       .03832     1.69  .0910      -.01033    .13989
--------+--------------------------------------------------------------------
```

As before, the structural parameters do not suggest what the implied parameters will look like. For these data, the estimated WTP values for the first 10 individuals (copied from *wtp_i*) are

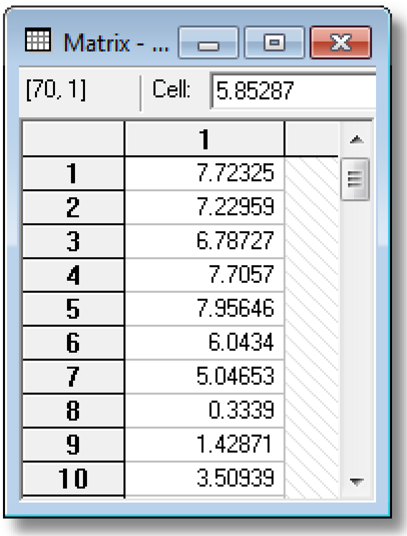


Figure N29.8 WTP Estimates

The overall average computed by averaging the 70 values in the matrix with

MATRIX ; List ; 1/70*1'wtp_i $

is 5.23934. This is in $/minute.

The WTP values are saved in the matrix *wtp_i* as shown in Figure N29.8. You may also expand the matrix into variable(s) in the data set as follows:

1. Use **CREATE** or **NAMELIST ; (New);…** to create the variable or variables if more than one.
2. Change **;WTP = definition** to **; WTP (variable or namelist) = definition**.

For example, the following will create a new variable, *timewtp* with the matrix *wtp_i*:

```
CREATE       ; timewtp $
RPLOGIT …    ; WTP (timewtp) = mttme / hinca
```

N29.9 Applications

The preceding sections and Section N29.10 contain numerous examples of the mixed logit model. The applications below show a few of the most basic procedures. This is a basic formulation with two random parameters and heterogeneity in the means as a function of household income. The observations are not grouped in this application – this is the cross section approach. We use 50 Halton draws for replicability.

```
RPLOGIT      ; Lhs = mode
             ; Choices = air,train,bus,car
             ; Rhs = gc,ttme ; Rh2 = one
             ; RPL = hinc ; Fcn = gc(n),ttme(n)
             ; Effects: gc(air)
             ; Halton ; Pts = 50 $
```

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                     MODE
Log likelihood function          -182.77116
Restricted log likelihood        -291.12182
Chi squared [   9 d.f.]           216.70131
Significance level                   .00000
McFadden Pseudo R-squared          .3721832
Estimation based on N =    210, K =    9
Inf.Cr.AIC  =    383.5 AIC/N =    1.826
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3722 .3631
Constants only   -283.7588  .3559 .3466
At start values  -199.9766  .0860 .0728
Response data are given as ind. choices
Replications for simulated probs. =  50
Halton sequences used for simulations
Number of obs.=   210, skipped    0 obs
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.01645          .01683     -.98  .3283     -.04943    .01653
    TTME|    -.17263***       .04157    -4.15  .0000     -.25409   -.09116
        |Nonrandom parameters in utility functions
   A_AIR|    10.7938***      2.02127     5.34  .0000      6.8322   14.7555
 A_TRAIN|    9.01315***      1.90238     4.74  .0000     5.28455  12.74174
   A_BUS|    8.00157***      1.83915     4.35  .0000     4.39690  11.60624
        |Heterogeneity in mean, Parameter:Variable
  GC:HIN|    -.00028          .00035     -.80  .4252     -.00097    .00041
TTME:HIN|    -.00055          .00063     -.87  .3830     -.00179    .00069
        |Distns. of RPs. Std.Devs or limits of triangular
    NsGC|     .00312          .05160      .06  .9518     -.09802    .10425
  NsTTME|     .11565***       .03706     3.12  .0018      .04303    .18828
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Parameter Matrix for Heterogeneity in Means.
--------+--------------
Delta   |          HINC
--------+--------------
      GC|  -.281194E-03
    TTME|  -.551868E-03
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |     AIR    TRAIN      BUS      CAR
--------+------------------------------------
     AIR|  -.7894    .8715   1.0384    .2573
```

This is a two level hierarchical model. There are no random parameters, but the coefficients on *gc* and *ttme* are modeled as linear functions of a constant and household income.

RPLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = gc,ttme ; Rh2 = one
; RPL = hinc ; Fcn = gc(c),ttme(c) $

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                   MODE
Log likelihood function      -198.39597
Restricted log likelihood    -291.12182
Chi squared [   7 d.f.]       185.45170
Significance level               .00000
McFadden Pseudo R-squared      .3185122
Estimation based on N =    210, K =   7
Inf.Cr.AIC  =    410.8 AIC/N =    1.956
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .3185 .3109
Constants only   -283.7588  .3008 .2930
At start values  -199.9766  .0079-.0032
Response data are given as ind. choices
Replications for simulated probs. = 500
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
```

```
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.01140         .00921    -1.24  .2158     -.02944    .00665
    TTME|    -.08786***      .01175    -7.48  .0000     -.11088   -.06484
        |Nonrandom parameters in utility functions
   A_AIR|    5.84415***      .65860     8.87  .0000     4.55331   7.13499
 A_TRAIN|    3.96546***      .44225     8.97  .0000     3.09866   4.83225
   A_BUS|    3.25638***      .45030     7.23  .0000     2.37381   4.13895
        |Heterogeneity in mean, Parameter:Variable
  GC:HIN|    -.00010         .00021     -.48  .6302     -.00051    .00031
TTME:HIN|    -.00028         .00018    -1.57  .1165     -.00063    .00007
        |Distns. of RPs. Std.Devs or limits of triangular
    CsGC|        0.0    .....(Fixed Parameter).....
  CsTTME|        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.
Parameter Matrix for Heterogeneity in Means.
--------+-------------
Delta   |         HINC
--------+-------------
      GC|  -.100937E-03
    TTME|  -.281317E-03
```

N29.10 Panel Data

The random parameters model includes a treatment for panel data. Two forms are accommodated. For a simple clustering of T_i choice situations by the same individual, for example, a stated preference survey in which several different scenarios are offered, then a random effects type of treatment might be appropriate. For example, the sequencing of choices might be unknown. In this case, the usual random effects setup would apply

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_{it} + \boldsymbol{\Gamma}\mathbf{v}_i$$

where '*t*' indexes the multiple observations for individual '*i*.' The connection to 'time' might not hold here, but we use the same index regardless. Note that the heterogeneity in the mean may change from one observation to the next (or not, depending on your situation), but the random term, $\mathbf{v}_i$ is the same for all observations. As in all panel data situations in *NLOGIT*, the number of observations, T_i on individual i may vary by individual. An alternative situation might arise when choice situations are observed in sequence, and there is a long enough lag between situations that the effect of the passage of time might be to allow preferences to evolve – consider, for example, cases in which habit persistence influences the choice (mode of travel to work), but new information enters the system. In such a case, an autoregressive arrangement might be appropriate;

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_{it} + \boldsymbol{\Gamma}\mathbf{v}_{it}$$
$$\mathbf{v}_{it} = \mathbf{R}\mathbf{v}_{i,t-1} + \mathbf{u}_{it}$$

where $\mathbf{R}$ is a diagonal matrix of autocorrelation coefficients and $\mathbf{u}_{it}$ constitutes the primitive randomness in the system.

The two situations are requested by first specifying the panel as usual with

; Pds = Ti

where T_i is either a fixed number of observations or a variable which gives the number of observations. (Note, we used this format in several of the earlier examples. See the application at the end of Section N29.8.1 for example.) In this setting, the panel consists of groups of T_i sets of J_i observations. In all cases, T_i tells the number of *groups* of data. You may have a variable number of observations *and* a variable number of choices within a group or any of the other three possible combinations. In our examples below, $J = 4$ – a fixed number of choices. In one case, $T_i = 3$, so in this case, there are 12 rows of data for each person. In the other case, there are six observations in a group, so 24 rows of data per person. If the number of observations in a group varies, so T_i is the name of a count variable, this count is repeated on every row of data within an observation, and for every observation in the group.

The autoregressive model is requested by adding

; AR1

to the **NLOGIT** command. You may also constrain the autoregressive model with

; AR1 = list of values

where the list may contain symbols for free parameters or specific numerical values, including zero if you do not wish for specific coefficients to evolve in this fashion.

To illustrate the panel data models, we will artificially treat our clogit data as if it were a panel. (It is not.) For the first model, we collect the observations in groups of three, and treat it as a random effects model. For the second, we collect the observations in groups of six, and fit an AR1 model to them.

N29.10.1 Random Effects Model

This example specifies the full parameter vector to be random, in the first form above, including the constant terms. As such, this is a true random effects model in the familiar form, that is, with a free term for each constant, in addition to the random variation in the slope parameters. The very small number of replication points was used to speed up convergence in this numerical example. Normally, you would use many more than this.

```
NLOGIT   ; Lhs = mode ; Choices = air,train,bus,car
         ; Rhs = gc,ttme ; Rh2 = one
         ; RPL= hinc
         ; Fcn = a_air(n),a_train(n),a_bus(n),gc(n),ttme(n)
         ; Correlation
         ; Parameters
         ; Pds = 3 ; Pts = 10 ; Halton $
```

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                    MODE
Log likelihood function         -121.64722
Restricted log likelihood       -291.12182
Chi squared [  25 d.f.]          338.94919
Significance level                  .00000
McFadden Pseudo R-squared         .5821432
Estimation based on N =    210, K =  25
Inf.Cr.AIC  =    293.3 AIC/N =    1.397
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .5821 .5649
Constants only   -283.7588  .5713 .5536
At start values  -199.9766  .3917 .3666
Response data are given as ind. choices
Replications for simulated probs. =  10
Halton sequences used for simulations
RPL model with panel has      70 groups
Fixed number of obsrvs./group=        3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
   A_AIR|    13.6504**      5.48059     2.49  .0128      2.9086   24.3921
 A_TRAIN|    24.5939***     6.87658     3.58  .0003     11.1161   38.0718
   A_BUS|    18.5641***     5.32865     3.48  .0005      8.1202   29.0081
      GC|    -.22146***      .07033    -3.15  .0016     -.35931   -.08361
    TTME|    -.30761***      .09578    -3.21  .0013     -.49533   -.11989
        |Heterogeneity in mean, Parameter:Variable
A_AI:HIN|      .13402        .12239     1.10  .2735     -.10585    .37390
A_TR:HIN|    -.25590**       .10410    -2.46  .0140     -.45992   -.05187
A_BU:HIN|    -.06356         .07498     -.85  .3966     -.21052    .08340
  GC:HIN|     .00432***      .00132     3.28  .0010      .00174    .00690
TTME:HIN|    -.00202         .00163    -1.24  .2158     -.00522    .00118
        |Diagonal values in Cholesky matrix, L.
 NsA_AIR|    23.8645***     7.70618     3.10  .0020      8.7607   38.9683
NsA_TRAI|    7.62594***     2.83788     2.69  .0072     2.06380  13.18807
 NsA_BUS|     .31976         .71775      .45  .6560    -1.08700   1.72652
    NsGC|     .01452         .02118      .69  .4929     -.02699    .05604
  NsTTME|     .06874***      .02413     2.85  .0044      .02144    .11603
        |Below diagonal values in L matrix. V = L*Lt
A_TR:A_A|    2.38370        2.64644      .90  .3677    -2.80322   7.57062
A_BU:A_A|   -4.83451*       2.72165    -1.78  .0757   -10.16885    .49983
A_BU:A_T|   -1.75285        1.29967    -1.35  .1774    -4.30015    .79445
  GC:A_A|    -.15494***      .04478    -3.46  .0005     -.24270   -.06717
  GC:A_T|     .10763**       .04663     2.31  .0210      .01624    .19902
  GC:A_B|     .04408**       .02081     2.12  .0341      .00330    .08486
TTME:A_A|     .22548***      .07884     2.86  .0042      .07096    .38000
TTME:A_T|    -.10454***      .03709    -2.82  .0048     -.17724   -.03184
TTME:A_B|    -.09187***      .03330    -2.76  .0058     -.15715   -.02660
 TTME:GC|    -.17369***      .05106    -3.40  .0007     -.27377   -.07362
        |Standard deviations of parameter distributions
 sdA_AIR|    23.8645***     7.70618     3.10  .0020      8.7607   38.9683
sdA_TRAI|    7.98980***     2.77044     2.88  .0039     2.55984  13.41977
 sdA_BUS|    5.15240*       2.73787     1.88  .0598     -.21372  10.51852
    sdGC|     .19428***      .02957     6.57  .0000      .13632    .25224
  sdTTME|     .32420***      .03151    10.29  .0000      .26243    .38596
```

```
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------

Parameter Matrix for Heterogeneity in Means.

--------+--------------
Delta   |          HINC
--------+--------------
   A_AIR|       .134023
 A_TRAIN|      -.255895
   A_BUS|     -.0635635
      GC|     .00432172
    TTME|    -.00201867
Correlation Matrix for Random Parameters
--------+------------------------------------------------------------------
Cor.Mat.|         A_AIR       A_TRAIN         A_BUS            GC          TTME
--------+------------------------------------------------------------------
   A_AIR|       1.00000       .298343      -.938303      -.797499       .695497
 A_TRAIN|       .298343       1.00000      -.604643       .290848      -.100279
   A_BUS|      -.938303      -.604643       1.00000       .573904      -.560473
      GC|      -.797499       .290848       .573904       1.00000      -.837658
    TTME|       .695497      -.100279      -.560473      -.837658       1.00000
```

N29.10.2 Error Components Model

The error components model presented in Section N29.5 (and Chapter N30) is also a random effects model. Without the nesting arrangement, in its simplest form, the model would be

$$\text{Prob}(y_{it} = j) \;=\; \frac{\exp\left[\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{jit} + d_j\theta_j E_{ij}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_q + \boldsymbol{\beta}'\mathbf{x}_{qit} + d_q\theta_q E_{iq}\right]}$$

where d_j equals one if the utility function for alternative j contains a random effect, and zero if not. To fit the model in this form, without random parameters, we would use the **ECLOGIT** command described in Chapter N30. The command would appear

```
ECLOGIT   ; specification of the alternatives
          ; specification of the utilities
          ; ECM = (first alt),(second alt), ...
          ; Pds = specification of the panel $
```

with one alternative in each set of parentheses. An example follows:

```
ECLOGIT   ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,ttme
          ; Rh2 = one,hinc
          ; ECM = (air),(train),(bus),(car)
          ; Pds = 3 ; Pts = 50 ; Halton $
```

```
-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                   MODE
Log likelihood function      -161.29108
Restricted log likelihood    -291.12182
Chi squared [  12 d.f.]       259.66147
Significance level               .00000
McFadden Pseudo R-squared      .4459670
Estimation based on N =    210, K =  12
Inf.Cr.AIC  =    346.6 AIC/N =    1.650
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4460 .4352
Constants only   -283.7588  .4316 .4206
At start values  -188.8499  .1459 .1293
Response data are given as ind. choices
Replications for simulated probs. =  50
Halton sequences used for simulations
ECM model with panel has      70 groups
Fixed number of obsrvs./group=        3
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters in utility functions
      GC|    -.02851***      .00881    -3.24  .0012     -.04578   -.01124
    TTME|    -.13863***      .03339    -4.15  .0000     -.20408   -.07318
   A_AIR|    7.40339***     2.58545     2.86  .0042     2.33599  12.47079
AIR_HIN1|    -.00205         .02703     -.08  .9395     -.05504    .05094
 A_TRAIN|    8.30852***     2.48448     3.34  .0008     3.43902  13.17802
TRA_HIN2|    -.09093**       .03647    -2.49  .0126     -.16240   -.01946
   A_BUS|    6.14475***     2.27164     2.70  .0068     1.69242  10.59708
BUS_HIN3|    -.03228         .03829     -.84  .3992     -.10734    .04277
        |Standard deviations of latent random effects
SigmaE01|   -4.53122***     1.39842    -3.24  .0012    -7.27208  -1.79037
SigmaE02|    3.32860***     1.14234     2.91  .0036     1.08967   5.56754
SigmaE03|     .57089        2.16106      .26  .7916    -3.66471   4.80650
SigmaE04|    1.14709        1.47766      .78  .4376    -1.74907   4.04326
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

 Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative   E01 E02 E03 E04
+-------------+---+---+---+---+
|  AIR        | * |   |   |   |
+-------------+---+---+---+---+
|  TRAIN      |   | * |   |   |
+-------------+---+---+---+---+
|  BUS        |   |   | * |   |
+-------------+---+---+---+---+
|  CAR        |   |   |   | * |
+-------------+---+---+---+---+
```

N29.10.3 Autoregression Model

The second application allows the random effect to evolve with an AR(1) process. The number of periods was increased to six for this application. Since these data are not consistent with this model at all – they are a cross section – even the larger number of 'periods' was not sufficient to produce a meaningful set of estimates. For purposes of constructing a numerical example for the display, the iterations were stopped at 10.

NLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = gc,ttme ; Rh2 = one,hinc
; RPL ; Fcn = gc(t),ttme(t)
; Correlated ; Pts = 20 ; Pds = 6
; AR1 ; Maxit = 10 ; Halton $

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                 MODE
Log likelihood function      -161.96039
Restricted log likelihood    -291.12182
Chi squared [  13 d.f.]       258.32286
Significance level               .00000
McFadden Pseudo R-squared      .4436680
Estimation based on N =    210, K =  13
Inf.Cr.AIC  =    349.9 AIC/N =    1.666
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .4437 .4319
Constants only   -283.7588  .4292 .4172
At start values  -189.5252  .1454 .1274
Response data are given as ind. choices
Replications for simulated probs. =  20
Halton sequences used for simulations
RPL model with panel has      35 groups
Fixed number of obsrvs./group=        6
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|    -.01415*        .00806    -1.76  .0790     -.02994    .00164
    TTME|    -.11237***      .03656    -3.07  .0021     -.18403   -.04071
        |Nonrandom parameters in utility functions
   A_AIR|    5.79452***     1.40263     4.13  .0000     3.04542   8.54362
AIR_HIN1|     .01081         .02924      .37  .7116     -.04649    .06811
 A_TRAIN|    6.10465***     1.26930     4.81  .0000     3.61686   8.59243
TRA_HIN2|    -.04142**       .01913    -2.17  .0303     -.07891   -.00393
   A_BUS|    4.34065***     1.49668     2.90  .0037     1.40722   7.27408
BUS_HIN3|    -.00899         .03543     -.25  .7998     -.07844    .06046
        |Diagonal values in Cholesky matrix, L.
    TsGC|     .00262         .03652      .07  .9429     -.06896    .07419
  TsTTME|     .03833         .12860      .30  .7657     -.21372    .29037
        |Below diagonal values in L matrix. V = L*Lt
 TTME:GC|    -.11219*        .06208    -1.81  .0707     -.23386    .00948
```

```
        |Autocorrelation parameters for AR(1) model
  ar[GC]|    -.00161       692.5725      .00 1.0000 -1357.41869  1357.41548
ar[TTME]|     .10571       12.12075      .01  .9930   -23.65052  23.86194
        |Standard deviations of parameter distributions
    sdGC|     .00262         .03652      .07  .9429     -.06896    .07419
  sdTTME|     .11856***      .04462     2.66  .0079      .03110    .20602
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Correlation Matrix for Random Parameters
--------+----------------------------
Cor.Mat.|            GC          TTME
--------+----------------------------
      GC|       1.00000      -.946304
    TTME|      -.946304       1.00000
--------+----------------------------
```

N29.10.4 Berry, Levinsohn and Pakes RP Logit Model

The Berry, Levinsohn and Pakes model is a random parameters logit model for market share data. The core of the specification is a multinomial logit model for the market shares, with fixed brand effects. The estimation alternates between a GMM estimator for $\boldsymbol{\beta}$ and an intermediate step to reconcile the fixed effects with the observed market share data. The multinomial logit model describes the utility of consumer *i*'s choice of brand *j* in market period *t*;

$$U_{ijt} = \mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt} + \varepsilon_{ijt},$$
$$U_{i0t} = \varepsilon_{i0t} \quad \text{for the outside good.}$$

The assumptions of the model produce the conditional (on $\boldsymbol{\beta}_i$) probability,

$$\text{Prob(consumer } i \text{ chooses brand } j \text{ in market } t) = s_j(\mathbf{X}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_i) = \frac{\exp(\mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt})}{1 + \Sigma_{m=1}^{J}\exp(\mathbf{x}'_{mt}\boldsymbol{\beta}_i + \alpha_{mt})}$$

This is a mixed logit model at this point, though it is based on market share data. Estimation of the model parameters is complicated by two factors:

1. Some attributes are endogenous due to omitted factors. In the BLP application, price is included in the model but features of the models that consumers respond to and which affect the price are not included.

2. The fixed brand effects must be estimated. The estimation procedure alternates between two steps, the 'outer' estimation, GMM conditioned on the fixed effects and the 'inner' method of moments equating market shares to theoretical market shares to calibrate the fixed effects.

The individual utility parameters β_i are distributed across individuals with CDF that is built on structural parameters $\boldsymbol{\theta}$. The predicted market share of brand *j* in market *t* will be

$$E_i[s_j(\mathbf{X}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_i)] = \int \frac{\exp(\mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt})}{1 + \Sigma_{m=1}^{J}\exp(\mathbf{x}'_{mt}\boldsymbol{\beta}_i + \alpha_{mt})} dF(\beta_i \mid \theta).$$

The estimator is built on Lee and Seo (2015) who propose a faster alternative to the original 'contraction mapping' method developed by Berry at al. Technical background on the estimation method is sketched below and can be found in Lee and Seo. (See, as well, Nevo (2000) and Greene (2015) for pedagogical material.)

Data for the estimator consist of a set of (fixed) *J* market shares for each of *T* periods (or 'markets.' The market shares for each market will be provided in a set of *J*+1 shares including the base brand. The model command is

```
BLPLOGIT    ; Lhs  = market shares variable
            ; Rhs  = attributes with nonrandom parameters
            ; Rpl  = attributes with random parameters
            ; Inst = list of instrumental variables
            ; Markets = T the number of markets $
```

An example based on simulated data is as follows:

```
SAMPLE      ; 1-1000 $
CREATE      ; market = Trn(10,0) $ (There will be 10 brands in each market) $
SETPANEL    ; Group = market ; Pds = mkt $
CREATE      ; x1 = Rnn(0,.2) ; x2 = Rnn(0,.1) ; x3 =Rnn(0,.1) $ (Attributes and price) $
CREATE      ; u = Rnn(1,.05)*x2+x1+Rnn(-.5,.3)*x3+.6 $ (Utilities with RPs) $
CREATE      ; eu = Exp(u) $
CREATE      ; iv = Group Sums(eu, Pds = mkt) $ (Inclusive value) $
CREATE      ; iv = iv+1 $ (Includes base opt-out brand) $
CREATE      ; eu = Rnu(.7,.9)*eu/iv $
CREATE      ; total = Group Sums(eu, Pds = mkt) $
CREATE      ; shares = eu $  (Random market shares) $
? Instruments are 2 exogenous variables and products of exogenous attributes.
CREATE      ; z1 = Rnu(1,1.5) ; z2 = x1 ; z3 = x2 ; z4 = x1*x2 ; z5 = x1*x1 ; z6 = x2*x2 $
CREATE      ; z7 = Rnn(0,1) ; z8 = z7*z7 $
NAMELIST    ; z = one,z* $ (All instruments) $
NAMELIST    ; x = one,x1,x2,x3 $ (All attributes) $
BLPLOGIT    ; Lhs = shares ; Rhs = one,x1 ; RPL = x2,x3
            ; Inst = z
            ; Draws = 50
            ; markets = 10 $
```

```
Iterative procedure has converged
Normal exit:   7 iterations. Status=0, F=    .1434121D-02 (Outer GMM then update)
ABLP outer iteration: 1||Change in Theta|| = .1153874D-01 (fixed effects)
Maximum of   25 iterations. Exit iterations with status=1
ABLP outer iteration: 2||Change in Theta|| = .1542259D+00 (again)
Maximum of   25 iterations. Exit iterations with status=1
ABLP outer iteration: 3||Change in Theta|| = .3972603D+00
Iterative procedure has converged
Normal exit:  13 iterations. Status=0, F=    .8732352D-01
ABLP outer iteration: 4||Change in Theta|| = .8082955D-02
Iterative procedure has converged
Normal exit:   7 iterations. Status=0, F=    .8737990D-01
ABLP outer iteration: 5||Change in Theta|| = .7679579D-03
```

```
Iterative procedure has converged
Normal exit:   7 iterations. Status=0, F=    .8738937D-01
ABLP outer iteration: 6||Change in Theta|| = .1133957D-03
Iterative procedure has converged
Normal exit:   4 iterations. Status=0, F=    .8739076D-01
ABLP outer iteration: 7||Change in Theta|| = .9635730D-06 (Last time.)
Parameter estimates have converged

-----------------------------------------------------------------------------
Random Parameters Logit (BLP)  Model
-----------------------------------------------------------------------------

--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  SHARES|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Variables with Nonrandom Parameters in Utility Function.............
Constant|   -1.18826       265.6695      .00  .9964  -521.89096  519.51444
      X1|     .92254***      .31271     2.95  .0032      .30965   1.53543
        |Variables with Random Parameters in Utility Function................
      X2|     .92450        1.01950      .91  .3645    -1.07369   2.92269
      X3|    -.14799        1.21830     -.12  .9033    -2.53581   2.23983
        |Standard Deviations of Random Parameters............................
     sX2|     .48469         .93317      .52  .6035    -1.34430   2.31368
     sX3|     .01339         .06622      .20  .8397     -.11640    .14318
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Technical Details on BLPLOGIT Estimation

BLPLOGIT uses the method developed by Lee and Seo (2015). (Further pedagogical notes may be found at Nevo (2000) and Greene (2015).) Lee and Seo's method (an alternative to BLP's contraction mapping) iterates between two steps:

1. Given the values of α_{jt}, this step estimates the random parameters logit model, $\boldsymbol{\theta}$, using GMM.

2. Given the structural parameters, $\boldsymbol{\theta}$, the values of α_{jt} are determined to equate the predicted market shares to the actual market shares. A first order Taylor series approximation that produces a Newton-like iteration is used at this step.

Convergence occurs when the values of $\boldsymbol{\theta}$ stabilize. The model is a random parameters logit model using market share data;

$$U_{ijt} = \mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt} + \varepsilon_{ijt}$$
$$U_{i0t} = \varepsilon_{i0t} \text{ (Outside good)}$$
$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{v}_i,\ \boldsymbol{\Gamma} = \text{diagonal}(\sigma_1,...)$$
$$\varepsilon_{ijt} \sim \text{Type I extreme value, IID across all choices}$$

The market shares are

$$\text{Market Shares: } s_j(\mathbf{X}_t, \alpha_t : \boldsymbol{\beta}_i) = \frac{\exp(\mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt})}{1 + \sum_{m=1}^{J} \exp(\mathbf{x}'_{mt}\boldsymbol{\beta}_i + \alpha_{mt})}, j = 1,..., J_t$$

$$\text{Expected Share: E}[s_j(\mathbf{X}_t, \alpha_t : \boldsymbol{\beta}_i)] = \int_{\boldsymbol{\beta}_i} \frac{\exp(\mathbf{x}'_{jt}\boldsymbol{\beta}_i + \alpha_{jt})}{1 + \sum_{m=1}^{J} \exp(\mathbf{x}'_{mt}\boldsymbol{\beta}_i + \alpha_{mt})} \text{f}(\boldsymbol{\beta}_i)\text{d}\boldsymbol{\beta}_i, j = 1,..., J_t$$

The expected shares are estimated using simulation:

$$\hat{s}_j(\mathbf{X}_t, \alpha_t : \boldsymbol{\beta}, \Gamma) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp(\mathbf{x}'_{jt}\boldsymbol{\beta}_{ir} + \alpha_{jt})}{1 + \sum_{m=1}^{J} \exp(\mathbf{x}'_{mt}\boldsymbol{\beta}_{ir} + \alpha_{mt})}$$

We have instruments, $\mathbf{z}_{jt}$ such that $E[\alpha_{jt}(\boldsymbol{\beta},\boldsymbol{\Gamma})\mathbf{z}_{jt}] = \mathbf{0}$; α_{jt} is obtained from an inverse mapping by equating the fitted market shares $\hat{\mathbf{s}}_t$ to the observed market shares, $\mathbf{s}_t$. Thus,

$$\hat{s}_j(\mathbf{X}_t, \alpha_t : \boldsymbol{\beta}_i) = \mathbf{s}_t, \text{ is inverted so } \hat{\boldsymbol{\alpha}}_t = \hat{\mathbf{s}}_t^{-1}(\mathbf{X}_t, \mathbf{s}_t : \boldsymbol{\beta}, \boldsymbol{\Gamma})$$

Lee provides further details on the inversion algorithm and the GMM estimator used at step 1.

N29.11 Technical Details on RP Estimation

This section will describe the procedures used in fitting the RPL model. This, with the random parameters models, constitutes what is probably the most intricate part of the estimation machinery in the software. We will present this in several parts, including formulation of the likelihood, drawing the replications for the simulations, and computing the gradients and Hessians for the optimization procedures.

N29.11.1 The Simulated Log Likelihood

We will formulate the model in the following general form: Conditioned on the unobserved latent effects, $\mathbf{v}_i$, and the other components in the model, denoted '*,' the probability for the observed outcome is

$$\text{Prob}(y_i = j \mid *, \mathbf{v}_i) = \frac{\exp(\boldsymbol{\beta}_i'\mathbf{x}_{ji})}{\sum_{m=1}^{J} \exp(\boldsymbol{\beta}_i'\mathbf{x}_{mi})},$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$$

$\mathbf{v}_i$ ~ with mean vector **0** and diagonal covariance matrix with known values on the diagonal. (It is not always **I** because we allow some distributions such as the uniform with variances that differ from one. As long as the scale is known, its precise value is immaterial. The scaling can be undone if needed when final results are reported.)

(We use the simplest possible formulation for this development. The more involved models, such as the error components models and the heteroscedastic models, are treated with the same basic procedures.) The log likelihood must be formulated in terms of observables. The unconditional probability is obtained by integrating the random terms out of the probability;

$$\text{Prob}(y_i = j|*) = \int_{\mathbf{v}_i} \text{Prob}(y_i = j \mid *, \mathbf{v}_i) g(\mathbf{v}_i) d\mathbf{v}_i .$$

As $\mathbf{v}_i$ may have many components, this is understood to be a multidimensional integral. The random variables in $\mathbf{v}_i$ are assumed to be independent, so the joint density, $g(\mathbf{v}_i)$, is the product of the individual densities. The integral will, in general, have no closed form. However, the integral is an expected value, so it can be approximated by simulation. Assuming that $\mathbf{v}_{ir}$, $r = 1,...,R$ constitutes a random sample from the underlying population $\mathbf{v}_i$, under certain conditions (see, e.g., Train (2009)), including that the function $f(\mathbf{v}_i)$ be 'smooth,' we have the property that

$$\text{plim } \frac{1}{R}\sum_{r=1}^{R} f(\mathbf{v}_{ir}) = E[f(\mathbf{v}_i)].$$

This is the fundamental result that underlies the approach to estimation used here. We will use a random number generator (or Halton draws) to produce the random samples. For each individual in the sample, the simulated unconditional probability for their observed choice is

$$\text{Prob}_S(y_i = j \mid *) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp(\boldsymbol{\beta}_{ir}' \text{'} \mathbf{x}_{ji})}{\sum_{m=1}^{J}\exp(\boldsymbol{\beta}_{ir}'\mathbf{x}_{mi})}$$

$$= \frac{1}{R}\sum_{r=1}^{R} \text{Prob}(y_i = j \mid *, \mathbf{v}_{ir}) .$$

$$\boldsymbol{\beta}_{ir} = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir}$$

$\mathbf{v}_{ir}$ ~ a random draw from the population generating $\mathbf{v}_i$.

The simulated log likelihood is then

$$\log L_S = \sum_{i=1}^{N} \log \text{Prob}_S(y_i = j \mid *) .$$

This function is then to be maximized with respect to the structural parameters, ($\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$) and, if a panel data model with autoregression is specified, ($\rho_1,...,\rho_K$). We will return to the panel data case below.

N29.11.2 Random Draws for the Simulations

The elements of $\mathbf{v}_{ir}$ are drawn as follows: We begin with a random vector $\mathbf{w}_{ir}$ which is either K independent draws from the standard uniform [0,1] distribution or K Halton draws from the mth Halton sequence, where m is the mth prime number in the sequence of K prime numbers beginning with 2. The Halton values are also distributed in the unit interval. They are described in detail below. This primitive draw is then transformed to the distribution specified in the **; Fcn** specification, as follows:

Uniform[-1,1]: $v_{k,ir} = 2w_{k,ir} - 1$

Tent [-1,1] $v_{k,ir} = \mathbf{1}(w_{k,ir} \leq .5)[\sqrt{2w_{k,ir}} - 1] +$

$\mathbf{1}(w_{k,ir} > .5)[1 - \sqrt{2(1-w_{k,ir})}]$

Normal[0,1] $v_{k,ir} = \Phi^{-1}(w_{k,ir})$

We note a consideration which is crucial in this sort of estimation. The random sequence used for the model estimation must be the same each time a probability or a function of that probability, such as a derivative, is computed 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 not, 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 *NLOGIT* 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 *NLOGIT*.

In the preceding derivation, $\mathbf{\Omega} = \mathbf{\Gamma\Gamma'}$ is the covariance matrix of $\mathbf{\Gamma v}_{ir}$ only for the standard normal case. For the other two cases, a further scaling is needed. The variance of the uniform [-1,1] is the squared width over 12, or 1/3, so its standard deviation is $1/\sqrt{3} = .57735$. The variance of the standardized tent distribution is 1/6. (Since this is a density with discontinuous derivative, this takes a bit of derivation to show.) It can be shown by partitioning the distribution. The density of u in this case is

$$f(u) = 2(1+u) \text{ for } u \leq 0 \text{ and } 2(1-u) \text{ for } u > 0.$$

The probability in each section is 1/2. The mean is obviously zero (by construction). The two conditional means are -1/3 and +1/3 for the left and right halves. The conditional variances can be found by simple integration to be 1/18 in each half. The variance equals the variance of the conditional mean plus the expected value of the conditional variance, which gives 1/9 for the former and 1/18 for the latter, which sum to 1/6. The standard deviation is therefore .40824. This implicit scaling is undone at the time the results are reported.

N29.11.3 Halton Draws for the Simulations

Conventional simulation based estimation uses a random number to produce a large number of draws from a specified distribution. The central component of the standard approach is draws from the standard continuous uniform distribution, $U[0,1]$. (*NLOGIT's* random number generator is described in Appendix R5A.3.) Draws from other distributions are obtained from these draws by using transformations. In particular, where u_i is one draw from $U[0,1]$,

Normal [0,1]: $v_i = \Phi^{-1}(u_i)$

Uniform[-1,1]: $v_i = 2u_i - 1$

Tent: $v_i = \sqrt{2u_i} - 1$ if $u_i \leq 0.5$, $v_i = 1 - \sqrt{2u_i - 1}$ otherwise.

Given that the initial draws satisfy the assumptions necessary, the central issue for purposes of specifying the simulation is the number of draws. Results differ on the number needed in a given application, but the general finding is that when simulation is done in this fashion, the number is large. A consequence of this is that for large scale problems, the amount of computation time in simulation based estimation can be extremely long.

Procedures have been devised in the numerical analysis literature for taking 'intelligent' draws from the uniform distribution, rather than random ones. (See Train (1999) and Bhat (2001) for extensive discussion and further references.) These procedures appear vastly to reduce the number of draws needed for estimation (by a factor of 90% or more) and reduce the simulation error associated with a given number of draws. In one application of the method to be discussed here, Bhat (2001) found that 100 Halton draws produced lower simulation error than 1,000 random numbers. The procedure described here is labeled Halton sequences. (See Train (1999).) The Halton sequence is generated as follows: Let r be a prime number larger than 2. Expand the sequence of integers $g = 1,...$ in terms of the base r as

$$g = \sum_{i=0}^{I} b_i r^i \text{ where by construction, } 0 \leq b_i \leq r - 1 \text{ and } r^I \leq g < r^{I+1}.$$

The Halton sequence of values that corresponds to this series is

$$H(g) = \sum_{i=0}^{I} b_i r^{-i-1}$$

For example, using base 5, the integer 37 has $b_0 = 2$, $b_1 = 2$, and $b_3 = 1$. Then

$$H(37) = 2\times5^{-1} + 2\times5^{-2} + 1\times5^{-3} = 0.448.$$

The sequence of Halton values is efficiently spread over the unit interval. The sequence is not random as the sequence of pseudo-random numbers is. The figures below show two sequences of Halton draws and two sequences of pseudorandom draws. The Halton draws are based on r = 7 and r = 9. The clumping evident in the first figure is the feature (among others) that mandates large samples for simulations.

We use the prime numbers in order beginning with 3. If a model requires K random draws, we use the first K prime numbers to generate the sequences. Within each series, the first 10 draws are discarded, as these draws tend to be highly correlated. Using Halton sequences instead of random draws can bring large savings in estimation time. Request this simply by adding **; Halton** to the **RPLOGIT** command. You will be able to reduce somewhat the number of replications when you do so.

```
SAMPLE      ; 1-1000 $
CREATE      ; h1 = Hlt(7) ; h2 = Hlt(9) ; x1 = Rnu(0,1) ; x2 = Rnu(0,1) $
PLOT        ; Lhs = h1 ; Rhs = h2 ; Limits = 0,1 ; Endpoints = 0,1
            ; Title = Plot of 1000 Draws Halton(7) vs. Halton(9) $
PLOT        ; Lhs = x1 ; Rhs = x2 ; Limits = 0,1 ; Endpoints = 0,1
            ; Title = Plot of 1000 Pairs of Pseudorandom Draws $
```

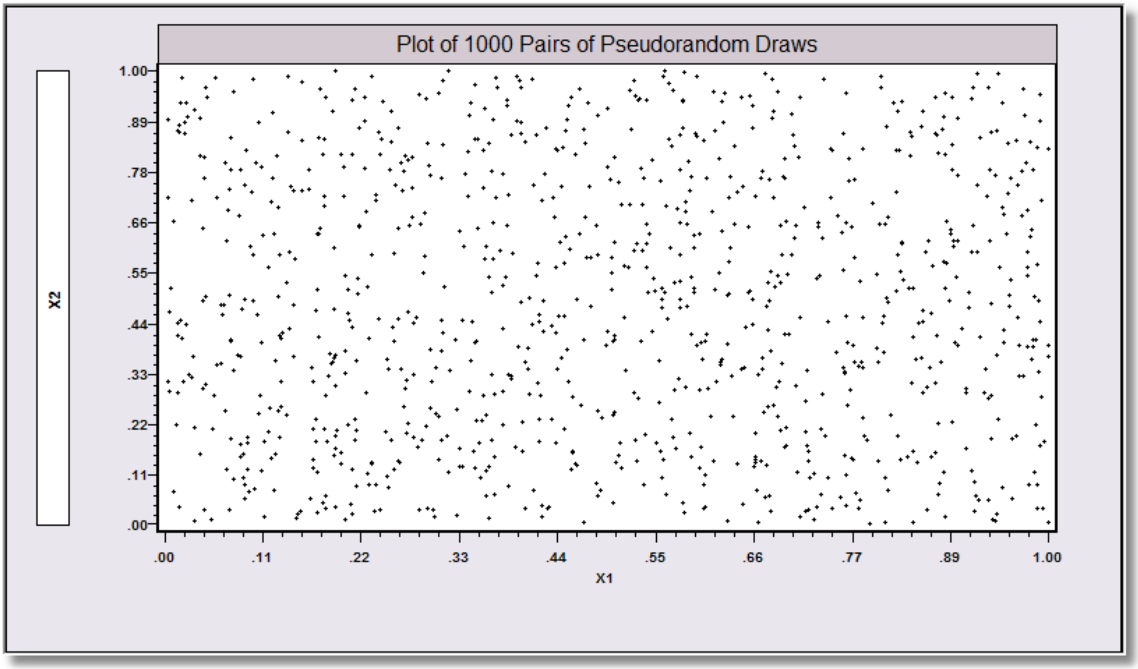


Figure N29.9 Bivariate Scatter Plot of Random Uniform Draws

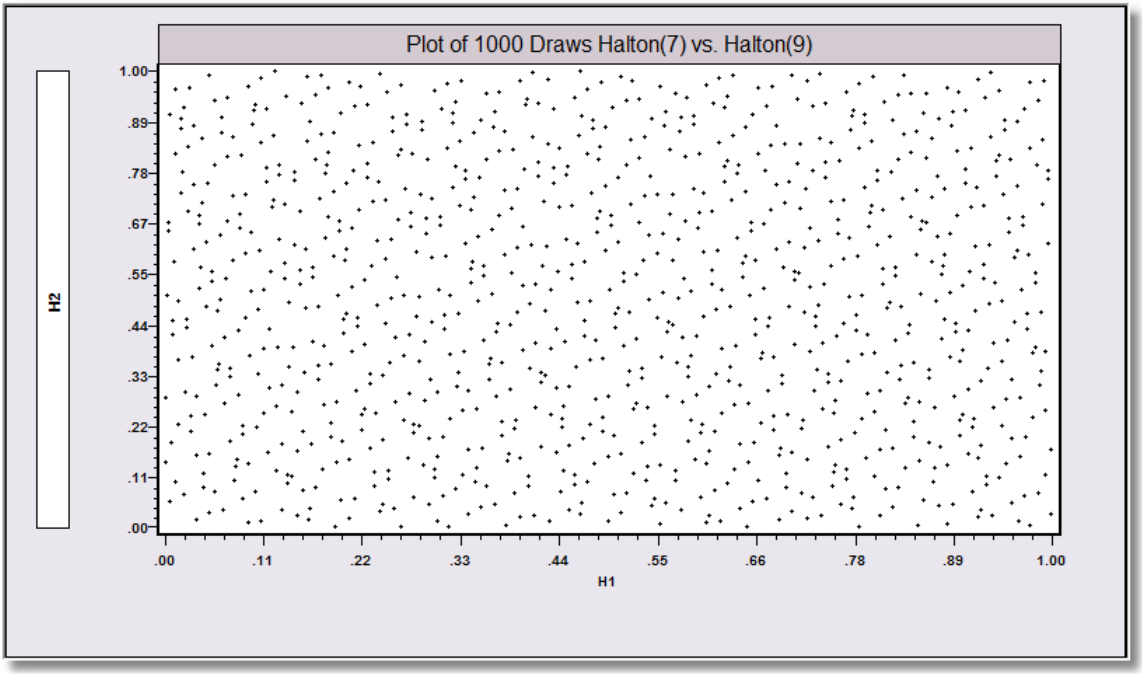


Figure N29.10 Bivariate Scatter Plot of Halton (7) and Halton (9)

N29.11.4 Functions and Gradients

We will develop the function and gradient for the basic case in which there are no error components and the variances are homoscedastic. These additional features are treated in essentially the same fashion, though they do add additional sources of complexity in the computations. We also build the results for a panel data (repeated choice situations) case. The simulated log likelihood is

$$\log L_S = \sum_{i=1}^{N} \log P_S(y_{i1} = j_1, ..., y_{iT_i} = j_{iT_i} \mid *) = \sum_{i=1}^{N} \log L_{S,i}$$

where

$$P_S(y_{i1} = j_1, ..., y_{iT_i} = j_{iT_i} \mid *) = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i} \frac{\exp(\boldsymbol{\beta}_{ir}'\mathbf{x}_{jit})}{\sum_{q=1}^{J_i}\exp(\boldsymbol{\beta}_{ir}'\mathbf{x}_{qit})}$$

$$= \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i} F_{it,r}$$

$$\boldsymbol{\beta}_{ir} = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir}$$

$\mathbf{v}_{ir} \sim$ a random draw from the population generating $\mathbf{v}_i$.

We seek

$$\mathbf{g} = \frac{\partial \log L_S}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} = \sum_{i=1}^{N} \frac{\partial \log L_{S,i}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} = \sum_{i=1}^{N} \mathbf{g}_i$$

This is obtained using

$$\begin{aligned}\mathbf{g} &= \sum_{i=1}^{N} \frac{\partial \log P_S(y_{i1}=j_1,...,y_{iT_i}=j_{iT_i}\,|\,*)}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} \\ &= \sum_{i=1}^{N} \frac{1}{P_S(y_{i1}=j_1,...,y_{iT_i}=j_{iT_i}\,|\,*)}\frac{\partial P_S(y_{i1}=j_1,...,y_{iT_i}=j_{iT_i}\,|\,*)}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} \\ &= \sum_{i=1}^{N} \frac{1}{P_S(y_{i1}=j_1,...,y_{iT_i}=j_{iT_i}\,|\,*)}\frac{1}{R}\sum_{r=1}^{R}\frac{\partial\prod_{t=1}^{T_i}F_{it,r}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})}.\end{aligned}$$

The latter term simplifies to

$$\begin{aligned}\frac{\partial\prod_{t=1}^{T_i}F_{it,r}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} &= \sum_{t=1}^{T_i}\left(\prod_{s\neq t}F_{isr}\right)\frac{\partial F_{itr}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} \\ &= \sum_{t=1}^{T_i}\left(\prod_{s=1}^{T_i}F_{is,r}\right)\frac{1}{F_{it,r}}\frac{\partial F_{it,r}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} \\ &= \left(\prod_{s=1}^{T_i}F_{isr}\right)\sum_{t=1}^{T_i}\frac{\partial\log F_{it,r}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})}\end{aligned}$$

To complete the derivation at this point, we require the innermost terms, the derivatives of logs of the multinomial logit probabilities with respect to the structural parameters. To obtain these, we use the following results: For each parameter in the vector $\boldsymbol{\beta}_{ir}$, which enters $P_S(y_{i1}=j_1,...,y_{iT_i}=j_{iT_i}\,|\,*)$, which we'll denote $\beta_{k,ir}$ we have the result that

$$\beta_{k,ir} = \beta_k + \boldsymbol{\delta}_k'\mathbf{z}_i + \boldsymbol{\Gamma}_k'\mathbf{v}_{ir}$$

where $\boldsymbol{\delta}_k$ is the kth row of $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}_k$ is the kth row of $\boldsymbol{\Gamma}$, and at this point, there is no overlap in the structural parameters that underlie different elements of $\boldsymbol{\beta}_{ir}$. If the parameters have been assumed to be uncorrelated, then $\boldsymbol{\Gamma}_k'\mathbf{v}_{ir}$ has only the diagonal term, and equals $\sigma_k v_{k,ir}$. If the parameters are correlated, then $\boldsymbol{\Gamma}$ is a lower triangular matrix, so that

$$\boldsymbol{\Gamma}_1'\mathbf{v}_{ir} = \sigma_1 v_{1,ir}$$

$$\boldsymbol{\Gamma}_2'\mathbf{v}_{ir} = \sigma_2 v_{2,ir} + \Gamma_{21} v_{1,ir}$$

and so on. The necessary derivatives can be found as follows:

$$\frac{\partial\log F_{it,r}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})} = \frac{\partial\log F_{it,r}}{\partial\beta_{k,ir}}\frac{\partial\beta_{k,ir}}{\partial(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{\Gamma})}.$$

The left (outer) part of this derivative is a familiar result in this context,

$$\frac{\partial \log F_{it,r}}{\partial \beta_{k,ir}} = x_{k,ji} - \bar{x}_{k,ir}$$

where $\bar{x}_{k,ir} = \sum_{q=1}^{J_i} F_{q,it,r} x_{k,ji}$ and $F_{q,it,r} = \dfrac{\exp(\boldsymbol{\beta}_{ir}'\mathbf{x}_{qit})}{\sum_{s=1}^{J_i}\exp(\boldsymbol{\beta}_{ir}'\mathbf{x}_{sit})}$.

The inner derivative is trivial, since $\beta_{k,ir}$ is linear in the terms of interest. Combining terms,

$$\frac{\partial \log F_{it,r}}{\partial(\beta_k, \boldsymbol{\delta}_k, \boldsymbol{\Gamma}_k)} = \left(x_{k,ji} - \bar{x}_{k,ir}\right)\begin{pmatrix}1\\ \mathbf{z}_i \\ \mathbf{v}_{k,ir}\end{pmatrix}$$

where we include the subscript k in $\mathbf{v}_{k,ir}$ to indicate that the number of elements in this vector is different for each k if the parameters are correlated, and it equals, simply, $v_{k,ir}$ when they are uncorrelated. These are then stacked for the full set of structural parameters. Collecting all terms, finally,

$$\frac{\partial \log L_S}{\partial(\beta_k, \boldsymbol{\delta}_k, \boldsymbol{\gamma}_k)} = \sum_{i=1}^{N} \frac{1}{P_S(y_{i1} = j_1,\ldots, y_{iT_i} = j_{iT_i} \mid *)}\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i} F_{it,r}\right)\sum_{t=1}^{T_i}\left(x_{k,jit} - \bar{x}_{k,it,r}\right)\begin{pmatrix}1\\ \mathbf{z}_i \\ \mathbf{v}_{k,ir}\end{pmatrix}.$$

N29.11.5 Hessians

Given the complexity of the preceding, the Hessians promise to be formidable. In fact, the results are surprisingly simple. We first write the first derivatives as

$$\mathbf{g} = \sum_{i=1}^{N} \mathbf{g}_i$$

$$= \sum_{i=1}^{N} \frac{1}{P_S(y_{i1} = j_1,\ldots, y_{iT_i} = j_{iT_i} \mid *)}\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i} F_{it,r}\right)\sum_{t=1}^{T_i} \mathbf{g}_{it,r}$$

It follows that the second derivatives matrix can be written as

$$\mathbf{H} = -\sum_{i=1}^{N} \mathbf{g}_i\mathbf{g}_i'$$

$$+ \sum_{i=1}^{N} \frac{1}{P_S(y_{i1} = j_1,\ldots, y_{iT_i} = j_{iT_i} \mid *)}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i} F_{it,r}\right)\left(\sum_{t=1}^{T_i}\mathbf{g}_{it,r}\right)\left(\sum_{t=1}^{T_i}\mathbf{g}_{it,r}\right)'$$

$$+ \sum_{i=1}^{N} \frac{1}{P_S(y_{i1} = j_1,\ldots, y_{iT_i} = j_{iT_i} \mid *)}\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i} F_{it,r}\right)\sum_{t=1}^{T_i} \mathbf{H}_{it,r}$$

where $\mathbf{H}_{it,r} = \dfrac{\partial^2 \log F_{it,r}}{\partial(\beta_k, \boldsymbol{\delta}_k, \boldsymbol{\Gamma}_k)^2}$.

The first two terms have already been derived. The last involves the second derivatives matrix of the log of the individual simulated probabilities. The notation at this point becomes excessively cumbersome. The terms in the rightmost second derivative in this expression are parts of Kronecker products involving the matrix

$$A_{k,m,ir} = \sum_{q=1}^{J_i} F_{it,q,t}\left(x_{k,qit} - \bar{x}_{k,i,r}\right)\left(x_{m,ji} - \bar{x}_{m,i,r}\right)$$

(again, a familiar result in the logit model) and a second matrix

$$\mathbf{B}_{km,ir} = \begin{pmatrix} 1 \\ \mathbf{z}_i \\ \mathbf{v}_{k,ir} \end{pmatrix}\begin{pmatrix} 1 \\ \mathbf{z}_i \\ \mathbf{v}_{m,ir} \end{pmatrix}'$$

The first and third terms in the Hessian are negative definite ($\mathbf{H}_{it,r}$ is negative semidefinite) and the second is positive definite. In a finite sample, the sum of the three need not be negative definite, which means that in a finite sample, the estimated asymptotic covariance based on the second derivatives might not be positive definite. However, in theory, the second and third terms above should sum to zero (at least in large samples). Therefore, the BHHH estimator in the first line is a valid estimator of the asymptotic covariance matrix for the maximum likelihood estimators of the parameters in this model. We use this estimator when the full Hessian turns out not to be negative definite. The results for the model will sometimes contain an indication of this condition. This does not indicate that something has gone wrong – this is a finite sample result that can be ignored (assuming that estimation was otherwise successful).

N29.11.6 Panel Data and Autocorrelation

The preceding results must be modified slightly for the panel data estimators. The draws are obtained as follows: The primitive draws are taken as usual, now denoted $\mathbf{u}_{itr}$ to indicate the role of time explicitly in the equations. In the standard, random effects case examined in the previous section, $\mathbf{v}_{itr} = \mathbf{u}_{irt}$ for all t and $\mathbf{u}_{irt}$ is drawn once for the entire set of T_i periods. The effect is the same in every period. The various terms in the log likelihood and its derivatives are be obtained by summation over periods within the summation over observations, and nothing else need be changed. The autoregression model is rather more involved. The random variable generation process is

$$v_{k,ir1} = \left(1/\sqrt{1-\rho_k^2}\right)u_{k,ir1}$$

$$v_{k,irt} = \rho_k\, v_{k,i,t-1,r} + u_{k,irt}$$

This is the standard first order autocorrelation treatment, with the Prais-Winsten treatment for the first observation – this is done to avoid losing any observations due to differencing.

Generation of the probabilities and the log likelihood are straightforward, given the results already presented. The substantial new complication arises in computing the derivatives. The first derivatives with respect to the other parameters in the model as shown in Section N29.11.4 are not changed, save for the addition of a time index in the summations, and summation over periods inside the summation over individuals. Then, the derivatives with respect to the parameters described earlier are as already stated. However, the derivatives with respect to the autocorrelation parameters remain. Consider, first, the simpler case in which there is no correlation across parameters. In the gradient, we will require, in addition to the terms already derived,

$$\frac{\partial \log F_{it,r}(*, \mathbf{v}_{irt})}{\partial \rho_k} = \left(x_{k,ji} - \bar{x}_{k,i}\right)\frac{\partial \beta_{k,irt}}{\partial \rho_k}$$

$$\frac{\partial \beta_{k,irt}}{\partial \rho_k} = \sigma_k \frac{\partial v_{k,irt}}{\partial \rho_k}$$

$$\frac{\partial v_{k,irt}}{\partial \rho_k} = v_{k,ir,t-1} + \rho_k \frac{\partial v_{k,ir,t-1}}{\partial \rho_k}$$

The first order difference equation in the third term begins with the Prais-Winsten transformed first random term,

$$\frac{\partial v_{k,ir1}}{\partial \rho_k} = v_{k,ir1} / (1 - \rho_k^2).$$

The second derivative is complex, but relies on the same kind of iterations. When parameters are correlated, then each parameter involves one or more of the autocorrelation coefficients. The derivative of the log probability in this instance must be accumulated by summing several such terms.

N30: Error Components Multinomial Logit Model

N30.1 Introduction

The error components model is an extension of the multinomial logit model that resembles the random effects model in other settings. The model is well suited to repeated choice situations, or panel data applications, though it can be applied in a single cross section as well. The simplest form of the model underlying the observed data would be assumed to be the following random utility specification:

$$\begin{aligned} U_{it}(\text{choice } j) &= U_{jit} \\ &= \boldsymbol{\beta}'\mathbf{x}_{jit} + \varepsilon_{jit} + \theta_j E_{ji},\ j = 1,\ldots,J_i,\ t = 1,\ldots,T_i. \end{aligned}$$

The random, individual specific terms, $(\varepsilon_{1it},\varepsilon_{2it},\ldots,\varepsilon_{Jit})$ are the same type 1 extreme value terms assumed in the basic MNL model. The 'error components,' E_{ji} are alternative specific random individual effects that account for choice situation invariant variation that is unobserved and not accounted for by the other model components. (The parameter θ_j is the standard deviation, made explicit for convenience so it is assumed that $\text{Var}[E_{ji}] = 1$. The means are assumed to equal zero.) As noted, this resembles a random effects model for panel data. The extensions noted below will take this somewhat beyond this specification. The conditional probability for choice j under the IID assumption on ε_{jit} is

$$\text{Prob}(y_{it} = j|E_{1i},E_{2i},\ldots) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{jit} + \theta_j E_{ji}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{qit} + \theta_q E_{qi}\right)},$$

where y_{it} is the index of the choice made. As seen below, under general assumptions, this model relaxes the IIA assumptions of the multinomial logit model. Because the unobserved random effects appear in the probabilities, the model above is not suitable for estimation. It is necessary to integrate the random effects out of the likelihood function. We use the method of maximum simulated likelihood.

N30.2 Command for the Error Components MNL Model

The simplest form of the command for the error components multinomial logit model is

```
ECLOGIT    ; Lhs = variable which indicates the choice made
           ; Choices = a set of J names for the set of choices
           ; ... utility functions, specified by ; Model: ... or ; Rhs and ; Rh2
           ; ECM = the specification of the error components $
```

The 'error components' are individual specific random effects that are distributed across alternatives according to a tree structure. This is somewhat similar to a random constants model, except that in that case, the random terms would be alternative specific – here they need not be. The simple form of the model has one component per alternative, which would be specified with

; ECM = (alternative 1), (alternative 2), ..., (alternative J)

The number of effects in the model is limited to 10 altogether, though in practice, the true limit will be the number of alternatives if that is less than 10. The model structure allows you to capture correlation across alternatives by arranging the error components in a tree structure, with branches that may overlap. It takes the same form as the nested logit model described in Chapter N28. For example, in the model below, all three error components appear in more than one utility function.

$$
\begin{aligned}
U_{i,air} &= \alpha_{air} + \beta_1 gc_{i,air} + \beta_2 ttme_{i,air} + \varepsilon_{i,air} + \theta_1 E_{1,i} \\
U_{i,train} &= \alpha_{train} + \beta_1 gc_{i,train} + \beta_2 ttme_{i,train} + \varepsilon_{i,train} + \theta_2 E_{2,i} + \theta_3 E_{3,i} \\
U_{i,bus} &= \alpha_{bus} + \beta_1 gc_{i,bus} + \beta_2 ttme_{i,bus} + \varepsilon_{i,bus} + \theta_2 \mathrm{E}_{2,i} + \theta_3 E_{3,i} \\
U_{i,car} &= \beta_1 gc_{i,car} + \beta_2 ttme_{i,car} + \varepsilon_{i,car} + \theta_1 E_{1,i} + \theta_3 \mathrm{E}_{3,i}
\end{aligned}
$$

Four of the six correlations in the 4×4 correlation matrix, $\rho(train,bus)$, $\rho(air,car)$ and $\rho(train,car) = \rho(bus,car)$, are nonzero. The specification for this model is

; ECM = (air,car),(train,bus),(train,bus,car)

The error components model may be layered on top of the random parameters (mixed) logit model that is described in Chapter N29. If you are fitting an RPL model, just add the ECM specification, with

; ECM = the specification of the error components

exactly as above. This form of the model is described in Section N29.5.

The full set of options and features for the multinomial logit model and the random parameters model are used in this setting as well. That includes fitted probabilities, inclusive values, all display options described and the simulator described in Chapters N19-N22. Do note, however, that although this model is closely related to the RP model, there is but one parameter vector, and hence, **; Par** has no effect here. The specification

; SDE = list of symbols or values

can be used in the same fashion as **; Rst = list** to constrain the standard deviations of the error components to equal each other or fixed values. For example, with four components, the specification

; SDE = 1,1,ss,ss

forces the first two to equal one and the third and fourth to equal each other. Two other specifications are available.

; SDE = a single value

forces all error components to be equal to that value. Finally, in any specification, if the value is enclosed in parentheses, then the value is merely used to provide the starting value for the estimator, it does not impose any constraints on the final estimates.

N30.3 Heteroscedastic Error Components

In the preceding, the underlying random effects are assumed to be normally distributed with mean zero and standard deviation one; σ_m is the variance parameter for each one. You can specify a heteroscedastic model of the usual form,

$$\text{Var}[E_{im}] = \exp(\boldsymbol{\gamma}_m \mathbf{h}_i)$$

by specifying the set of variables in

; Hfe = list of variables

When the model is fully specified with multiple random effects and numerous variables in the heteroscedasticity function, you may wish to specify which variables appear in the variances of the components. This is done with a modification of the **; ECM** specification. We will detail it with an example. Suppose the specification is

```
; ECM = (air,car),
        (train,bus),
        (train,bus,car)
; Hfe   = hinc,psize
```

Suppose we wish to specify that only *hinc* appears in the first function, only *psize* in the second, and both in the third. The **; ECM** specification would be modified to

```
; ECM = (air,car ! 10) ,
        (train,bus ! 01),
        (train,bus,car ! 11)
```

An exclamation point inside the parentheses after the last name signals that a specification of the heteroscedastic function is to follow. The succeeding specification is a set of zeros and ones where a one indicates that the variable appears in the variance and a zero indicates that it does not. The number of zeros and ones provided is exactly the number of variables that appear in the Hfe list. One abbreviation is available. If you wish for an effect to be homoscedastic, that is, for none of the Hfe variables to appear in the variance, then just end the specification with the exclamation point. For example,

```
; ECM = (air,car ! ) ...
```

specifies that the first of the three effects is homoscedastic. A caution is in order. It is possible to specify a model in which you specify a set of variables in the Hfe list, but remove one or more of these variables from all of the functions. *NLOGIT* cannot verify for you that you have done this. However, such a model cannot be estimated. The most likely outcome is an excessive number of iterations followed after exit with a warning that the Hessian was singular and could not be inverted.

N30.4 General Form of the Error Components Model

Under the preceding assumptions, the general form of the error components multinomial logit model is as follows:

$$U_{jit} = \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \sum_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}.$$

In this form, the model components are

$\mathbf{x}_{jit}$ = attributes and characteristics that enter the utility function of individual i in choice situation t for alternative j,

$\boldsymbol{\beta}_i$ = the parameters for individual i.

We write the model in the random parameters form because the error components model may be added to the random parameters model. This development is continued in Chapter N29.

E_{im} = the individual specific random error components, $m = 1,...,M$.

Note that the error components are not necessarily identified with specific alternatives, though they may be. That depends on your specification of the model. It will be assumed from here onward that the error components have standard normal distributions,

$$E_{im} \sim \text{N}[0,1].$$

The variance of the error component that enters the model is

$$\theta_{im}^2 = \theta_m^2[\exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)]^2,$$

where $\mathbf{h}_i$ = individual choice invariant characteristics that produce heterogeneity in the variances of the error components,

$\boldsymbol{\gamma}_m$ = parameters that enter the heteroscedasticity in the variances of the error components,

and d_{jm} = 1 if E_{im} appears in the utility function for alternative j and 0 otherwise.

Conditional probabilities are built up in the fashion shown in Section N30.1. Estimation of the model is considered in Section N30.7.

N30.5 Results for the Error Components MNL Model

Results for the error components multinomial logit model will consist of the standard model results and any additional descriptive output you have requested. The application below will display the full set of available results. 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

The model results appear generally the same as those for the multinomial logit model. The difference will be the specific results for the error components. For example, the model specified above produces the following results:

```
-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                   MODE
Log likelihood function      -172.88527
Replications for simulated probs. =  50
Halton sequences used for simulations
ECM model with panel has       70 groups
Fixed number of obsrvs./group=         3
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters in utility functions
      GC|    -.04152***      .00684    -6.07  .0000     -.05493   -.02810
    TTME|    -.11186***      .01519    -7.36  .0000     -.14163   -.08209
   A_AIR|    5.21772***     1.35047     3.86  .0001     2.57085   7.86458
 A_TRAIN|    5.53789***      .84435     6.56  .0000     3.88300   7.19278
   A_BUS|    4.41685***      .88537     4.99  .0000     2.68156   6.15214
        |Standard deviations of latent random effects
SigmaE01|    1.26980        1.30876     -.97  .3319    -3.83492   1.29531
SigmaE02|     .64732        1.87332      .35  .7297    -3.02432   4.31896
SigmaE03|    5.07437***     1.45435     3.49  .0005     2.22390   7.92484
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative   E01 E02 E03
+-------------+---+---+---+
|  AIR        | * |   |   |
+-------------+---+---+---+
|  TRAIN      |   | * | * |
+-------------+---+---+---+
|  BUS        |   | * | * |
+-------------+---+---+---+
|  CAR        | * |   | * |
+-------------+---+---+---+
```

The marginal effects in the multinomial logit model are computed as the derivatives of the probability of choice *j* with respect to attribute *k* in alternative *m*. This is

$$\frac{\partial P_j}{\partial x_{km}} = \left[\mathbf{1}(j=m) - P_m\right] P_j \beta_k ,$$

where the function $\mathbf{1}(j = m)$ equals one if *j* equals *m* and zero otherwise. Derivatives and elasticities are obtained by averaging the observation specific values, rather than by computing them at the sample means. The listing reports the sample mean (average partial effect) and the sample standard deviation. Alternative approaches are discussed in Chapter N21. The elasticities in the MNL model display one of the signature features of the IIA assumptions, that cross elasticities are all equal. The error components logit model does not impose that set of assumptions throughout the model. The probabilities are the expected values over the error components, and do not display this characteristic. For example, the specification for the model estimated above produces the following sets of elasticities. Note that two of the elasticities with respect to gc_{air} are the same.

```
-----------------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in AIR
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
     AIR|    -.98551***      .02647    -37.23  .0000    -1.03740    -.93362
   TRAIN|     .35118***      .01489     23.59  .0000      .32200     .38035
     BUS|     .35118***      .01489     23.59  .0000      .32200     .38035
     CAR|     .42718***      .01644     25.99  .0000      .39496     .45940
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+--------------------------------------------------------------------
     AIR|     .39644***      .02337     16.97  .0000      .35065     .44224
   TRAIN|   -3.25055***      .16698    -19.47  .0000    -3.57782   -2.92327
     BUS|    1.92116***      .07857     24.45  .0000     1.76716    2.07516
     CAR|    1.16170***      .06328     18.36  .0000     1.03767    1.28574
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+--------------------------------------------------------------------
     AIR|     .20206***      .01973     10.24  .0000      .16340     .24073
   TRAIN|     .99208***      .07638     12.99  .0000      .84238    1.14178
     BUS|   -3.58608***      .12537    -28.60  .0000    -3.83180   -3.34036
     CAR|     .59892***      .05438     11.01  .0000      .49234     .70549
--------+--------------------------------------------------------------------
```

```
Average elasticity      of prob(alt) wrt GC       in CAR
--------+------------------------------------------------------------------
     AIR|      .36741***       .01962    18.73  .0000       .32896    .40587
   TRAIN|     1.00528***       .06214    16.18  .0000       .88348   1.12709
     BUS|     1.00528***       .06214    16.18  .0000       .88348   1.12709
     CAR|    -1.95363***       .09288   -21.03  .0000    -2.13566  -1.77159
--------+------------------------------------------------------------------
Average elasticity      of prob(alt) wrt TTME     in AIR
--------+------------------------------------------------------------------
     AIR|    -1.78359***       .05257   -33.93  .0000    -1.88663  -1.68055
   TRAIN|      .52235***       .01481    35.28  .0000       .49333    .55137
     BUS|      .52235***       .01481    35.28  .0000       .49333    .55137
     CAR|      .64488***       .01528    42.20  .0000       .61493    .67484
--------+------------------------------------------------------------------
Average elasticity      of prob(alt) wrt TTME     in TRAIN
--------+------------------------------------------------------------------
     AIR|      .30284***       .01404    21.57  .0000       .27532    .33037
   TRAIN|    -2.51160***       .10572   -23.76  .0000    -2.71881  -2.30440
     BUS|     1.53504***       .05514    27.84  .0000      1.42696   1.64312
     CAR|      .88835***       .03761    23.62  .0000       .81463    .96206
--------+------------------------------------------------------------------
Average elasticity      of prob(alt) wrt TTME     in BUS
--------+------------------------------------------------------------------
     AIR|      .14922***       .01008    14.81  .0000       .12947    .16897
   TRAIN|      .77910***       .04298    18.13  .0000       .69486    .86334
     BUS|    -3.94405***       .11847   -33.29  .0000    -4.17625  -3.71185
     CAR|      .44643***       .02814    15.86  .0000       .39127    .50158
--------+------------------------------------------------------------------
Average elasticity      of prob(alt) wrt TTME     in CAR
--------+------------------------------------------------------------------
     AIR|         0.0    .....(Fixed Parameter).....
   TRAIN|         0.0    .....(Fixed Parameter).....
     BUS|         0.0    .....(Fixed Parameter).....
     CAR|         0.0    .....(Fixed Parameter).....
--------+------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|   -.9855    .3512    .3512    .4272
   TRAIN|    .3964  -3.2505   1.9212   1.1617
     BUS|    .2021    .9921  -3.5861    .5989
     CAR|    .3674   1.0053   1.0053  -1.9536

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
TTME    |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -1.7836    .5223    .5223    .6449
   TRAIN|    .3028  -2.5116   1.5350    .8883
     BUS|    .1492    .7791  -3.9441    .4464
     CAR|    .0000    .0000    .0000    .0000
```

N30.6 Application

The following shows the complete set of results for an error components model. This is the full model that has been displayed in parts in the preceding sections.

```
ECLOGIT ; Lhs = mode
        ; Choices = air,train,bus,car
        ; Rhs = one,gc,ttme
        ; ECM = (air,car),(bus,train),(car,bus,train)
        ; Hfe = psize  ? (This is the traveling party size.)
        ; Halton
        ; Pts = 50
        ; Effects: gc(*) $
```

```
-----------------------------------------------------------------------------
Random Parms/Error Comps. Logit Model
Dependent variable                  MODE
Log likelihood function       -195.72367
Replications for simulated probs. =  50
Halton sequences used for simulations
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
    MODE|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters in utility functions
      GC|    -.02550***      .00569    -4.48  .0000     -.03665   -.01434
    TTME|    -.11407***      .01429    -7.98  .0000     -.14208   -.08606
   A_AIR|    5.99505***      .87988     6.81  .0000     4.27051   7.71959
 A_TRAIN|    4.93783***      .70771     6.98  .0000     3.55075   6.32491
   A_BUS|    4.08945***      .68594     5.96  .0000     2.74503   5.43388
        |Standard deviations of latent random effects
SigmaE01|     .18704        7.81521      .02  .9809   -15.13049  15.50458
SigmaE02|     .03909        3.92463      .01  .9921    -7.65305   7.73123
SigmaE03|    7.09290**      3.13987     2.26  .0239      .93886  13.24693
E01PSIZE|    -.09792       26.08617      .00  .9970   -51.22586  51.03003
        |Heterogeneity in variance of latent random effects
E02PSIZE|     .27601       31.04402      .01  .9929   -60.56915  61.12118
E03PSIZE|    -.69217*        .37582    -1.84  .0655    -1.42876    .04442
--------+--------------------------------------------------------------------
```

```
Random Effects Logit Model
 Appearance of Latent Random Effects in Utilities
 Alternative    E01 E02 E03
+-------------+---+---+---+
|  AIR        | * |   |   |
+-------------+---+---+---+
|  TRAIN      |   | * | * |
+-------------+---+---+---+
|  BUS        |   | * | * |
+-------------+---+---+---+
|  CAR        | * |   | * |
+-------------+---+---+---+
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+------------------------------------
     AIR| -1.1618     .4183     .4183     .4196
   TRAIN|   .4857  -2.3989     .9206     .9119
     BUS|   .2187     .4721  -2.4664     .4678
     CAR|   .4818     .8222     .8222  -1.6014
```

N30.7 Technical Details on Maximum Likelihood Estimation

The error components multinomial logit model is estimated by maximum simulated likelihood. The log likelihood is built as follows: The conditional choice probabilities for individual i are

$$\text{Prob}(y_{it} = j|E_{1i},E_{2i},...) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)}.$$

Conditioned on the error components, the choices are independent, so the contribution of individual i to the conditional likelihood function is

$$L_{i,C} = \prod_{t=1}^{T_i} \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)}.$$

The components must be integrated out of the conditional likelihood to obtain the unconditional likelihood that will be maximized. Thus,

$$L_i = \int_{E_{i1}}...\int_{E_{iM}} \prod_{t=1}^{T_i} \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im}\right)} \phi(E_{iM})...\phi(E_{i1})dE_{iM}...dE_{i1}.$$

The integrals cannot be expressed in closed form. However, the form of the likelihood is particularly convenient, since the error components are independent standard normal. We use simulation instead. The simulated likelihood function for individual i is

$$L_{i,S} = \frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im,r}\right)}{\sum_{q=1}^{J_i}\exp\left(\boldsymbol{\beta}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m \exp(\boldsymbol{\gamma}_m'\mathbf{h}_i)E_{im,r}\right)},$$

in which $E_{im,r}$ is a set of M independent standard random normal draws. These may be pseudorandom draws or Halton sequences. The function to be maximized is

$$\log L_S = \sum_{i=1}^{N} \log L_{i,S}.$$

Analysis of this model follows precisely along the lines of the random parameters models described in Chapter N29. The function and analytic first derivatives and second derivatives are obtained by simulation. (The derivatives are surprisingly simple in spite of the formidable appearance of the function.) The BFGS method is used for optimization. Starting values are the MNL values for the slopes, and zeros for all variance terms.

N31: Nonlinear Random Parameters Logit Model

N31.1 Introduction

The nonlinear random parameters logit model provides several extensions of the mixed logit model described in Chapter N29. The central feature is the specification of the utility functions, which take the general form

$$U_{it}\,(alt=j) = \sigma_i\,[V(\boldsymbol{\beta}_i,\mathbf{x}_{it,j}) + EC_i\,] + \varepsilon_{it,j},$$

$$V(\boldsymbol{\beta}_i,\mathbf{x}_{it,j}) = \text{any nonlinear function of attributes, characteristics, parameters,}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{w}_i \qquad \text{(random parameters),}$$

$$EC_i = \sum\nolimits_{c=1}^{C} d_{ic}E_{ic} \qquad \text{(error components),}$$

$$\sigma_i = \exp(\boldsymbol{\delta}'\mathbf{r}_i + \tau v_i - \tau^2/2) \qquad \text{(scaled MNL).}$$

The IID assumption is maintained for the random component of the utility functions. The overall model is a multinomial logit with this extended form of the utility functions;

$$\text{Prob}(alt=j) = \frac{\exp\left\{\sigma_i\left[V(\boldsymbol{\beta}_i, x_{it,j}) + \Sigma_{c=1}^{C} d_{ic}E_{ic}\right]\right\}}{\sum_{j=1}^{J_{it}} \exp\left\{\sigma_i\left[V(\boldsymbol{\beta}_i, x_{it,j}) + \Sigma_{c=1}^{C} d_{ic}E_{ic}\right]\right\}}.$$

This specification combines the random parameters specification of Chapter N29, the error components logit model of Chapter N30 and the scaled multinomial logit model of Chapter N24, and extends the utility functions beyond the linear specification assumed up to this point.

N31.2 Model Command for Nonlinear RP Models

The command structure for this general model requires four parts, the choice set, the random parameters, the nonlinear parts of the utility functions and the utility functions themselves. The general format of the basic command is

```
NLRPLOGIT  ; Lhs =          choice
           ; Choices = …  as usual
           ; Fcn = …       definition of random parameters
           ; Fn1 = …       definition of nonlinear components of utilities
           ; Model: …      definition of utility functions $
```

The choice set definition is the same as for all other model specification in *NLOGIT*. The other essential parts and various options are described below.

N31.2.1 Parameter Definition

The parameters definition begins with same sort of definition used in the nonlinear optimization (**MAXIMIZE**, **NLSQ**, etc.) commands. You must provide the names you will be using for the parameters in the model and starting values.

; Labels = the set of labels
; Start = starting values for means of random parameters)

The parameters in the model take the form

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{w}_i.$$

The starting values for any nonzero elements of **Δ** and **Γ** will be zero. You provide the values for **β**. The remainder of the parameter definition uses the same features as described in Chapter N29 for the linear utilities, random parameters model. The setup uses

; Fcn = definitions of random parameters, with all other features from Chapter N29
; Correlated
; RPL = list for observed heterogeneity in means

The parameter heteroscedasticity, **; Hfn = list**, and the **; AR1** features are not built into this form of the random parameters model. Section N29.3.7 describes a device, **; SDV = list**, that can be used to impose certain restrictions on the standard deviations of the random parameters.

In the **; Fcn** definition, use the labels that you defined in your **; Labels** definition. Note that it is not necessary for all parameters named in the **; Labels** definition to be random. Use **; Fcn** to define the distributions only for those parameters that are actually random in the model.

The set of features is restricted a bit. The distributions that may be used in the **; Fcn** setup are '*c*' (constant), '*n*' (normal), '*u*' (uniform), '*t*' (triangular), '*o*' (one sided triangular), '*z*' (truncated normal) and '*s*' (skew normal). The '*l*' (lognormal) distribution is not supported in the command, however, if you require a lognormal parameter, γ, you can use exp(β) where β is normally distributed. (See the technical details for an additional note on this usage.)

N31.2.2 Nonlinear Components

The nonlinear utility functions are constructed from nonlinear functions, which can be constructed recursively. These functions are defined to be built into the utility functions.

; Fn1 = alias1 = a nonlinear function of parameters and data
; Fn2 = alias2 = a nonlinear function of parameters and data
… up to
; Fn50 = alias50 = a nonlinear function of parameters and data

Note, these are not necessarily the utility functions. Utility functions are constructed from these parts.

For example,

; Fn1 = frstpart = a0 + exp(a1*x1 + a2*x2)
; Fn2 = scndpart = frstpart + c2*c3*z
; Fn3 = both = frstpart * scndpart

might be followed with

; Model: U(alt1,alt2) = both / U(alt3) = frstpart $

The definitions of the nonlinear components follow the construction used in **MAXIMIZE**, **NLSQ**, etc. They may involve any number of layers of parentheses, functions such as Log, Exp, Phi, etc., and the usual operators, +, -, *, /, ^.

N31.2.3 Utility Functions

The utility functions are defined to equal one of the nonlinear components. Any one of the nonlinear functions may be the applicable utility function, as shown in the example above.

; Model: U(list of alts) = alias … /
U(list of alts) = alias …

Each utility function is defined to equal one of the nonlinear functions. The utility functions do not specify any more mathematics. The utility function only identifies which of the nonlinear components should be used for that alternative. (See the technical details for a note about using information about specific nonlinear components in the utility functions to speed up the computations.)

N31.2.4 The Error Components Model

The utility functions may contain error components (random effects), as in the **ECLOGIT** model Section N29.5 and Chapter N30. Use

; ECM = definition of error components.

N31.2.5 Scaling function, σ_i – The Scaled Nonlinear RP Model

The overall common scaling parameter is added to the model with

; SMNL

Without this specification, $\sigma_i = 1$. With **; SMNL**, $\sigma_i = \exp(\boldsymbol{\delta}'\mathbf{r}_i + \tau v_i - \tau^2/2)$ where $v_i \sim N[0,1]$. You can specify a starting value or a fixed value for τ with

; Tau = value for starting value or [value] to fix.

The variance may also depend on observed variables by adding

; Hfr = list of r variables in σ_i

to the specification. The specification **; SCV** that is used in Chapter N33 to allow for correlation between the random parts of σ_i and β_{ki} is not available here.

N31.2.6 Panel Data

Panel data or repeated choice data are identified using

; Pds = definition if this is a panel application

as in the random parameters and latent class models. The treatment is the same as in the other models. The parameters are time invariant and the likelihood function is computed accordingly when the parameters are estimated.

N31.2.7 Ignored Attributes

Section N18.9 describes a device with which you can account for cases in which some but not all individuals in the sample indicate that they have not considered specific attributes in their decision. The modeling response to this case is to replace the relevant coefficient, not the attribute, with zero, for this person, and adjust the log likelihood and derivatives accordingly. In the nonlinear utility function of this chapter, there is no exact correspondence between coefficients and variables, so this device cannot be used. That is, when the utility function is U(choice) = $\beta_1 x_1 + \beta_2 x_2 + \ldots \varepsilon$, our modeling strategy is to set, e.g., β_1 to zero when we find x_1 equals -888. But, when the utility function is merely of the form U(choice)=V(x,β) + ε, finding that a particular x equals -888 does not imply that a particular β should equal zero. However, you can make the association yourself, explicitly as follows. The syntax,

; 888: (attribute / label), …

states that when variable *attribute* equals -888, then coefficient *label* is to be set to zero. The coefficient label is one that appears in the **; Labels = list** part of the command.

N31.3 Results

Standard results, as shown in the applications below, include the usual statistical output – diagnostic statistics, coefficient estimates, and so on. Descriptive results from

; List
; Crosstab
and **; Show**

may all be used as with other multinomial choice models. See Section N19.3 for details.

N31.3.1 Individual Specific Parameters

Individual specific parameters and standard deviations are saved with

; Parameters.

Figure N31.1 below shows the results from the first application below.

Matrix - BETA_I

[70, 3] Cell: -0.0572491

	1	2	3
1	-0.0452201	0.0257665	-0.119295
2	-0.0235284	0.0213341	-0.0383062
3	-0.0101252	-0.0292803	0.0503272
4	-0.0757293	-0.00701655	-0.0463007
5	-0.0391433	0.0937987	-0.188775
6	-0.0228655	-0.0186891	0.043203
7	-0.0625737	-0.0330299	0.0542642
8	-0.064325	-0.0340054	0.133556
9	-0.014717	-0.000960952	0.0258255
10	0.00185599	-0.0417622	0.0681077

Matrix - SDBETA_I

[70, 3] Cell: 0.00343419

	1	2	3
1	0.00343419	0.0226245	0.0157433
2	0.0105913	0.0342231	0.0727215
3	0.00647983	0.00996981	0.0317199
4	0.0147673	0.079081	0.117234
5	0.0228172	0.056897	0.137721
6	0.0102192	0.0261084	0.0420834
7	0.0148523	0.0433434	0.0460317
8	0.00841665	0.0172	0.0321703
9	0.00536789	0.00463855	0.00857757
10	0.00703209	0.024138	0.0255353

Figure N31.1 Matrices Saved with ; Par

The simulator for analyzing scenarios and changes in market shares that is described in Chapter N22 is used in exactly the same way for this nonlinear model. All aspects of the command are identical here.

N31.3.2 Willingness to Pay

In principle, willingness to pay is computed as

WTP = *marginal utility of attribute / marginal utility of income*

When income is not in the data set, researchers often use a cost variable as a surrogate, with the negative of the disutility of cost being a surrogate for the marginal utility of income. Thus,

WTP = - *marginal utility of attribute / marginal disutility of cost*.

When the utility functions are linear and have generic coefficients, WTP is typically computed as a ratio of coefficients. These may be fixed, as in the MNLOGIT model, or they may be random, as shown in Section N29.8.4.

In the nonlinear model of this chapter, this is an ambiguous calculation because the marginal utility (derivative of utility) with respect to an attribute is likely to depend on which utility function is used (that is, which one is differentiated). We do not have a right answer to propose in this case. You can specify how the computation is to be done by using

; WTP = choice [attribute / cost]

where the choice gives the name of the utility function to be differentiated. The attribute and cost variables define which two variables are to be the denominators of the derivative. You may have up to five of these in the command. Each provides a column in the matrices *wtp_i* and *sdwtp_i* that are created by this procedure.

The WTP values are saved in the matrix *wtp_i*. You may also expand the matrix into variable(s) in the data set as follows:

1. Use **CREATE** or **NAMELIST ; (New);…** to create the variable or variables if more than one.
2. Change **; WTP = definition** to **; WTP (variable or namelist) = definition**.

For example, the following will create a new variable, *invtwtp* with the matrix *wtp_i*:

```
CREATE          ; invtwtp $
NLRPLOGIT …     ; WTP (invtwtp) = car[invt / invc]
```

N31.4 Application

The following small contrived example illustrates the structure of the model command and shows several of the options.

```
SAMPLE       ; 1-840 $
CREATE       ; zrpl = Rnu(0,1) $
NLRPLOGIT    ; Lhs = mode
             ; Choices = air,train,bus,car
             ; Pds = 3 ; Labels = a0,b1,b2,b3
             ; Start 8.530310,-.12119,-.03512,.17651
             ; Fcn = b1(n),b2(n),b3(n)
             ; Halton
             ; Draws = 25 ; Correlated
             ; RPL = zrpl
             ; Fn1 = utility1 = a0+b1*gc+b2*ttme+b2*b3*invc+b2*(1+b3)*invt
             ; Fn2 = utility2 =    b1*gc+b2*ttme+b2*b3*invc+b2*(1+b3)*invt
             ; Model: U(train,bus,car) = utility1 / U(air) = utility2
             ; Effects: gc(*)
             ; Full $
```

```
-----------------------------------------------------------------------------
Nonlinear Utility Mixed Logit Model
Dependent variable                  MODE
Log likelihood function      -195.14005
Restricted log likelihood    -291.12182
Chi squared [  13 d.f.]       191.96354
Significance level               .00000
McFadden Pseudo R-squared      .3296962
Estimation based on N =    210, K =  13
Inf.Cr.AIC  =    416.3 AIC/N =    1.982
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .32971.0000
Constants only   -283.7588  .31231.0000
At start values -2281.0900  .91451.0000
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
NLM model with panel has      70 groups
Fixed number of obsrvs./group=        3
Hessian is not PD. Using BHHH estimator
Number of obs.=   210, skipped    0 obs

--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      B1|     .26132***      .07479     3.49  .0005      .11473    .40792
      B2|    -.03404         .02190    -1.55  .1201     -.07696    .00888
      B3|    1.04016        1.17787      .88  .3772    -1.26843   3.34875
        |Nonrandom parameters in utility functions
      A0|    19.9877***     2.60009     7.69  .0000     14.8916   25.0838
        |Heterogeneity in mean, Parameter:Variable
  B1:ZRP|    -.16776         .11919    -1.41  .1593     -.40137    .06585
  B2:ZRP|     .02450         .02718      .90  .3674     -.02877    .07777
  B3:ZRP|    -.95112         .91132    -1.04  .2966    -2.73728    .83503
        |Diagonal values in Cholesky matrix, L.
    NsB1|     .23644***      .05321     4.44  .0000      .13215    .34072
    NsB2|     .04972*        .02665     1.87  .0621     -.00251    .10194
    NsB3|     .09373         .12515      .75  .4539     -.15155    .33902
        |Below diagonal values in L matrix. V = L*Lt
   B2:B1| .74092D-04         .00520      .01  .9886 -.10117D-01  .10266D-01
   B3:B1|    -.17472         .43851     -.40  .6903    -1.03419    .68474
   B3:B2|    -.25865         .28191     -.92  .3589     -.81119    .29388
        |Standard deviations of parameter distributions
    sdB1|     .23644***      .05321     4.44  .0000      .13215    .34072
    sdB2|     .04972*        .02664     1.87  .0620     -.00250    .10194
    sdB3|     .32591         .27935     1.17  .2433     -.22160    .87342
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Parameter Matrix for Heterogeneity in Means.
--------+--------------
Delta   |          ZRPL
--------+--------------
      B1|      -.167756
      B2|      .0244996
      B3|      -.951124
Correlation Matrix for Random Parameters
--------+--------------------------------------------
Cor.Mat.|            B1            B2            B3
--------+--------------------------------------------
      B1|       1.00000     .00149026      -.536113
      B2|     .00149026       1.00000      -.794438
      B3|      -.536113      -.794438       1.00000
--------+--------------------------------------------
+-----------------------------------------------------+
| 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|    1.58235***      .12494    12.67  .0000     1.33748   1.82722
   TRAIN|    -.56845***      .04624   -12.29  .0000     -.65908   -.47782
     BUS|    -.34267***      .03449    -9.93  .0000     -.41027   -.27507
     CAR|    -.35495***      .02320   -15.30  .0000     -.40041   -.30949
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+--------------------------------------------------------------------
     AIR|    -.88959***      .11050    -8.05  .0000    -1.10616   -.67302
   TRAIN|    6.19651***      .29665    20.89  .0000     5.61509   6.77793
     BUS|   -3.02442***      .16857   -17.94  .0000    -3.35480  -2.69403
     CAR|   -1.96623***      .09900   -19.86  .0000    -2.16026  -1.77220
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+--------------------------------------------------------------------
     AIR|    -.21430***      .03789    -5.66  .0000     -.28857   -.14004
   TRAIN|   -2.27244***      .12462   -18.23  .0000    -2.51670  -2.02819
     BUS|    6.08597***      .32191    18.91  .0000     5.45503   6.71691
     CAR|   -1.84942***      .12706   -14.56  .0000    -2.09845  -1.60040
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in CAR
--------+--------------------------------------------------------------------
     AIR|    -.42448***      .04668    -9.09  .0000     -.51598   -.33298
   TRAIN|   -2.08423***      .10344   -20.15  .0000    -2.28696  -1.88150
     BUS|   -2.55957***      .14983   -17.08  .0000    -2.85323  -2.26592
     CAR|    3.67551***      .23676    15.52  .0000     3.21147   4.13955
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
     GC |
--------+-----------------------------------
     AIR|  1.5824   -.5685   -.3427   -.3550
   TRAIN|  -.8896   6.1965  -3.0244  -1.9662
     BUS|  -.2143  -2.2724   6.0860  -1.8494
     CAR|  -.4245  -2.0842  -2.5596   3.6755
```

This example adds the scaled MNL feature to the model above, including heteroscedasticity based on household income.

```
NLRPLOGIT ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Labels = a0,b1,b2,b3
          ; Start 8.530310,-.12119,-.03512,.17651
          ; Fn1 = utility1 = a0+b1*gc+b2*ttme+b2*b3*invc+b2*(1+b3)*invt
          ; Fn2 = utility2 =    b1*gc+b2*ttme+b2*b3*invc+b2*(1+b3)*invt
          ; Model: U(train,bus,car) = utility1 / U(air) = utility2
          ; Fcn = b1(n),b2(n),b3(n)
          ; Correlated
          ; RPL = zrpl ; Halton
          ; Pds = 3 ; Draws = 25
          ; SMNL ; Hfr = hinc $
```

```
-----------------------------------------------------------------------------
Nonlinear Utility Mixed Logit Model
Dependent variable                    MODE
Log likelihood function         -205.21019
Restricted log likelihood       -291.12182
Chi squared [  15 d.f.]          171.82325
Significance level                  .00000
McFadden Pseudo R-squared         .2951054
Estimation based on N =    210, K =  15
Inf.Cr.AIC  =    440.4 AIC/N =    2.097
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .29511.0000
Constants only   -283.7588  .27681.0000
At start values -2057.9615  .90031.0000
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
NLM model with panel has      70 groups
Fixed number of obsrvs./group=        3
Hessian is not PD. Using BHHH estimator
Variable IV parameters are denoted s_...
Number of obs.=   210, skipped    0 obs
```

```
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      B1|      .00886         .01190      .74  .4569     -.01447    .03219
      B2|     -.02324         .01623    -1.43  .1522     -.05504    .00857
      B3|     -.16142         .50856     -.32  .7509    -1.15818    .83534
        |Nonrandom parameters in utility functions
      A0|     5.97289***     2.07038     2.88  .0039     1.91502  10.03075
        |Heterogeneity in mean, Parameter:Variable
  B1:ZRP|     -.06099         .03800    -1.61  .1084     -.13546    .01348
  B2:ZRP|      .05019         .03535     1.42  .1556     -.01909    .11947
  B3:ZRP|    -1.52411*        .90703    -1.68  .0929    -3.30185    .25364
        |Diagonal values in Cholesky matrix, L.
    NsB1|      .03945         .02710     1.46  .1455     -.01367    .09257
    NsB2|      .01591         .01216     1.31  .1907     -.00793    .03975
    NsB3|      .08381         .21933      .38  .7024     -.34607    .51368
        |Below diagonal values in L matrix. V = L*Lt
   B2:B1|      .02152*        .01279     1.68  .0926     -.00356    .04659
   B3:B1|     -.10448         .13597     -.77  .4422     -.37097    .16201
   B3:B2|     -.49465         .34683    -1.43  .1538    -1.17443    .18512
        |Heteroscedasticity in NLRPLRP scale factor
  sdHINC|      .02729*        .01637     1.67  .0956     -.00480    .05938
        |Variance parameter tau in GMX scale parameter
TauScale|     1.50229***      .51522     2.92  .0035      .49248   2.51210
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     2.37693        3.64049      .65  .5138    -4.75830   9.51217
        |Standard deviations of parameter distributions
    sdB1|      .03945         .02710     1.46  .1455     -.01367    .09257
    sdB2|      .02676         .01652     1.62  .1052     -.00562    .05914
    sdB3|      .51247         .35101     1.46  .1443     -.17549   1.20042
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
Parameter Matrix for Heterogeneity in Means.
--------+--------------
Delta   |          ZRPL
--------+--------------
      B1|     -.0609927
      B2|      .0501894
      B3|      -1.52411
Correlation Matrix for Random Parameters
--------+------------------------------------------
Cor.Mat.|            B1            B2            B3
--------+------------------------------------------
      B1|       1.00000       .803992      -.203880
      B2|       .803992       1.00000      -.737888
      B3|      -.203880      -.737888       1.00000
--------+------------------------------------------
+----------------------------------------------------+
| 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|    -.56131***      .05836    -9.62  .0000     -.67569   -.44693
   TRAIN|     .14415***      .02368     6.09  .0000      .09773    .19056
     BUS|     .14940***      .02027     7.37  .0000      .10967    .18913
     CAR|     .10596***      .01476     7.18  .0000      .07702    .13489
--------+-----------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+-----------------------------------------------------------------
     AIR|     .16399***      .01828     8.97  .0000      .12817    .19981
   TRAIN|   -1.17338***      .16860    -6.96  .0000    -1.50384   -.84292
     BUS|     .31627***      .04916     6.43  .0000      .21993    .41261
     CAR|     .27626***      .03927     7.03  .0000      .19929    .35323
--------+-----------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+-----------------------------------------------------------------
     AIR|     .16347***      .01709     9.57  .0000      .12997    .19696
   TRAIN|     .30197***      .04769     6.33  .0000      .20850    .39543
     BUS|   -1.16451***      .15899    -7.32  .0000    -1.47612   -.85289
     CAR|     .35290***      .05104     6.91  .0000      .25286    .45293
--------+-----------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in CAR
--------+-----------------------------------------------------------------
     AIR|     .19369***      .01961     9.88  .0000      .15526    .23212
   TRAIN|     .43537***      .06028     7.22  .0000      .31722    .55352
     BUS|     .54706***      .07713     7.09  .0000      .39588    .69824
     CAR|    -.64829***      .09277    -6.99  .0000     -.83012   -.46646
--------------------------------------------------------------------------
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
     GC |
--------+------------------------------------
     AIR|  -.5613    .1441    .1494    .1060
   TRAIN|   .1640  -1.1734    .3163    .2763
     BUS|   .1635    .3020  -1.1645    .3529
     CAR|   .1937    .4354    .5471   -.6483
```

N31.4.1 Elasticities and Partial Effects

In principle, elasticities and partial effects are computed the same here as in other models. However, one difference is that the marginal utility part of the partial effect is not a constant coefficient, it is the derivative of the nonlinear utility function with respect to a variable. This can differ across the choices when a generic coefficient would not. The result is that there is likely to be somewhat more variation in cross elasticities than might otherwise be the case. The example below computes **; Effects: gc(*)** for the model in the previous example.

```
+----------------------------------------------------+
| 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|    -.56131***      .05836    -9.62  .0000     -.67569   -.44693
   TRAIN|     .14415***      .02368     6.09  .0000      .09773    .19056
     BUS|     .14940***      .02027     7.37  .0000      .10967    .18913
     CAR|     .10596***      .01476     7.18  .0000      .07702    .13489
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in TRAIN
--------+--------------------------------------------------------------------
     AIR|     .16399***      .01828     8.97  .0000      .12817    .19981
   TRAIN|   -1.17338***      .16860    -6.96  .0000    -1.50384   -.84292
     BUS|     .31627***      .04916     6.43  .0000      .21993    .41261
     CAR|     .27626***      .03927     7.03  .0000      .19929    .35323
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in BUS
--------+--------------------------------------------------------------------
     AIR|     .16347***      .01709     9.57  .0000      .12997    .19696
   TRAIN|     .30197***      .04769     6.33  .0000      .20850    .39543
     BUS|   -1.16451***      .15899    -7.32  .0000    -1.47612   -.85289
     CAR|     .35290***      .05104     6.91  .0000      .25286    .45293
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in CAR
--------+--------------------------------------------------------------------
     AIR|     .19369***      .01961     9.88  .0000      .15526    .23212
   TRAIN|     .43537***      .06028     7.22  .0000      .31722    .55352
     BUS|     .54706***      .07713     7.09  .0000      .39588    .69824
     CAR|    -.64829***      .09277    -6.99  .0000     -.83012   -.46646
-----------------------------------------------------------------------------
Elasticity wrt change of X in row choice on Prob[column choice]
--------+------------------------------------
     GC |
--------+------------------------------------
     AIR|  -.5613     .1441     .1494     .1060
   TRAIN|   .1640   -1.1734     .3163     .2763
     BUS|   .1635     .3020   -1.1645     .3529
     CAR|   .1937     .4354     .5471    -.6483
```

N31.4.2 Variables Saved in the Data Set

The standard set of retained results can be retained for the nonlinear model as well. These include

; IVB = variable	to retain the log sum variable in the probabilities
; Actualy = variable	replicates in each row of an observation the actual outcome chosen by the individual
; Fittedy = variable	replicates in each row of an observation the index of the alternative that has the highest probability
; Utility = variable	retains in each row of an observation the utility computed for that alternative for the individual.

It is also possible to retain individual specific partial effects or elasticities with the standard syntax,

; Effects: attribute(choice) = variable

for a single attribute and a single alternative.

N31.5 Technical Details

Computation Time

This model is fit by maximum simulated likelihood. Because it is necessary to approximate the derivatives and simulate them as well, this is a very time consuming estimator. Every possible avenue to minimize the amount of computation is taken internally. There is a point at which you can reduce the computation in a way that is not evident to the program internally. In order to compute the full set of probabilities needed to compute the log likelihood, *NLOGIT* must fill a matrix that contains all *M* nonlinear component functions for every one of the *J* utility functions. However, it may be that some of the nonlinear components are not needed for some of the utilities. Consider an example

```
; Fn1 = K1        = a1 + a2*x2*x2 + a3*x3
; Fn2 = K2        = b1 + b2*q
; Fn3 = K3        = c1 + c2*xc
; Fn4 = Ratio     = K2/K3
; Model: U(alt1) = K1 / U(alt2) = K2 / U(alt3) = Ratio
```

By default, *NLOGIT* will compute all four functions for each of the 3 alternatives. But, in fact, functions *K*2, *K*3 and *Ratio* are not needed for *U*(*alt1*) and *K*1, *K*3 and *Ratio* are not needed for *U*(alt2) and *K*1 is not needed for *U*(*alt3*). An extension of the model command that can increase the speed of the computations considerably is to specify the utility functions and name explicitly the other functions needed to compute it. For this example, we could use

```
; Model: U(alt1) = K1 [.] /
         U(alt2) = K2 [.] /
         U(alt3) = Ratio [K2,K3]
```

The [.] indicates that no other functions are needed to compute this one. This would bring a substantial time saving (greater than 50%), as only 5, rather than all 12 functions×utilities are computed.

Lognormal Parameters

It is noted earlier, the available set of distributions for the random parameters does not include the lognormal. You can exponentiate a normally distributed parameter to achieve the same result. However, the long, thick tail of the lognormal distribution can produce extreme values of the parameters and implausible results, as well as instability in the estimator itself. You can dampen this effect by using the truncated normal, '(z)' specification instead of the normal '(n).' This distribution removes the upper and lower 2.5% of the distribution, which is where the mischief resides. Defining beta(z) in ';Fcn, then exp(beta) in your model may produce better results.

Internal Limits

There are a few technical constraints and internal limits on this estimator.

- $T \leq 20$ in panel data sets;
- $J \leq 100$ – choice sets are limited to 100 alternatives;
- $K \leq 100$ parameters – this limit is set in the **; Labels** specification;
- Data must be individual, not shares, ranks or frequencies;
- Data on attributes and characteristics must be in the multiple line format;
- Utilities must be specified using **; Model: U(…) = …,** not with **; Rhs** and **; Rh2**;
- **; Wts** may not be used;
- **; MCS** is not supported – it is not possible to simulate the outcome variable;
- **; Checkdata** is not available;
- **; IAS** and **; Choices = …(alts)…** for the IIA tests are not available.

Controlling the Simulation

As in earlier cases (**RPLOGIT** and **SMNLOGIT**), you can control the simulations in part with

; Halton	to use Halton sequences rather than pseudorandom draws
; Draws =	**number of draws or Halton values**
; Shuffled	to use shuffled pseudorandom or Halton draws.

(The third of these has relatively limited impact on the results.)

N32: Latent Class Random Parameters Model

N32.1 Introduction

The latent class random parameters (LCRP) model combines the latent class model described in Chapter N25 with the random parameters model in Chapter N29. It accommodates two layers of heterogeneity. The model assumes that there are Q distinct classes of individuals in the population distinguished by different distributions of parameters within the classes. The random parameters aspect of the model specifies a continuous distribution of model parameters within each class. The classes are distinguished by the characteristics of these distributions. Thus, in a two class model, we have distributions of random parameters with means $\boldsymbol{\beta}_q$ and covariance matrices $\boldsymbol{\Sigma}_q$, $q = 1,\ldots,Q$. Full details on the model specification and the estimator appear in Section N32.4.

N32.2 Command

The command for the latent class random parameters model consists of

LCRPLOGIT ; Lhs = choice set definition
; Choices = list of choice labels
; Pds = definition if this is a panel, or stated choice experiment
; … definition of the utility functions
; … definition of the random parameters model
; … definition of the latent class model
; … other options $

The utility functions are specified using either **; Rhs = … list… ; Rh2 = … list …** or **; Model: U(…) = ….** All of the features used for other models such as **RPLOGIT**, **CLOGIT**, **NLOGIT**, etc. are used here in the same way.

The random parameters are formulated using a restricted version of the RPLOGIT model. Use

; RPL
; Fcn = name (type), name(type),… for $\boldsymbol{\beta}_{qi} = \boldsymbol{\beta}_q + \boldsymbol{\Gamma}_q\mathbf{w}_{iq}$

or **; RPL = list of variables**
; Fcn = name (type), name(type),… for $\boldsymbol{\beta}_{qi} = \boldsymbol{\beta}_q + \boldsymbol{\Delta}_q\mathbf{z}_i + \boldsymbol{\Gamma}_q\mathbf{w}_{iq}$.

If you are using the first form, then the **; RPL** may be omitted (as it is implied by the model name). This estimator does not support heterogeneity in variances of the random parameters, so **; Hfn = list** is not used here. The types may be 'N' (normal), 'T' (triangular), 'O' (one sided triangular) or 'C' (constant, which is the same as not including that parameter in the **; Fcn** list). Draws for the simulation are controlled with

; Halton
; Draws = number of draws

Note that this command looks for **; Draws** rather than **; Pts** for the number of replicates for the simulation – **; Pts** is used to specify the number of latent classes.

The latent class model is specified as in Chapter N25, with

; LCM
; Pts = number of classes

or **; LCM = list of variables that appear in the class probabilities**
; Pts = the number of classes.

Since it is implied by the model name, you may omit the **;LCM** if you are using the first form of the model.

N32.2.1 Output Options

The following general options for display of the model and data are available:

; Describe
; Crosstab
; Covariance

N32.2.2 Post Estimation

The following general options are based on the results of the estimation step. In general, where a parameter vector is called for, the program uses the individual specific estimate of E[β_{qi}|i]. To begin, the estimated posterior probabilities,

$$\hat{\pi}(q\,|\,i)=\frac{\hat{\pi}_{qi}\hat{L}_{qi}}{\sum_{q=1}^{Q}\hat{\pi}_{qi}\hat{L}_{qi}}$$

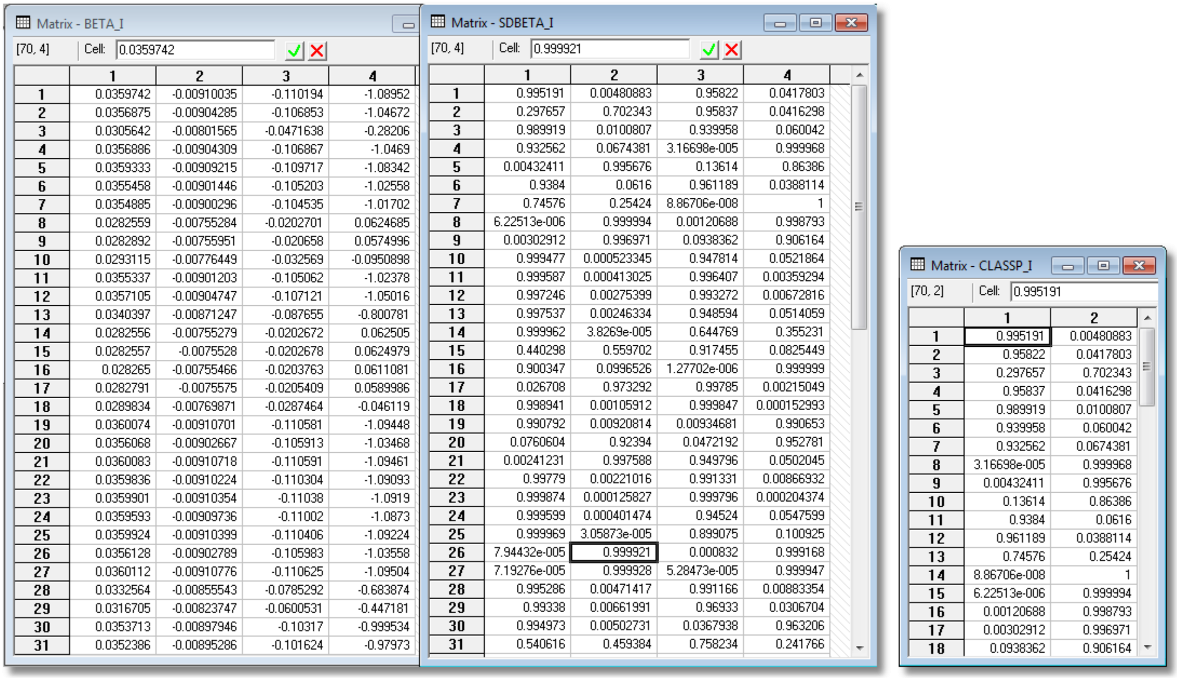
are computed. The contributions to the likelihood within the classes are

$$\hat{L}_{qi}=\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i}\text{Prob}(choice_{it}\,|\,\hat{\boldsymbol{\beta}}_q,\mathbf{x}_{it})$$

where $\hat{\boldsymbol{\beta}}_q$ is the full set of parameters in class q, i.e., $[\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Gamma}},\hat{\boldsymbol{\Delta}}]$ Once the posterior probabilities are obtained, the estimate of the applicable class is the one with the largest posterior probability. The parameter vector for that class is estimated using the method described in Section N29.8. This class specific, person specific estimator is used in the computations that follow. Post estimation results generally available are as follows:

; List	requests a list of choice probabilities
; Effects:…	requests the partial effects or elasticities
; Actualy = name	replicates the actual choice in each row of the observation
; Fittedy = name	replicates the index of the choice with the highest probability
; Prob = name	stores the predicted probabilities
; Utility = name	stores the utility functions based on the estimated parameters

The setting **; Parameters** saves three matrices, *beta_i* and *sdbeta_i* are for the individual specific estimates described above; *classp_i* contains the estimated posterior probabilities. Figure N32.1 shows the results for the first of the applications in Section N32.3.



Matrix - BETA_I

[70, 4] Cell: 0.0359742

	1	2	3	4
1	0.0359742	-0.00910035	-0.110194	-1.08952
2	0.0356875	-0.00904285	-0.106853	-1.04672
3	0.0305642	-0.00801565	-0.0471638	-0.28206
4	0.0356886	-0.00904309	-0.106867	-1.0469
5	0.0359333	-0.00909215	-0.109717	-1.08342
6	0.0355458	-0.00901446	-0.105203	-1.02558
7	0.0354885	-0.00900296	-0.104535	-1.01702
8	0.0282559	-0.00755284	-0.0202701	0.0624685
9	0.0282892	-0.00755951	-0.020658	0.0574996
10	0.0293115	-0.00776449	-0.032569	-0.0950898
11	0.0355337	-0.00901203	-0.105062	-1.02378
12	0.0357105	-0.00904747	-0.107121	-1.05016
13	0.0340397	-0.00871247	-0.087655	-0.800781
14	0.0282556	-0.00755279	-0.0202672	0.062505
15	0.0282557	-0.0075528	-0.0202678	0.0624979
16	0.028265	-0.00755466	-0.0203763	0.0611081
17	0.0282791	-0.0075575	-0.0205409	0.0589986
18	0.0289834	-0.00769871	-0.0287464	-0.046119
19	0.0360074	-0.00910701	-0.110581	-1.09448
20	0.0356068	-0.00902667	-0.105913	-1.03468
21	0.0360083	-0.00910718	-0.110591	-1.09461
22	0.0359836	-0.00910224	-0.110304	-1.09093
23	0.0359901	-0.00910354	-0.11038	-1.0919
24	0.0359593	-0.00909736	-0.11002	-1.0873
25	0.0359924	-0.00910399	-0.110406	-1.09224
26	0.0356128	-0.00902789	-0.105983	-1.03558
27	0.0360112	-0.00910776	-0.110625	-1.09504
28	0.0332564	-0.00855543	-0.0785292	-0.683874
29	0.0316705	-0.00823747	-0.0600531	-0.447181
30	0.0353713	-0.00897946	-0.10317	-0.999534
31	0.0352386	-0.00895286	-0.101624	-0.97973

Matrix - SDBETA_I

[70, 4] Cell: 0.999921

	1	2	3	4
1	0.995191	0.00480883	0.95822	0.0417803
2	0.297657	0.702343	0.95837	0.0416298
3	0.989919	0.0100807	0.939958	0.060042
4	0.932562	0.0674381	3.16698e-005	0.999968
5	0.00432411	0.995676	0.13614	0.86386
6	0.9384	0.0616	0.961189	0.0388114
7	0.74576	0.25424	8.86706e-008	1
8	6.22513e-006	0.999994	0.00120688	0.998793
9	0.00302912	0.996971	0.0938362	0.906164
10	0.999477	0.000523345	0.947814	0.0521864
11	0.999587	0.000413025	0.996407	0.00359294
12	0.997246	0.00275399	0.993272	0.00672816
13	0.997537	0.00246334	0.948594	0.0514059
14	0.999962	3.8269e-005	0.644769	0.355231
15	0.440298	0.559702	0.917455	0.0825449
16	0.900347	0.0996526	1.27702e-006	0.999999
17	0.026708	0.973292	0.99785	0.00215049
18	0.998941	0.00105912	0.999847	0.000152993
19	0.990792	0.00920814	0.00934681	0.990653
20	0.0760604	0.92394	0.0472192	0.952781
21	0.00241231	0.997588	0.949796	0.0502045
22	0.99779	0.00221016	0.991331	0.00866932
23	0.999874	0.000125827	0.999796	0.000204374
24	0.999599	0.000401474	0.94524	0.0547599
25	0.999969	3.05873e-005	0.899075	0.100925
26	7.94432e-005	0.999921	0.000832	0.999168
27	7.19276e-005	0.999928	5.28473e-005	0.999947
28	0.995286	0.00471417	0.991166	0.00883354
29	0.99338	0.00661991	0.96933	0.0306704
30	0.994973	0.00502731	0.0367938	0.963206
31	0.540616	0.459384	0.758234	0.241766

Matrix - CLASSP_I

[70, 2] Cell: 0.995191

	1	2
1	0.995191	0.00480883
2	0.95822	0.0417803
3	0.297657	0.702343
4	0.95837	0.0416298
5	0.989919	0.0100807
6	0.939958	0.060042
7	0.932562	0.0674381
8	3.16698e-005	0.999968
9	0.00432411	0.995676
10	0.13614	0.86386
11	0.9384	0.0616
12	0.961189	0.0388114
13	0.74576	0.25424
14	8.86706e-008	1
15	6.22513e-006	0.999994
16	0.00120688	0.998793
17	0.00302912	0.996971
18	0.0938362	0.906164

Figure N32.1 Posterior Estimates of Parameters and Class Probabilities

N32.3 Applications

The following demonstrates the LCRP model with a fairly sparse specification. The data are actually a cross section, but for purpose of the example, we have grouped the observations into a panel of 70 sets of three. Nonetheless, the model appears to be overspecified for this data set. Nearly all of the improvement in the log likelihood function over the basic MNL results from the latent class specification.

We note, as emerges from estimation in this example, the LCRP model is somewhat volatile, and identification is a bit fragile.

```
LCRPLOGIT ; Lhs = mode
          ; Choices = air,train,bus,car
          ; Rhs = gc,invc,invt,casc
          ; Pds = 3
          ; Rpl ; Fcn = gc(n),invt(n) ; Draws = 500 ; Halton
          ; LCM ; Pts = 2
          ; Effects: gc(*) / invt(*) ; Full ; Par $
```

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable              Choice
Log likelihood function     -275.34264
Estimation based on N =    210, K =   4
Inf.Cr.AIC  =    558.7 AIC/N =    2.660
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .0297 .0092
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|     .02463*        .01347     1.83  .0674     -.00177    .05103
    INVT|    -.00580***      .00188    -3.08  .0020     -.00949   -.00211
    INVC|    -.04417***      .01525    -2.90  .0038     -.07406   -.01427
    CASC|    -.19710         .21268     -.93  .3541     -.61395    .21975
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Latent Class Mixed (RP) Logit Model
Dependent variable                MODE
Log likelihood function     -237.65976
Restricted log likelihood   -291.12182
Chi squared [  13 d.f.]      106.92411
Significance level              .00000
McFadden Pseudo R-squared     .1836415
Estimation based on N =    210, K =  13
Inf.Cr.AIC  =    501.3 AIC/N =    2.387
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients  -291.1218  .1836 .1664
Constants only   -283.7588  .1625 .1448
At start values  -275.3443  .1369 .1187
Response data are given as ind. choices
Replications for simulated probs. = 500
Halton sequences used for simulations
Number of latent classes =           2
Average Class Probabilities
     .611  .389
LCM model with panel has      70 groups
Fixed number of obsrvs./group=       3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Estimated latent class probabilities
 PrbCls1|     .61093***      .07954     7.68  .0000      .45503    .76683
 PrbCls2|     .38907***      .07954     4.89  .0000      .23317    .54497
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Random Parameters Logit Model for Class  1
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Random parameters in utility functions
      GC|      .03601          .02775     1.30  .1944     -.01838    .09040
    INVT|    -.00911**         .00398    -2.29  .0222     -.01691   -.00130
        |Nonrandom parameters in utility functions
    INVC|    -.11063***        .03446    -3.21  .0013     -.17818   -.04308
    CASC|   -1.09509***        .40728    -2.69  .0072    -1.89334   -.29683
        |Distns. of RPs. Std.Devs or limits of triangular
    NsGC| .61362D-05           .00559      .00  .9991 -.10958D-01  .10970D-01
  NsINVT| .65061D-06           .00054      .00  .9990 -.10587D-02  .10600D-02
--------+-----------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
Random Parameters Logit Model for Class  2
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Random parameters in utility functions
      GC|      .02826          .03926      .72  .4717     -.04869    .10520
    INVT|    -.00755           .00589    -1.28  .1998     -.01910    .00399
        |Nonrandom parameters in utility functions
    INVC|    -.02027           .04298     -.47  .6372     -.10450    .06396
    CASC|      .06251          .64507      .10  .9228    -1.20181   1.32682
        |Distns. of RPs. Std.Devs or limits of triangular
    NsGC| .19471D-04           .00694      .00  .9978 -.13590D-01  .13629D-01
  NsINVT| .54962D-05           .00065      .01  .9932 -.12610D-02  .12720D-02
--------+-----------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------

+---------------------------------------------------+
| Elasticity             averaged over observations.|
| Effects on probabilities of all choices in model: |
+---------------------------------------------------+
--------------------------------------------------------------------------------
Average elasticity      of prob(alt) wrt GC       in AIR
--------+-----------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  Choice|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
     AIR|    2.22541***        .09590    23.21  .0000     2.03745   2.41337
   TRAIN|    -.69945***        .05308   -13.18  .0000     -.80349   -.59541
     BUS|    -.71462***        .06642   -10.76  .0000     -.84479   -.58445
     CAR|   -1.01229***        .11398    -8.88  .0000    -1.23569   -.78889
--------+-----------------------------------------------------------------------
(Results omitted)
```

```
--------+--------------------------------------------------------------------
Average elasticity      of prob(alt) wrt INVT     in AIR
--------+--------------------------------------------------------------------
     AIR|    -.79152***       .03909   -20.25  .0000      -.86812    -.71491
   TRAIN|     .22241***       .01120    19.86  .0000       .20045     .24436
     BUS|     .21523***       .01507    14.28  .0000       .18570     .24477
     CAR|     .29078***       .02380    12.22  .0000       .24413     .33743
--------+--------------------------------------------------------------------
(Results omitted)
-----------------------------------------------------------------------------
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  2.2254   -.6994   -.7146  -1.0123
   TRAIN|  -.6559   2.7363   -.5195   -.6015
     BUS|  -.8410   -.7176   2.4200  -1.0315
     CAR|  -.7826   -.6792   -.8638   2.2465
--------+-----------------------------------
INVT    |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.7915    .2224    .2152    .2908
   TRAIN|   .8851  -3.2469    .7302    .8229
     BUS|  1.2095   1.0357  -3.3599   1.4583
     CAR|  1.1958   1.0281   1.3029  -3.4886

Clogit
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-----------------------------------
GC      |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  1.7153   -.8132   -.8132   -.8132
   TRAIN|  -.4983   2.7090   -.4983   -.4983
     BUS|  -.6247   -.6247   2.2145   -.6247
     CAR|  -.6214   -.6214   -.6214   1.7290
--------+-----------------------------------
INVT    |      AIR    TRAIN      BUS      CAR
--------+-----------------------------------
     AIR|  -.5202    .2554    .2554    .2554
   TRAIN|   .5678  -2.9605    .5678    .5678
     BUS|   .8017    .8017  -2.8494    .8017
     CAR|   .8742    .8742    .8742  -2.4506
```

This second application is based on a simulated data set in which the responses are a stated choice experiment with 8 repetitions based on a four choice setting, 3 unlabeled alternatives and 'none.' There are 400 observations grouped by three latent classes. Data consist of the choice outcome, data on two attributes, A and B, a price variable and its square, and demographics, sex and three age groups, young, middle, old. The model fit is a random effects latent class model;

$$U_{it,q}\,(type1) = \beta_{1,q}\,A(1)_{it} + \beta_{2,q}\,B(1)_{it} + \beta_{3,q}\,p(1)_{it} + \beta_{4,q}\,p(1)_{it}^2 + \gamma_q + \sigma_q w_{iq} + \varepsilon(1)_{it}$$
$$U_{it,q}\,(type2) = \beta_{1,q}\,A(2)_{it} + \beta_{2,q}\,B(2)_{it} + \beta_{3,q}\,p(2)_{it} + \beta_{4,q}\,p(2)_{it}^2 + \gamma_q + \sigma_q w_{iq} + \varepsilon(2)_{it}$$
$$U_{it,q}\,(type3) = \beta_{1,q}\,A(3)_{it} + \beta_{2,q}\,B(3)_{it} + \beta_{3,q}\,p(3)_{it} + \beta_{4,q}\,p(3)_{it}^2 + \gamma_q + \sigma_q w_{iq} + \varepsilon(3)_{it}$$
$$U_{it,q}\,(none) = \varepsilon(none)_{it}$$

The utility functions for the three non-null choices include a common random effect, w_{iq} which is time and choice invariant – this carries an unmeasured characteristic of the person. The command for the full model is

```
LCRPLOGIT  ; Lhs = choice
           ; Choices = type1,type2,type3,none
           ; Pds = 8
           ; Model: U(type*) = b1*attrA+b2*attrB+b3*price+b4*pricesq+c*type
           ; LCM = sex,young,middle ; Pts = 3
           ; Rpl ; Fcn = c(n) ; Draws = 25 ; Halton $
```

The model is refit with only the latent class specification by eliminating the random parameters specification and changing the model request:

```
LCLOGIT    ; Lhs = choice
           ; Choices = type1,type2,type3,none
           ; Pds = 8
           ; Model: U(type*) = b1*attrA+b2*attrB+b3*price+b4*pricesq+c*type
           ; LCM = sex,young,middle ; Pts = 3 $
```

The third specification is the random parameters (random effect) model obtained by eliminating the latent class request:

```
RPLOGIT    ; Lhs = choice
           ; Choices = type1,type2,type3,none
           ; Pds = 8
           ; Model: U(type*) = b1*attrA+b2*attrB+b3*price+b4*pricesq+c*type
           ; Rpl ; Fcn = c(n) ; Draws = 25 ; Halton $
```

The log likelihood functions for the four models are

-3648.66419	for the latent class random parameters model,
-3648.66560	for the latent class model,
-4145.12849	for the random effects model,
-4145.19725	for the multinomial logit model,

respectively. The implication is that almost no additional fit is obtained by adding the random parameters component to the latent class model, while nearly all of the additional fit over the multinomial logit model is added by the latent class model.

```
-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function      -4145.19725
Estimation based on N =   3200, K =   5
Inf.Cr.AIC  = 8300.395 AIC/N =    2.594
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only  -4391.1804  .0560 .0535
Response data are given as ind. choices
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
       C|    -.92843***      .19292    -4.81  .0000    -1.30655   -.55030
      B1|    1.46579***      .06746    21.73  .0000     1.33358   1.59800
      B2|    1.04267***      .06451    16.16  .0000      .91624   1.16909
      B3|    4.05938        3.23373     1.26  .2094    -2.27861  10.39737
      B4|   -61.0613***    12.11106    -5.04  .0000    -84.7985  -37.3240
--------+--------------------------------------------------------------------
-----------------------------------------------------------------------------
Latent Class Mixed (RP) Logit Model
Dependent variable                CHOICE
Log likelihood function      -3648.66419
Replications for simulated probs. =  25
Halton sequences used for simulations
Number of latent classes =             3
Average Class Probabilities
     .505  .237  .258
LCM model with panel has     400 groups
Fixed number of obsrvs./group=         8
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |This is THETA(01) in class probability model.
Constant|    -.93006**       .37552    -2.48  .0133    -1.66605   -.19406
   _SEX|1|     .66750*       .36297     1.84  .0659     -.04392   1.37891
 _YOUNG|1|    2.13774***     .32185     6.64  .0000     1.50694   2.76855
 _MIDDL|1|     .69623        .43518     1.60  .1096     -.15670   1.54917
        |This is THETA(02) in class probability model.
Constant|     .36431         .34476     1.06  .2906     -.31141   1.04004
   _SEX|2|   -2.78195***     .69797    -3.99  .0001    -4.14995  -1.41394
 _YOUNG|2|    -.14938        .54763     -.27  .7850    -1.22272    .92397
 _MIDDL|2|    1.96666***     .71585     2.75  .0060      .56361   3.36971
        |This is THETA(03) in class probability model.
Constant|        0.0    .....(Fixed Parameter).....
   _SEX|3|        0.0    .....(Fixed Parameter).....
 _YOUNG|3|        0.0    .....(Fixed Parameter).....
 _MIDDL|3|        0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Parameters Logit Model for Class  1
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
       C|   -1.44332***      .38488    -3.75  .0002    -2.19768   -.68897
        |Nonrandom parameters in utility functions
      B1|    3.01430***      .14702    20.50  .0000     2.72614   3.30246
      B2|    -.07439         .12736     -.58  .5591     -.32402    .17523
      B3|   -6.94557        6.48173    -1.07  .2839   -19.64952   5.75838
      B4|   -10.3168       23.80017     -.43  .6647    -56.9643   36.3307
        |Distns. of RPs. Std.Devs or limits of triangular
     NsC|     .00015         .05506      .00  .9978     -.10777    .10807
--------+--------------------------------------------------------------------
Random Parameters Logit Model for Class  2
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
       C|     .74693*        .39677     1.88  .0598     -.03072   1.52458
        |Nonrandom parameters in utility functions
      B1|    1.22106***      .16382     7.45  .0000      .89997   1.54214
      B2|    1.10763***      .16489     6.72  .0000      .78445   1.43081
      B3|   -19.8414***     6.85353    -2.90  .0038    -33.2741   -6.4088
      B4|    22.6733       25.13052      .90  .3669    -26.5816   71.9282
        |Distns. of RPs. Std.Devs or limits of triangular
     NsC|     .00322         .08544      .04  .9700     -.16423    .17067
--------+--------------------------------------------------------------------
-----------------------------------------------------------------------------
Random Parameters Logit Model for Class  3
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
       C|    -.29439         .41625     -.71  .4794    -1.11023    .52146
        |Nonrandom parameters in utility functions
      B1|    -.16334         .16638     -.98  .3263     -.48945    .16277
      B2|    2.70227***      .18006    15.01  .0000     2.34935   3.05519
      B3|   -6.86567        7.42419     -.92  .3551   -21.41681   7.68547
      B4|   -8.26246       27.65433     -.30  .7651   -62.46394  45.93903
        |Distns. of RPs. Std.Devs or limits of triangular
     NsC|     .00075         .09731      .01  .9938     -.18996    .19147
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Latent Class Logit Model
Dependent variable                CHOICE
Log likelihood function      -3648.66560
Number of latent classes =             3
Average Class Probabilities
     .505  .237  .258
LCM model with panel has     400 groups
Fixed number of obsrvs./group=         8
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Utility parameters in latent class -->> 1
    B1|1|    3.01443***      .14704    20.50  .0000     2.72624   3.30262
    B2|1|    -.07457         .12737     -.59  .5582     -.32421    .17507
    B3|1|   -6.97451        6.48123    -1.08  .2819   -19.67748   5.72847
    B4|1|   -10.2076       23.79811     -.43  .6680    -56.8510   36.4358
     C|1|   -1.44175***      .38483    -3.75  .0002    -2.19599   -.68750
        |Utility parameters in latent class -->> 2
    B1|2|    1.22082***      .16381     7.45  .0000      .89975   1.54188
    B2|2|    1.10766***      .16487     6.72  .0000      .78452   1.43080
    B3|2|   -19.7732***     6.85471    -2.88  .0039    -33.2082   -6.3382
    B4|2|    22.4120       25.13694      .89  .3726    -26.8555   71.6795
     C|2|     .74323*        .39686     1.87  .0611     -.03460   1.52106
        |Utility parameters in latent class -->> 3
    B1|3|    -.16351         .16641     -.98  .3258     -.48966    .16265
    B2|3|    2.70297***      .18014    15.00  .0000     2.34990   3.05604
    B3|3|   -6.95426        7.42439     -.94  .3489   -21.50580   7.59729
    B4|3|   -7.92518       27.65361     -.29  .7744   -62.12525  46.27489
     C|3|    -.28994         .41617     -.70  .4860    -1.10561    .52573
        |This is THETA(01) in class probability model.
Constant|    -.92984**       .37555    -2.48  .0133    -1.66590   -.19379
  _SEX|1|     .66719*        .36300     1.84  .0661     -.04429   1.37866
_YOUNG|1|    2.13778***      .32185     6.64  .0000     1.50697   2.76859
_MIDDL|1|     .69660         .43521     1.60  .1095     -.15639   1.54960
        |This is THETA(02) in class probability model.
Constant|     .36443         .34484     1.06  .2906     -.31144   1.04029
  _SEX|2|   -2.78223***      .69819    -3.98  .0001    -4.15065  -1.41380
_YOUNG|2|    -.14880         .54764     -.27  .7858    -1.22215    .92455
_MIDDL|2|    1.96741***      .71611     2.75  .0060      .56385   3.37096
        |This is THETA(03) in class probability model.
Constant|        0.0    .....(Fixed Parameter).....
  _SEX|3|        0.0    .....(Fixed Parameter).....
_YOUNG|3|        0.0    .....(Fixed Parameter).....
_MIDDL|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.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                CHOICE
Log likelihood function      -4145.12489
Restricted log likelihood    -4436.14196
Chi squared [   6 d.f.]        582.03414
Significance level               .00000
McFadden Pseudo R-squared      .0656014
Estimation based on N =   3200, K =   6
Inf.Cr.AIC  = 8302.250 AIC/N =    2.594
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients -4436.1420  .0656 .0650
Constants only  -4391.1804  .0560 .0554
At start values -4145.1973  .0000-.0006
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has     400 groups
Fixed number of obsrvs./group=        8
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
       C|    -.92698***      .19316    -4.80  .0000    -1.30556   -.54840
        |Nonrandom parameters in utility functions
      B1|    1.46610***      .06747    21.73  .0000     1.33386   1.59835
      B2|    1.04286***      .06451    16.16  .0000      .91641   1.16930
      B3|    4.06333        3.23406     1.26  .2090    -2.27530  10.40196
      B4|   -61.0922***    12.11305    -5.04  .0000    -84.8334  -37.3511
        |Distns. of RPs. Std.Devs or limits of triangular
     NsC|      .10674        .19459      .55  .5833     -.27464    .48812
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

N32.4 Technical Details

The latent class model assumes that parameter vectors, $\boldsymbol{\beta}_i$ are distributed among individuals with a discrete distribution, rather than the continuous distribution that lies behind the mixed logit model. It is assumed that the population consists of a finite number, Q, of groups of individuals. The groups are heterogeneous, with common parameters, $\boldsymbol{\beta}_q$, for the members of the group, but the groups themselves are different from one another. The classes are distinguished by the different parameter vectors, though the fundamental data generating process, the probability density for the interesting variable under study, is the same.

The full specification of the latent class structure for a generic data generating process is

$$Prob(choice_i|\mathbf{x}_i,\text{class} = q) = g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_q)$$

$$Prob(\text{class} = q) = \pi_q(\boldsymbol{\theta}),\ q = 1,...,Q.$$

The unconditional probability attached to an observation is obtained by integrating out the heterogeneity due to the distribution across classes,

$$Prob(choice_i|\mathbf{x}_i) = \Sigma_q\, \pi_q(\boldsymbol{\theta})g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_q).$$

The latent class random parameters model allows for heterogeneity both within and across the classes. To accommodate the two layers of heterogeneity, we allow for continuous variation of the parameters within classes. The latent class aspect of the model is

$$Prob(choice_i|\mathbf{x}_i,\text{class} = q) = g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_{i|q})$$

$$Prob(\text{class} = q) = \pi_q(\boldsymbol{\theta}),\ q = 1,...,Q.$$

This is the model developed in Chapter N25. The within-class heterogeneity is structured as set up in Chapter N29 for the random parameters model,

$$\boldsymbol{\beta}_{i|q} = \boldsymbol{\beta}_q + \mathbf{w}_{i|q}$$

$$\mathbf{w}_{i|q} \sim E[\mathbf{w}_{i|q}|\mathbf{X}] = \mathbf{0},\ \text{Var}[\mathbf{w}_{i|q} \mid \mathbf{X}] = \boldsymbol{\Sigma}_q$$

where the $\mathbf{X}$ indicates that $\mathbf{w}_{i|q}$ is uncorrelated with all exogenous data in the sample.

We will assume below that the underlying distribution for the within-class heterogeneity has mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. In a given application, it may be appropriate to further assume that certain rows and corresponding columns of $\boldsymbol{\Sigma}_q$ equal zero, indicating that the variation of the corresponding parameter is entirely across classes.

The contribution of individual *i* to the log likelihood for the model is obtained for each individual in the sample by integrating out the within-class heterogeneity and then the class heterogeneity. We allow for a panel data setting, hence the observed vector of outcomes is denoted $\mathbf{y}_i$ and the observed data on exogenous variables are collected in $\mathbf{X}_i = [\mathbf{X}_{i1},..,\mathbf{X}_{iT_i}]$. An individual is assumed to engage in T_i choice situations, where $T_i \geq 1$. The generic model is

$$Prob(choice_i|\mathbf{X}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_Q,\boldsymbol{\theta},\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_Q) = \sum_{q=1}^{Q} \pi_q(\boldsymbol{\theta})\int_{\mathbf{w}_i} \prod_{t=1}^{T_i} f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}]h(\mathbf{w}_i \mid \boldsymbol{\Sigma}_q)d\mathbf{w}_i)$$

The class probabilities are parameterizes using a multinomial logit formulation to impose the adding up and positivity restrictions on $\pi_q(\boldsymbol{\theta})$. Thus,

$$\pi_q(\boldsymbol{\theta}) = \frac{\exp(\theta_q)}{\Sigma_{q=1}^{Q}\exp(\theta_q)},\ q = 1,...,Q;\ \theta_Q = 0.$$

A useful refinement of the class probabilities model is to allow the probabilities to be dependent on individual data, such as demographics. The class probability model becomes

$$\pi_{iq}(\mathbf{z}_i, \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)}{\Sigma_{q=1}^{Q}\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)},\ q = 1,...,Q;\ \boldsymbol{\theta}_Q = \mathbf{0}.$$

The resulting model employed in this application is a latent class, random parameters multinomial logit (LCRPLOGIT) model. Individual i chooses among J alternatives with conditional probabilities

$$f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}] = \frac{\exp[\Sigma_{j=1}^{J} y_{it,j} (\boldsymbol{\beta}_q + \mathbf{w}_i)' \mathbf{x}_{it,j}]}{\sum_{j=1}^{J} \exp[\Sigma_{j=1}^{J} y_{it,j} (\boldsymbol{\beta}_q + \mathbf{w}_i)' \mathbf{x}_{it,j}]}, \; j = 1,...,J,$$

$y_{it,j} = 1$ for the j corresponding to the alternative chosen and 0 for all others, and $\mathbf{x}_{it,j}$ is the vector of attributes of alternative j for individual i in choice situation t.

We use maximum simulated likelihood to evaluate the terms in the log likelihood expression. The contribution of individual i to the simulated log likelihood is

$$f^S(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_Q,\boldsymbol{\theta},\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_Q) \;=\; \sum_{q=1}^{Q} \pi_{iq}(\mathbf{z}_i, \boldsymbol{\theta}) \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_{i,r}), \mathbf{X}_{it}]$$

$\mathbf{w}_{i,r}$ is the rth of R random draws on the random vector $\mathbf{w}_i$. Collecting all terms, the simulated log likelihood is

$$\log L^S = \sum_{i=1}^{N} \log\left[\sum_{q=1}^{Q} \pi_{iq}(\mathbf{z}_i, \boldsymbol{\theta}) \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_{i,r}), \mathbf{X}_{it}]\right].$$

N33: Generalized Mixed Logit Model

N33.1 Introduction

The generalized mixed logit model developed by Fiebig, Keane, Louviere and Wasi (2010) encompasses several of the models shown in the earlier chapters. We have added several additional layers to the specification in our implementation. The essential format in their paper is a random parameters logit model,

$$U_{it}(j) = \boldsymbol{\beta}_i'\mathbf{x}_{it,j} + \varepsilon_{it,j};$$

$$\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\Gamma}\mathbf{w}_i,\ \mathbf{w}_\mathrm{i} \sim N[\mathbf{0},\mathbf{I}],\ 0 \leq \gamma \leq 1;$$

$$\sigma_i = \exp(-\tau^2/2 + \tau v_i),\ \ v_i \sim N[0,1].$$

The central form is the multinomial logit model based on the extreme value distribution of $\varepsilon_{it,j}$. The general form combines the scaled MNL (Chapter N24) with the random parameters model of Chapter N29. The random scaling factor, σ_i has mean 1 and variance $\exp(\tau^2 - 1)$. There are several interesting special cases:

- $\tau = 0$ implies the random parameters model, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_i$
- $\gamma = 0$ implies a scaled random parameters logit model, $\boldsymbol{\beta}_i = \sigma_i[\boldsymbol{\beta} + \sigma_i\boldsymbol{\Gamma}\mathbf{w}_i]$
- $\gamma = 1$ implies a hybrid model, $\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_\mathrm{i}$

Note that τ is crucial to the model formulation; if τ equals zero, γ is not identified (i.e., not estimable).

- $\tau = 0$, $\boldsymbol{\Gamma} = \mathbf{0}$ implies the original multinomial logit model with $\sigma_i = 0$
- $\gamma = 0$, $\boldsymbol{\Gamma} = \mathbf{0}$ implies the scaled multinomial logit model, $\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta}$

The model provided in *NLOGIT* extends this formulation with the following:

- The means of $\boldsymbol{\beta}_i$ may be heterogeneous, $\sigma_i[\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i] + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\Gamma}\mathbf{w}_i$
- σ_i may be heterogeneous based on observed data, $\sigma_i = \exp(-\tau^2/2 + \boldsymbol{\delta}'\mathbf{r}_i + \tau v_i)$ (see Fiebig et al. equation (12))
- The full menu of distributions is available for $\mathbf{w}_i$
- The implementation accommodates panel data (e.g., stated choice data).

N33.2 Commands

The minimal form of the command for the generalized mixed logit model is

GMXLOGIT ; Lhs = choice
; Choices = list ...
; Rhs and ; Rh2 to set up the utility functions or
; Model: sets up the utility functions
; Fcn = the usual specification of the random parameters $

The **; Fcn = specification** sets up the random parameters exactly as shown in Chapter N29 for the random parameters logit model. The GMXLOGIT model adds nonzero γ and τ to the random parameters model. Note, again, if $\tau = 0$, the model reverts to the original mixed logit model; γ is not estimable (does not exist) if $\tau = 0$. All forms of the random parameters model are available with GMXLOGIT. However, some specifications will act unpredictably when γ is nonzero. There are numerous options for modifying the GMXLOGIT model. Two important overall settings are

; Pds = number of choice situations, if more than 1

and **; Pts = number of points in simulation**
; Halton if desired.

The random parameters formulation to this point is

$$\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + [\gamma + \sigma_i(1 - \gamma)]\boldsymbol{\Gamma}\mathbf{w}_i.$$

This is the essential model, though it adds a bit to what is in Fiebig et al.'s paper. Note, for example, they have to apply the treatment to the entire parameter vector, while the preceding applies it to the parameters that you specify. They have a figure on page 31 with the various special cases. The spec above is for G-MNL at the top of the page. Use

; GMX = list of $\mathbf{z}_i$ variables in the mean

to produce $$\boldsymbol{\beta}_i = \sigma_i[\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i] + [\gamma + \sigma_i(1 - \gamma)]\boldsymbol{\Gamma}\mathbf{w}_i$$

The random parameters are assumed to be uncorrelated - $\boldsymbol{\Gamma}$ is a diagonal matrix. This assumption is relaxed by adding

; Correlation

to the command. With this in place, $\boldsymbol{\Gamma}$ is now a lower triangular matrix. Further restrictions on the correlations are described in Section N29.3.6. It is generally assumed that the heterogeneity in $\boldsymbol{\beta}_i$, that is $\mathbf{w}_i$, is uncorrelated with v_i, the heterogeneity in σ_i. This restriction is relaxed by adding

; SCV

to the model command. This adds a new set of parameters to the model, $\boldsymbol{\lambda} = \text{Cov}(\mathbf{w}_i, v_i)$. (We note, though the program does allow this, as a specification, it is probably a bad choice. In our experience, the estimator becomes rather unstable with this feature enabled.) Finally, heteroscedasticity can be introduced into σ_i with

; Hfr = list of $\mathbf{r}_i$ variables.

With this specification, the model becomes

$$\sigma_i = \exp(-\tau^2/2 + \boldsymbol{\delta}'\mathbf{r}_i + \tau v_i)$$

(This extension of the model is proposed in passing in equation (12) in Fiebig et al.)

N33.2.1 Controlling the GMXLOGIT Parameters

The GMXLOGIT model is created by the two parameters γ and τ. As noted, the model reverts to the random parameters model if τ equals zero. You can preset the values of these parameters as follows:

; Gamma = [value]

fixes γ at the value. Generally interesting values are 0 and 1, but any value from 0 to 1, inclusive may be specified. If you omit the square brackets, then the value is simply used as the starting value for the iterations. Referring to the figure on page 6 in Fiebig et al., the two interesting special cases here are

; Gamma = [1] produces G-MNL-I
; Gamma = [0] produces G-MNL-II

Any other value between 0 and 1 may be specified. The parameter τ is also controlled the same way. Use

; Tau = [value]

to fix τ. The $\tau = 0$ case, which implies $\gamma = 0$, produces Fiebig et al.'s MIXL variant of the model. With $\tau = 0$, the resulting value is the mixed (random parameters) logit model. We note one caution. If $\tau = 0$, then γ is not estimable. But, the command processor will allow you to specify a model in which τ equals zero but γ is a free parameter. The iterations will proceed, and *NLOGIT* may even claim convergence. However, because γ is not identified when $\tau = 0$, changes in γ will not change the log likelihood function. The end result is that the second derivatives matrix will be singular – the estimator will quit with a warning such as

```
Error  1027: Models - estimated variance matrix of estimates is singular.
```

N33.2.2 The Scaled MNL Model

The scaled MNL model of Chapter N24 results when no parameters are random and $\gamma = 0$. You should use the **SMNLOGIT** command to produce this case. It also results if you use the **GMXLOGIT** command with

; SMNL.

In this case, $\gamma = 0$ is imposed automatically. If you have specified any random parameters, they will be set as 'constant' parameters, that is type (C). This wastes computing time, however, as forcing parameters to be 'constant' forces the variance to be zero. It does not prevent the generation of the random draws. When a parameter is specified as type [C], then

$$\beta_i = \sigma_i[\beta + 0\, v_i]$$

The v_i is still drawn. Using **; SMNL** in this fashion produces the preceding type of (non)random parameter.

You cannot use a constrained distribution like (O) with **; SMNL**; (O) sets up a parameter in which the variance parameter is the same as the mean. It implies that $\beta(i) = \beta + \beta^* v(i)$. But, to set up the estimator internally, the second β is treated as a separate σ that equals β. It then becomes impossible to force σ to equal zero without forcing β to equal zero as well. You should not do this. The long and short of it is that (O) is incompatible with the scaled MNL model.

The scaled MNL model is one in which the only random parameter is σ_i. In general, you should use the major command, **SMNLOGIT** to fit this model, not **GMXLOGIT** with restrictions.

N33.2.3 Alternative Specific Constants

Fiebig et al. notes that the ASCs in the model produce special estimation problems. There are three possible strategies:

1. Leave them out of the FCN specification so they are fixed parameters.

2. Include them in the FCN specification and make them a part of the general model. Fiebig et al. observes that this frequently causes the estimator to fail.

3. Include them among the random parameters, but for them, force no special scaling, so the σ_i scaling parameter equals 1 and gamma = 0 for these parameters.

Case (3) requires a complicated special treatment internally. You can request this treatment by including ASCs using **; Rh2 = one**, then adding the following specification to the command,

; RPASC.

N33.2.4 Heteroscedasticity

The generalized mixed logit model preserves all of the elaborate models for the RPLOGIT case described in Chapter N29 except the heteroscedasticity model described in Section N29.4. In the random parameters model, there may be a separate $\theta_{ik} = \sigma_k \times \exp(\boldsymbol{\delta}'\mathbf{r}_i)$ for each random parameter. That model is no longer identified in the presence of σ_i in this model, so the heteroscedasticity that is supported resides completely in σ_i. See the definition of **; Hfr = list** above.

N33.3 Estimation in Willingness to Pay Space

A common calculation in choice modeling is the willingness to pay measure,

$$WTP = \textit{Marginal utility of attribute / Marginal utility of income}$$

or sometimes

$$WTP = -1 \times \textit{Marginal utility of attribute / Marginal disutility of cost}.$$

The empirical estimator is typically a ratio of coefficients,

$$WTP = b_{attribute} \;/\; b_{income\ or\ cost}.$$

Scarpa, Thiene and Train (2008) and Daly, Hess and Train (2012) argue (persuasively) that ratios of coefficients generally have infinite variances for most distributions of econometric estimators. Hence, WTP estimators such as the above do not have finite moments. In the case of the MNL, the implication is that the constant WTP estimate, which is the ratio of two asymptotically normal estimators, does not, itself, have a finite variance. The problem reappears in mixed logit models. Researchers often specify RP models so that the denominator in the WTP calculation is a nonrandom parameter. The resulting estimator takes the form

$$WTP(i) = -\, b_{attribute}(i) \;/\; b_{cost}.$$

However, this does not actually solve the problem. The distribution of the ratio is still problematic.

A reformulation of the utility function in the choice model suggests a solution. The model in ‘preference space’ is

$$U = \beta_{cost}\; cost + \beta_{x1}\, x1 + \beta_{x2}\, x2 + \varepsilon$$

where $x1$ and $x2$ are two attributes. The WTP computation is β_{x1}/β_{cost}. Using the familiar econometric estimates produces the problems noted earlier. The function can be trivially rewritten is

$$U = \beta_{cost}\,[cost + (\beta_{x1}/\beta_{cost})X1 + (\beta_{x2}/\beta_{cost})X2] + \varepsilon$$

$$U = \beta_{cost}\,[cost + \theta_1\, x1 + \theta_2\, x2] + \varepsilon.$$

For an MNL model, this is a trivial reformulation. It does create a nonlinearity in the model that was not there previously. However, the MLEs of the parameters will be identical because of the invariance of the MLE to a one to one transformation. This invariance does not carry over to a random parameters formulation. If θ_1 and θ_2 are random parameters, the results are not invariant to the transformation. This model in ‘WTP space,’

$$U = \beta_{cost}\,[cost + \theta_{1,i}\, x1 + \theta_{2,i}\, x2] + \varepsilon.$$

The $\theta_{k,i}$ parameters are already willingness to pay estimates.

You can choose one of the parameters in the GMXLOGIT model to have a coefficient of one and build this nonlinearity into the GMXLOGIT model by changing its type to (*type) in the **; Fcn = (*type)**,… specification. (Note, this device only works in the GMXLOGIT model. It doesn't work in the RPLOGIT or SMNLOGIT models.) The model results will appear as in the following contrived example:

GMXLOGIT ; Lhs = mode ; Choices = air,train,bus,car
; Rhs = gc,invt,invc ; Rh2 = one
; Fcn = gc(*n), invc(n) $

```
-----------------------------------------------------------------------------
Generalized Mixed (RP) Logit Model
Dependent variable               MODE
Log likelihood function    -277.71292
Response data are given as ind. choices
Replications for simulated probs. = 150
Halton sequences used for simulations
RPL model with panel has      70 groups
Fixed number of obsrvs./group=      3
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    MODE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      GC|        1.0    .....(Fixed Parameter).....
    INVC|    1.00030***      .00269   372.40  .0000      .99504   1.00557
        |Nonrandom parameters in utility functions
    INVT|   -1.72722***      .00116 -1486.90  .0000    -1.72950  -1.72494
   A_AIR|    -.00770         .43002     -.02  .9857     -.85053    .83512
 A_TRAIN|    -.53569***      .16878    -3.17  .0015     -.86649   -.20489
   A_BUS|     .80344***      .21489     3.74  .0002      .38227   1.22462
        |Distns. of RPs. Std.Devs or limits of triangular
    CsGC|        0.0    .....(Fixed Parameter).....
  NsINVC|     .15888***      .00220    72.18  .0000      .15456    .16319
        |Variance parameter tau in GMX scale parameter
TauScale|     .99961***      .00071  1403.83  .0000      .99822   1.00101
        |Weighting parameter gamma in GMX model
GammaMXL|        0.0    .....(Fixed Parameter).....
        |Coefficient on GC       in preference space form
Beta0WTP|   -2.30273***      .00522  -441.07  .0000    -2.31296  -2.29250
S_b0_WTP|    194.638***      .30603   636.02  .0000     194.038   195.237
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     .94743        1.03071      .92  .3580    -1.07272   2.96759
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

In the original specification, the coefficient *GC* is random, with type (*N). With the WTP specification, the coefficient on *GC* is forced to equal 1.0 and its standard deviation, *NsGC* equals zero. The new random parameter created to replace *GC* is *Beta0WTP* with standard deviation parameter *S_b0_WTP*.

N33.4 Results

Estimation results for the generalized mixed logit model are all the same as for the random parameters logit (RPLOGIT) model of Chapter N29. The two additional parameters, the estimators of γ and τ are reported with the other results. To illustrate, we will use the data used in the second application in Section N32.3. A generic GMXLOGIT model is estimated with

GMXLOGIT ; Lhs = choice ; Choices = type1,type2,type3,none ; Pds = 8
; Model: U(type*) = attra*attra + attrb*attrb
+ price*price + picktype*picktype
; Fcn = price(n),attra(n),attrb(n)
; Draws = 25 ; Halton ; Maxit = 20 $

There is an additional estimate reported in the results, *sigma*(*i*). This is not an additional parameter estimate. The results report the sample average of the computed values of σ_i. This is computed by averaging over the random draws for each individual then averaging across the individuals. The sample standard deviation reported is the standard deviation of the averages for the individuals.

```
-----------------------------------------------------------------------------
Generalized Mixed (RP) Logit Model
Dependent variable               CHOICE
Log likelihood function     -3917.16748
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has     400 groups
Fixed number of obsrvs./group=        8
Hessian is not PD. Using BHHH estimator
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  CHOICE|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
   PRICE|   -12.5275***     1.07357   -11.67  .0000    -14.6316   -10.4233
   ATTRA|    1.70406***      .12078    14.11  .0000     1.46732    1.94079
   ATTRB|     .92087***      .13659     6.74  .0000      .65316    1.18857
        |Nonrandom parameters in utility functions
PICKTYPE|    -.17822**       .07729    -2.31  .0211     -.32970    -.02675
        |Distns. of RPs. Std.Devs or limits of triangular
 NsPRICE|     .24302        1.29768      .19  .8514    -2.30038    2.78642
 NsATTRA|     .90498***      .08434    10.73  .0000      .73967    1.07029
 NsATTRB|    1.45085***      .10262    14.14  .0000     1.24973    1.65198
        |Variance parameter tau in GMX scale parameter
TauScale|     .46618***      .10750     4.34  .0000      .25548     .67688
        |Weighting parameter gamma in GMX model
GammaMXL|     .99999***      .20332     4.92  .0000      .60149    1.39849
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     .99133**       .47261     2.10  .0359      .06503    1.91763
```

This second example adds heterogeneity in the means to the previous model.

GMXLOGIT ; Lhs = choice ; Choices = type1,type2,type3,none ; Pds = 8
; Model: U(type*) = attra*attra + attrb*attrb
+ price*price + picktype*picktype
; Fcn = price(n),attra(n),attrb(n)
; GMX = sex,young,middle
; Draws = 25 ; Halton ; Maxit = 20 $

```
-----------------------------------------------------------------------------
Generalized Mixed (RP) Logit Model
Dependent variable               CHOICE
Log likelihood function     -3861.36607
Response data are given as ind. choices
Replications for simulated probs. =  25
Halton sequences used for simulations
RPL model with panel has     400 groups
Fixed number of obsrvs./group=         8
Hessian is not PD. Using BHHH estimator
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
   PRICE|   -10.0593***     1.72460    -5.83  .0000    -13.4395   -6.6791
   ATTRA|      .82432***     .20778     3.97  .0001      .41708   1.23156
   ATTRB|     1.63496***     .27831     5.87  .0000     1.08949   2.18044
        |Nonrandom parameters in utility functions
PICKTYPE|     -.19048**      .07649    -2.49  .0128     -.34040   -.04057
        |Heterogeneity in mean, Parameter:Variable
PRIC:SEX|   -1.07261        1.71500     -.63  .5317    -4.43394   2.28872
PRIC:YOU|   -3.24790        1.99533    -1.63  .1036    -7.15867    .66286
PRIC:MID|   -1.81243        2.22591     -.81  .4155    -6.17514   2.55027
ATTR:SEX|     .28880         .21653     1.33  .1823     -.13560    .71320
ATTR:YOU|    1.26832***      .26417     4.80  .0000      .75055   1.78609
ATTR:MID|     .55050**       .28011     1.97  .0494      .00149   1.09950
ATT0:SEX|    -.23672         .26569     -.89  .3729     -.75746    .28401
ATT0:YOU|   -1.08491***      .31233    -3.47  .0005    -1.69706   -.47276
ATT0:MID|    -.32916         .34939     -.94  .3461    -1.01395    .35564
        |Distns. of RPs. Std.Devs or limits of triangular
 NsPRICE|     .10201        1.33009      .08  .9389    -2.50491   2.70893
 NsATTRA|     .79072***      .08142     9.71  .0000      .63113    .95030
 NsATTRB|    1.25714***      .09500    13.23  .0000     1.07095   1.44334
        |Variance parameter tau in GMX scale parameter
TauScale|     .40424***      .09133     4.43  .0000      .22523    .58324
        |Weighting parameter gamma in GMX model
GammaMXL|     .99416***      .21229     4.68  .0000      .57807   1.41025
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     .99291**       .40709     2.44  .0147      .19503   1.79079
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

This final example estimates the preceding model in WTP space rather than preference space. The only change in the command is the addition of the '*' in the function definition of price.

GMXLOGIT ; Lhs = choice ; Choices = type1,type2,type3,none ; Pds = 8
; Model: U(type*) = attra*attra + attrb*attrb
+ price*price + picktype*picktype
; Fcn = price(*n),attra(n),attrb(n)
; GMX = sex,young,middle
; Draws = 25 ; Halton ; Maxit = 20 $

```
-----------------------------------------------------------------------------
Generalized Mixed (RP) Logit Model
Dependent variable               CHOICE
Log likelihood function     -3914.92659
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
   PRICE|        1.0    .....(Fixed Parameter).....
   ATTRA|    -.13179         .40820     -.32  .7468     -.93185    .66828
   ATTRB|   -1.64113        4.81765     -.34  .7334   -11.08356   7.80130
        |Nonrandom parameters in utility functions
PICKTYPE|    -.30699***      .07670    -4.00  .0001     -.45731   -.15666
        |Heterogeneity in mean, Parameter:Variable
PRIC:SEX|    17.7585       55.63561      .32  .7496    -91.2853  126.8023
PRIC:YOU|    20.4019       63.79981      .32  .7491   -104.6435  145.4472
PRIC:MID|    19.0938       60.07365      .32  .7506    -98.6484  136.8360
ATTR:SEX|   -2.10648        6.54684     -.32  .7476   -14.93805  10.72509
ATTR:YOU|   -3.36057       10.35383     -.32  .7455   -23.65370  16.93257
ATTR:MID|   -2.04350        6.39039     -.32  .7491   -14.56843  10.48144
ATT0:SEX|    -.54985        1.85642     -.30  .7671    -4.18836   3.08866
ATT0:YOU|    -.05851         .63358     -.09  .9264    -1.30030   1.18327
ATT0:MID|    -.81671        2.73296     -.30  .7651    -6.17321   4.53980
        |Distns. of RPs. Std.Devs or limits of triangular
 CsPRICE|        0.0    .....(Fixed Parameter).....
 NsATTRA|    2.23589        6.78855      .33  .7419   -11.06942  15.54120
 NsATTRB|    3.15281        9.59604      .33  .7425   -15.65508  21.96071
        |Variance parameter tau in GMX scale parameter
TauScale|     .33740***      .06723     5.02  .0000      .20563    .46916
        |Weighting parameter gamma in GMX model
GammaMXL|        0.0    .....(Fixed Parameter).....
        |Coefficient on PRICE    in preference space form
Beta0WTP|    -.44498        1.35338     -.33  .7423    -3.09757   2.20760
S_b0_WTP|     .00028         .03696      .01  .9939     -.07217    .07273
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|     .99441***      .33780     2.94  .0032      .33233   1.65650
--------+--------------------------------------------------------------------
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.
-----------------------------------------------------------------------------
```

N34: Diagnostics and Error Messages

N34.1 Introduction

The following is a complete list of diagnostics that will be issued by *NLOGIT*. Altogether, there are over 1,000 specific conditions that are picked up by the command translation and computation programs in *LIMDEP* and *NLOGIT*. Those listed here are specific to *NLOGIT*. The full set of diagnostics is given in Chapter R28. Nearly all of the error messages listed below identify problems in commands that you have provided for the command translator to parse and then to pass on to the computation programs.

Most diagnostics are self explanatory and will be obvious. For example,

```
82  ;LHS - variable in list is not in the variable names table.
```

states that your Lhs variable in a model command does not exist. No doubt this is due to a typographical error – the name is misspelled. Other diagnostics are more complicated, and in many cases, it is not quite possible to be precise about the error. Thus, in many cases, a diagnostic will say something like 'the following string contains an unidentified name' and a part of your command will be listed – the implication is that the error is somewhere in the listed string. Finally, some diagnostics are based on information that is specific to a variable or an observation at the point at which it occurs. In that case, the diagnostic may identify a particular observation or value. In the listing below, we use the conventions:

<AAAAAAAA>	indicates a variable name that will appear in the diagnostic,
<nnnnnnnnnnn>	indicates an integer value, often an observation number, that is given,
<xxxxxxxxxxxx>	indicates a specific value that may be invalid, such as a 'time' that is negative.

The listing below contains the diagnostics and, in some cases, additional points that may help you to find and/or fix the problem. The actual diagnostic you will see in your output window is shown in the Courier font, such as appears in diagnostic 82 above.

We note it should be extremely rare, but occasionally, an error message will occur for reasons that are not really related to the computation in progress. (We cannot give an example – if we knew where it was, we would remove the source before it occurred.) You will always know exactly what command produces a diagnostic – an echo of that command will appear directly above the error message in the output window. So, if an absolutely unfathomable error message shows up, try simplifying the command that precedes it to its bare essentials, and by building it up, reveal the source of the problem.

Finally, there are the 'program crashes.' Obviously, we hope that these never occur, but they do. The usual ones are division by zero and exponent overflow. Once again, we cannot give specific warnings about these, since if we could, we would fix the problem. If you do get one of these and you cannot get around it, please contact us at support@nlogit.com.

N34.2 Discrete Choice (CLOGIT) and NLOGIT

1000 `FIML/NLogit is not enabled in this program.`

1001 `Syntax problem in tree spec or expected ; or $ not found.`

1002 `Model defines too many alternatives (more than 100).`

1003 `A choice label appears more than once in the tree specification.`

1004 `Number of observations not a multiple of # of alternatives.`
This is expected when you have a fixed choice set.

1005 `Problem reading labels, or weights for choice based sample.`

1006 `Number of weights given does not match number of alternatives.`

1007 `A choice based sampling weight given is not between zero and one.`

1008 `The choice based sampling weights given do not sum to one.`

1009 `Expected [ in limb specification was not found.`

1010 `Expected ( in branch specification was not found.`

1011 `A branch label appears more than once in the tree.`

1012 `A choice label in a branch spec. is not in ;CHOICES list.`

1013 `Branch specifications are not separated by commas.`

1014 `One or more ;CHOICE labels does not appear in the tree.`

1015 `One or more ;CHOICE labels appears more than once in tree.`

1016 `The model must have either 1 or 3 LHS variables. Check spec.`

1017 `Nested logit model must include ;MODEL:... or ;RHS spec.`
`Found neither Model: nor RhS/Rh2.`
Your model specification is incomplete.

1018 `There is an unidentified variable name in the equation.`
In the ; Model: U (...) part of the command, one of your specified utility functions contains a variable name that is not in your data set.

1019 `Model specification exceeds an internal limit. See documentation.`
`RANK data can only be used for 1 level (nonnested) models.`
You have specified a nested logit model and requested rank data for the observed outcomes. The nested logit model cannot be estimated with ranks data.

1020 `Not used specifically. May show up with a self explanatory message.`

1021 `Using Box-Cox function on a variable that equals 0?`

1022 `Insufficient valid observations to fit a model.`

1023 `Mismatch between current and last models.`
This occurs when you are using the ; Simulation = ... part of *NLOGIT*.

1024 `Failure estimating DISCRETE CHOICE model.`
Since this occurs during an attempt to compute the starting values for other models, if it fails here, it won't succeed in the more complicated model.

1025 `Failed to fit model. See earlier diagnostic.`
This is a general diagnostic that precedes exit from the estimator. An error condition has occurred, generally during estimation, not setup.

1026 `Singular VC may mean model is unidentified. Check tree.`
What looks like convergence of a nested logit model may actually be an unidentified model. In this case, the covariance matrix will show up with at least one column of zeros. Sometimes it is more subtle than this. In a complicated model, the configuration of the tree may lead to nonidentification. A common source is too many constant terms in the model.

1027 `Models - estimated variance matrix of estimates is singular.`
`Non P.D. 2nd derivatives. Trying BHHH estimator instead.`
This is just a notice. In almost all cases, the Hessian for a model that is not the simple MNL model will fail to be positive definite at the starting values. This does not indicate any kind of problem.

1028 `In ;SIMULATION=list of alts, a name is unknown.`

1029 `Did not find closing ] in labels[list].`

1030 `Error in specification of list in ;Choices=...labels[list].`

1031 `List in ;Choices=...labels[list] must be 1 or NALT values.`

1032 `Merging SP and RP data. Not possible with 1 line data setup.`
`Merging SP and RP data requires LHS=choice,NALTi,ALTij form.`
`Check :MERGERPSP(id=variable, type=variable) for an error.`

1033 `Indiv. <nnnnnn> with ID= <nnnnn> has same ID as another individual.`
This makes it impossible to merge the data sets.

1034 `Specification error. Scenario must begin with a colon.`

1035 `Expected to find Scenario: specification = value.`

1036 `Unbalanced parentheses in scenario specified.`

1037 `Choice given in scenario: attr(choice...) is not in the model.`

1038 `Cannot identify attribute specified in scenario.`

1039 `Value after = in scenario spec is > 20 characters.`

1040 `Cannot identify RHS value in scenario spec.`

1041 `Transformation asks for divide by zero.`

1042 `Can only analyze 5 scenarios at a time.`

1043 `Did not find any valid observations for simulation.`

1044 `Expected to find ; LIST : name_x ( choices ). Not found.`

1045 `Did not find matching ( or [ in <scenario specification is given>.`

1046 `Cannot recognize the name  <AAAAAAAA> in <scenario specification is given>.`

1047 `Same as 1046.`

1048 `None of the attributes requested appear in the model.`

1049 `Model has no free parameters among slopes!`
This occurs during an attempt to fit the MNL model to obtain starting values for a nested logit or some other model.

1050 `DISC with RANKS. Obs= <nnnnnn>. Alt= <nn>. Bad rank given = <nnnn>.`
`DISC w/ RANKS. Incomplete set of ranks given for obs. <nnnnnn>.`
These are data problems with the coding of the Lhs variable.

1051 `Singular VC matrix trying to fit MNL model.`
When the MNL breaks down, it will be impossible to fit a more elaborate model such as a nested logit model.

1052 `You did not provide ;FCN=label(distn),... for RPL model.`

1053 Scaling option is not available with HEV, RPL, or MNP model.
Ranks data may not be used with HEV, RPL, or MNP model.
Nested models are not available with HEV, RPL, or MNP model.
Cannot keep cond. probs. or IVs with HEV, RPL, or MNP model.
Choice based sampling not useable in HEV, RPL, or MNP model.

These diagnostics are produced by problems setting up the scaling option for mixed data sets.

1054 Scaling option is not available with one line data setup.
Ranks data may not be used with one line data setup.
Choice set may not be variable with one line data setup.
One line data setup requires ;RHS and/or ;RH2 spec.
Nested models are not available with one line data setup.
Cannot keep probabilities or IVs with one line data setup.

1055 Did not find closing paren in ;SCALE(list) spec.
The list of variables to be scaled has an error.
Only 40 or fewer variables may be scaled.
You are attempting to scale the LHS variable.
The list of values given for SCALE grid is bad.
Grid must = Lo,Hi,N or Lo,Hi,N,N2. Check spec.
Grid must have Low > 0 and High > low. Check #s.
Number of grid points must be 2,3,... up to 20.

1056 Unidentified name in IIA list. Procedure omitted.

1057 More than 5 names in IIA list. Limit is 5.

1058 Size variables only available with (Nested) MNL.

1059 Cannot locate size variable specified.

1060 Model is too large: Number of betas up to 90.
Model is too large: Number of alphas up to 30.
Model is too large: Number of gammas up to 15.
Model is too large: Number of thetas up to 10.

1061 Number of RHS variables is not a multiple of # of choices.
This occurs when you are using a one line setup for your data.

1062 Expected ;FIX=name[...]. Did not find [or].

1063 In ;FIX=name[...], name does not exist: <name is given>.

1064 Error in fixed parameter given for <name is given>.

1065 Wrong number of start values given.
This occurs with nested logit and other models, not the random parameters logit model.

1066 `Command has both ;RHS and Model: U(alts). Inconsistent.`

1067 `Syntax problem in ;USET:(names list)= list of values.`

1068 `;USET: list of parms contains an unrecognized name.`

1069 `Warning, ;IUSET: # values not equal to # names.`

1070 `Warning, ;IUSET: # values not equal to # names.`

1071 `Spec for RPL model is label(type) or [type]. Type=N,C,or L.`

1072 `Expected ,;$ in COR/SDV/HFN/REM/AR1=list not found.`

1073 `Invalid value given for correl. or std.dev. in list.`

1074 `;COR/SDV=list did not give enough values for matrix.`

1075
```
Error. Expected [ in ;EQC=list[value] not found.
Error:Value in EQC=list[value] is not a correlation.
Error. Unrecognized alt name in ;EQC=list[value].
Error:List needs more than 1 name in EQC=list[value].
Error. A name is repeated in ;EQC=list[value].
```

1076 `Your model forces a free parameter equal to a fixed one.`

1077 `Covariance heterogeneity model needs nonconstant variables.`

1078
```
Covariance heterogeneity model not available with HEV model.
Covariance heterogeneity model is only for 2 level models.
Covariance heterogeneity model needs 2 or more branches.
```

1079 `At least one variance in the HEV model must be fixed.`
In *NLOGIT*, in the heteroscedastic extreme value, you have specified the model so that all the variances are free. But, for identification, one of them must be fixed.

1080 `Multiple observation RPL/MNP data must be individual.`

1081
```
Mismatch of # indivs. and number implied by groups.
WARNING   Halton method is limited to 25 random parameters.
```

1082 `Not used.`

1083 `MODEL followed by a colon was expected, not found.`

1084 `Expected equation specs. of form U(...) after MODEL.`

1085 `Unidentified name found in <string is given>.`
This occurs during translation of ; Model: U (...) specifications.

1086 U(list) must define only choices, branches, or limbs.

1087 An equals sign was not found where expected in utility function definition.

1088 Mismatched [or (in parameter value specification.

1089 Could not interpret string; expected to find number.

1090 Expected to find ;IVSET:=defn. at this point.

1091 Expected to find a list of names in parens in IVSET.

1092 IVSET:(list) ... Unidentified name appears in (list).

1093 You have given a spec for an IV parm that is fixed at 1.

1094 You have specified an IV parameter more than once.

1095 Count variable <nnnnnn> at row <nnnnnn> equals <nnnn>.
The peculiar value for the count variable has thrown off the counter that keeps track of where the estimator is in the data set.

1096 Choice variable <AAAAAAAA> at row <nnnnn>: Choice= <nnnnn>.
The most likely cause is a coding error. Check for bad data.

1097 Obs. <nnnnnn>: Choice set contains <nnnn> <nnnn> times.
The choice variable for individual data has more than one 1.0 in it. *NLOGIT* cannot determine which alternative is chosen.

1098 Obs. <nnnnnn> alt. <nnn> is not an integer nor a proportion.

1099 Obs. <nnnnnn> responses should sum to 1.0. Sum is <xxxxxx>.

1100 Cannot classify obs. <nnnnnn> as IND, PROPs, or FREQs.
Your data appear to be a mix of individual and frequency data. This occurs when an individual's Lhs variable data include zeros. It then becomes difficult to determine what kind of data you have. You can settle the question by including ; Frequencies in your command, if that is appropriate.

1101 # of parms in < list > greater than # choices in U(list).

1102 RANK data can only be used for 1 level (nonnested) models.

1103 Wrong number of variables given in ;CLASSP=list.
;CLASSP=list contains ONE. Cannot save P(j|i) in ONE.

1104 Negative value in NLRP;Tau=value is ignored
Negative value in GMXL;Tau=value is ignored
Value not in [0,1] in GMXL;Gamma=value
Unknown name in ;RPASC=list. Spec. ignored

The following diagnostics are returned by the ; CheckData program in *NLOGIT*: The reports includes the data row of the observation and the individual number in the current sample.

```
No choice was made by this individual
Sum of LHS variable for individual should equal 1. Does not
Sum of proportions does not equal 1
Sum of LHS variable for all alts is zero
Missing value found for LHS variable
Missing value found for weighting variable
Missing value found for characteristic or attribute in utility
Missing value found for utility in branch equation
Missing value found for utility in limb equation
Missing value found for utility in trunk equation
Missing value found for LHS variable, one line format
Restricted choice set. Indiv. chose one of the excluded alts
Missing value found for RPL=variable or MNP variance variable
Trying to take log of nonpositive value in twig utility
Box-Cox transformation is applied to nonpositive value in twig
Trying to take log of nonpositive value in branch utility
Box-Cox transformation applied to nonpositive value in branch
Trying to take log of nonpositive value in limb utility
Box-Cox transformation is applied to nonpositive value in limb
Trying to take log of nonpositive value in trunk utility
Box-Cox transformation is applied to nonpositive value in trunk
IIA excluded choices. Individual chose one of excluded alts.
Missing value for variables in covariance heterogeneity
One line setup. LHS variable value is not in choices list
Individual data, LHS variable is not 0 or 1
Universal choice set, ID var. (3rd LHS) takes same value twice
Proportions data for LHS. Value found not a proportion [0 to 1]
Frequency data for LHS. Value found not a nonnegative integer
Universal choice set. LHS variable is < 0 or > no. of alts
Variable choice set size.2nd LHS var. must be same for all alts
```

The following diagnostics are returned by the command parser for the nonlinear random parameters logit (NLRPLOGIT) model:

1121 Too many parameters in list (over 150)

1122 num_symbol, num negative or greater than 150

1123 No. of start values must equal no. of labels

1124 NLRPLogit requires ;Start=starting values

1125 Error reading starting values for NLRPLogit

1126 Error in ;FIX=list of labels for NLRPLogit

1127 Invalid parameter name (;label) <name> is a <type>

1128 Fn. name conflicts with var. or other name

1129 Unbalanced parentheses in function defn

1130 Table overflow. Function is too complex

1131 Error in function. See earlier error msg

1132 Expected to find ;Model:U(...) = name / ...

1133 Utility spec uses a function not in the table

1134 Expected ;Fnj=function name=function defnn

1135 Alternative function name may not use a label

1136 Expected ending] in name[...] was not found

1137 Unknown name appears in list in name[list]

1138 WTP setup for NLRP must be alt[xvar/xvar]

1139 Alt name in WTP spec for NLRP is unknown

1140 X var name in Alt[Xvar/Yvar] is unknown

1141 Y var name in Alt[Xvar/Yvar] is unknown

1142 Expected ;888:(xname,blabel) colon not found

1143 Expected (xname,bname) found incorrect specs

1144 Table full,25 specs for 888:(xname,bname)/...

1151 User fn. in RPMIN/MAX is nonpositive. Using Log(.)?

1152 Numerical underflow Product of F(i,r,t) is too small

1153 Numerical overflow Product of F(i,r,t) is too large

NLOGIT 6 References

Abramovitz, M. and Stegun, I. (1972) *Handbook of Mathematical Functions*, Dover Press, New York.

Allenby, G. and Ginter, J. (1995) 'The Effects of In-Store Displays and Feature Advertising on Consideration Sets,' *International Journal of Research in Marketing*, 12, pp. 67-80.

Angrist, J. and Pischke, J. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University

Bates, J. (1999) 'More Thoughts on Nested Logit,' Mimeo, John Bates Services, Oxford.

Beggs, J., Cardell, S., and Hausman, J. (1981) 'Assessing the Potential Demand for Electric Cars,' *Journal of Econometrics*, 17, pp. 1-19.

Bera, A., Jarque, C., and Lee, L. (1984) 'Testing the Normality Assumption in Limited Dependent Variable Models,' *International Economic Review*, 25, pp. 563-578.

Berry, S., Levinsohn, J., and Pakes, A. (1995) 'Automobile Prices in Market Equilibrium,' *Econometrica*, 63, pp. 841-890.

Bhat, C. (1995) 'A Heteroscedastic Extreme Value Model of Intercity Mode Choice,' *Transportation Research*, 29B, 6, pp. 471-483.

Bhat, C. (1996) 'Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling,' Working Paper, Department of Civil and Environmental Engineering, University of Massachusetts, Amherst.

Bhat, C. (2001) 'Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model,' *Transportation Research*, 35B, pp. 677-693.

Boyes, W., Hoffman, D., and Low, S. (1998) 'An Econometric Analysis of the Bank Credit Scoring Problem,' *Journal of Econometrics*, 40, pp. 3-14.

Brownstone, D. and Train, K. (1999) 'Forecasting New Product Penetration with Flexible Substitution Patterns,' *Journal of Econometrics*, 89, pp. 109-129.

Butler, J. and Chatterjee, S. (1997) 'Tests of the Specification of Univariate and Bivariate Ordered Probit,' *Review of Economics and Statistics*, 79, 2, pp. 343-347.

Chamberlain, G. (1980) 'Analysis of Covariance with Qualitative Data,' *Review of Economic Studies*, 47, pp. 225-238.

Chesher, A. and Irish, M. (1987) 'Residual Analysis in the Grouped Data and Censored Normal Linear Model,' *Journal of Econometrics*, 34, pp. 33-62.

Chorus, C. (2010) 'A New Model of Random Regret Minimization,' *European Journal of Transport and Infrastructure Research*, 10, 2, pp. 181-196.

Chorus, C., Greene, W., and Hensher, D. (2013) 'Random Regret Minimization or Random Utility Maximization: An Exploratory Analysis in the Context of Automobile Fuel Choice,' *Journal of Advanced Transportation*, 47, 7, pp. 667-678.

Christofides, L., Stengos, T., and Swidinsky, R. (1997) 'On the Calculation of Marginal Effects in the Bivariate Probit Model,' *Economics Letters*, 54, 3, pp. 203-208.

Daly, A., Hess, S., and Train, K. (2012) 'Assuring Finite Moments for Willingness to Pay in Random Coefficient Models,' *Transportation*, 39,1, pp. 19-31

Davidson, R. and MacKinnon, J. (1993) *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Estrella, A. (1998) 'A New Measure of Fit for Equations with Dichotomous Dependent Variables,' *Journal of Business and Economic Statistics*, 16, 2, pp. 198-205.

Fiebig, D., Keane, M., Louviere, J., and Wasi, N. (2010) 'The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity,' *Marketing Science*, 29, 3, pp. 393-421.

Fomby, T., Hill, R.C., and Johnson, S. (1984) *Advanced Econometric Methods*, Springer Verlag, Heidelberg.

Glewwe, P. (1997) 'A Test of the Normality Assumption in the Ordered Probit Model,' *Econometric Reviews*, 16, pp. 1-19.

Gong, X., van Soest, A., and Villagomez, E. (2000) 'Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico,' IZA Working Paper 213, Bonn.

Greene, W. (1992) 'A Statistical Model for Credit Scoring,' Working Paper 92-29, Department of Economics, Stern School of Business, New York University, New York.

Greene, W. (1998) 'Gender Economics Courses in Liberal Arts Colleges: Further Results,' *Journal of Economic Education*, 29, 4, pp. 291-300.

Greene, W. (2001) ' Fixed and Random Effects in Nonlinear Models,' Working Paper 01-01, Department of Economics, Stern School of Business, New York University, New York.

Greene, W. (2012) *Econometric Analysis*, 7th Edition, Prentice Hall, Englewood Cliffs.

Greene, W. (2015) 'Multinomial Choice Modeling with Aggregate Share Data,' http://people.stern.nyu.edu/wgreene/DiscreteChoice/2015/ME-5-3-BLP.pptx.

Greene, W. and Hensher, D. (2010) *Modeling Ordered Choices*, Cambridge University Press, Cambridge.

Greene, W. and Hensher, D. (2013) 'Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model,' *Applied Economics*, 24, 14, pp. 1897-902.

Greene, W. and McKenzie, C. (2015) 'An LM Test Based on Generalized Residuals for Random Effects in a Nonlinear Model,' *Economics Letters*, 126, pp. 47-50.

Harris, M. and Zhao, X. (2004) 'Modelling Tobacco Consumption with a Zero-Inflated Ordered Probit Model,' Working Paper 14/04, Department of Econometrics and Business Statistics, Monash University, Clayton.

Harris, M. and Zhao, X. (2007) 'A Zero Inflated Ordered Probit Model with an Application to Modeling Tobacco Consumption,' *Journal of Econometrics*, 141, pp. 1073-1099.

Hausman, J. and McFadden, D. (1984) 'Specification Tests for the Multinomial Logit Model,' *Econometrica*, 52, pp. 1219-1240.

Heckman, J. (1979) 'Sample Selection Bias as a Specification Error,' *Econometrica*, 47, pp. 153-161.

Heckman, J. (1981) 'The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,' in Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, pp. 114-178.

Heckman, J. and MaCurdy, T. (1980) 'A Life Cycle Model of Female Labor Supply,' *Review of Economic Studies*, 47, pp. 247-283.

Heckman, J. and Singer, B. (1984) 'Econometric Duration Analysis,' *Journal of Econometrics*, 24, pp. 63-132.

Hensher, D. and Greene, W. (2002) 'Specification and Estimation of Nested Logit Models,' *Transportation Research*, 36B, 1, pp. 1-18.

Hensher, D. and Greene, W. (2003) 'The Mixed Logit Model: The State of Practice,' *Transportation Research, B*, 30, pp. 133-176.

Hensher, D. and Johnson, N. (1981) *Applied Discrete Choice Modelling*, John Wiley and Sons, New York.

Hensher, D., Rose, J., and Greene, W. (2005) ‘The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes,’ *Transportation*, 32 (3), pp. 203-222.

Hensher, D., Rose, J., and Greene, W. (2011) ‘Accounting for Endogeneity of Attribute Non-Attendance in Valuing Travel Time Savings: A Note and a Warning for Stated Choice Experiment Design,’ MSP, Sydney University, ITLS.

Hensher, D., Rose, J., and Greene, W. (2015) *Applied Choice Analysis*, 2nd Edition, Cambridge University Press, Cambridge.

Horowitz, J. (1993) ‘Semiparametric Estimation of a Work-Trip Mode Choice Model,’ *Journal of Econometrics*, 58, pp. 49-70.

Hunt, G. (2000) ‘Alternative Nested Logit Model Structures and the Special Case of Partial Degeneracy,’ *Journal of Regional Science*, 40, pp. 89-113.

Hyslop, D. (1999) ‘State Dependence, Serial Correlation, and Heterogeneity in Labor Force Participation of Married Women,’ *Econometrica*, 67, 6, pp. 1255-1294.

Jain, D., Vilcassim, N., and Chintagunta, P. (1994) ‘A Random-Coefficients Logit Brand Choice Model Applied to Panel Data,’ *Journal of Business and Economic Statistics*, 12, 3, pp. 317-328.

Kim, H. and Pollard, J. (1990) ‘Cube Root Asymptotics,’ *Annals of Statistics*, pp. 191-219.

Klein, R. and Spady, R. (1993) ‘An Efficient Semiparametric Estimator for Discrete Choice Models,’ *Econometrica*, 61, pp. 387-421.

Krailo, M. and Pike, M. (1984) ‘Conditional Multivariate Logistic Analysis of Stratified Case-Control Studies,’ *Applied Statistics*, 44, 1, pp. 95-103.

Lee, J. and Seo, K. (2015) ‘A Computationally Fast Estimator for Random Coefficients Logit Demand Models Using Aggregate Data,’ *The RAND Journal of Economics*, 46, 1, pp. 86-102.

Lee, L. (1983) ‘Generalized Econometric Models with Selectivity,’ *Econometrica*, 51, pp. 507-512.

Lerman, S. and Manski, C. (1981) ‘On the Use of Simulated Frequencies to Approximate Choice Probabilities,’ in Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

Long, S. (1997) *Regression Models for Categorical and Limited Dependent Variables*, Sage Publications, Thousand Oaks.

Maddala, G. S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Manski, C. (1975) ‘Maximum Score Estimation of the Stochastic Utility Model,’ *Journal of Econometrics*, 3, pp. 205-228.

Manski, C. (1985) ‘Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,’ *Journal of Econometrics*, 27, pp. 313-333.

Manski, C. and McFadden, D. (eds.) (1981) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

Manski, C. and Thompson, S. (1985) ‘Operational Characteristics of Maximum Score Estimation,’ *Journal of Econometrics*, 32, pp. 85-108.

Manski, C. and Thompson, S. (1987) ‘MSCORE: A Program for Maximum Score Estimation of Linear Quantile Regressions from Binary Response Data With NPREG: A Program for Kernel Estimation of Univariate Nonparametric Regression Functions,’ Department of Economics, University of Wisconsin, Madison.

McFadden, D. (1981) ‘Econometric Models of Probabilistic Choice,’ in Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

McFadden, D. and Train, K. (2000) 'Mixed MNL Models for Discrete Response,' *Journal of Applied Econometrics*, 15, pp. 447-470.

Nerlove, M. and Press, J. (1973) 'Univariate and Multivariate Log-Linear and Logistic Models,' RAND Corporation Report R-1306-EDA/NIH.

Nevo, A. (2000) 'A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand,' *Journal of Economics and Management Strategy*, 9, 4, pp. 513-548.

Newey, W. (1987) 'Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables,' *Journal of Econometrics*, 36, pp. 231-250.

Pudney, S. and Shields, M. (2000) 'Gender, Race, Pay and Promotion in the British Nursing Profession: Estimation of a Generalized Ordered Probit Model,' *Journal of Applied Econometrics*, 15, 4, pp. 367-399.

Revelt, D. and Train, K. (1998) 'Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level,' *Review of Economics and Statistics*, 80, 4, pp. 647-657.

Scarpa, R., Thiene, M., and Train, K. (2008) 'Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps,' *American Journal of Agricultural Economics*, 90, 4, pp. 994-1010.

Schmidt, P. and Strauss, R. (1975) 'The Predictions of Occupation Using Multinomial Logit Models,' *International Economic Review*, 16, 2, pp. 471-486.

Small, K. and Hsiao, C. (1985) 'Multinomial Logit Specification Tests,' *International Economic Review*, 26, 5, pp. 619-626.

Train, K. (1998) 'Recreation Demand Models with Taste Differences over People,' *Land Economics*, 74, pp. 230-239.

Train, K. (1999) 'Halton Sequences for Mixed Logit,' Manuscript, Department of Economics, University of California, Berkeley.

Train, K. (2009) *Discrete Choice Models with Simulation*, 2nd Edition, Cambridge University Press, Cambridge.

Train, K., Hess, S., and Polak, J. (2004) 'On the Use of Randomly Shifted and Shuffled Uniform Vectors in the Estimation of a Mixed Logit Model for Vehicle Choice,' Paper 04-433, 83rd Annual Meeting of the Transportation Research Board, Washington DC.

Wong, W. (1983) 'On the Consistency of Cross-Validation in Kernel Nonparametric Regression,' *The Annals of Statistics*, 11, pp. 1136-1141.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

Wynand, P. and van Praag, B. (1981) 'The Demand for Deductibles in Private Health Insurance,' *Journal of Econometrics*, 17, pp. 229-252.

Zavoina, R. and McElvey, W. (1975) 'A Statistical Model for the Analysis of Ordinal Level Dependent Variables,' *Journal of Mathematical Sociology*, Summer, pp. 103-120.

NLOGIT 6 Index