NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS

Department of Economics

Econometric Analysis of Panel Data

Professor William Greene Phone: 212.998.0876
Office: KMC 7-90 Home page: people.stern.nyu.edu/wgreene
Email: wgreene@stern.nyu.edu
URL for course web page:
people.stern.nyu.edu/wgreene/Econometrics/PanelDataEconometrics.htm

Final Examination: Spring 2016

This is a 'take home' examination. Today is Tuesday, May 3, 2016. Your answers are due by Friday, June 3, 2016. You may use any resources you wish – textbooks, computer, the web, etc. – but please work alone and submit only your own answers to the questions.

The five parts of the exam are weighted as follows:

Part I.	**Literature**	**20**
Part II.	**The Mundlak Estimator**	**20**
Part III.	**Panel Data Regressions**	**50**
Part IV.	**Binary Choice Models**	**50**
Part V.	**A Loglinear Model**	**60**

Note, in parts of the exam in which you are asked to report the results of computation, please filter your response so that you present the numerical results as part of an organized discussion of the question. Do not submit long, unannotated pages of computer output. Some of the parts require you to do some computations. Use Stata, R, *NLOGIT*, MatLab or any other software you wish to use.

Part I. Literature

Locate a published study in a field that interests you that uses a panel data based methodology. Describe in no more than one page the study, the estimation method(s) used and the conclusion(s) reached by the author(s).

Part II. The Mundlak Approach inEstimation

Many recent studies have revived Mundlak's approach to modeling common effects in linear regression and nonlinear models. Describe in detail the standard common effects models. How is the Mundlak estimator motivated? How is it employed? Show how the estimator provides a constructive test for fixed vs. random effects.

Part III. Panel Data Regressions

The course website contains an abbreviated version of the WHO health outcomes data set,

http://people.stern.nyu.edu/wgreene/Econometrics/WHO-balanced-panel.csv

and as an nlogit project,

http://people.stern.nyu.edu/wgreene/Econometrics/WHO-balanced-panel.lpj

The csv file is a text, comma delimited file that should be directly readable by other programs such as Stata and R. The original data set contained 840 observations as an unbalanced panel for 191 countries. It also contained data for some internal political districts such as the 24 states of Mexico and the provinces of Canada and Australia. This panel retains the data for the 140 countries that contain all 5 years of data. The variables in the file are

COUNTRY	= Country name (text)
ID, STRATUM	= Country ID. Ignore STRATUM
YEAR	= 1993, …, 1997
COMP and LOGCOMP	= WHO health outcome measure and its log
DALE and LOGDALE	= WHO life expectance and its log
EDUC, LOGEDUC, LOGEDUC2	= Education, log and square of log
HLTHEXP, LOGHEXP, LOGHEXP2	= Health expenditure, log and square of log
PUBTHE	= Share of health expenditure paid by government
LOGED_EX	= LOGHEXP * LOGEDUC
GINI	= Gini coefficient income distribution
TROPICS	= Dummy variable for tropical country
POPDEN, LOGPOPDN	= Population density, people per square kilometer and log
GDPC, LOGGDPC	= Per capita GDP and log
T93,…,T97	= Year dummy variables
GEFF	= World bank measure of government effectiveness
VOICE	= World Bank measure of political efficacy
OECD	= OECD member dummy variable
MEANLCMP	= Country mean of log COMP
MEANLHC	= Country mean of log EDUC
MEANLHC2	= Country mean of log EDUC squared
MEANLEXP	= Country mean of log HEXP

Note that COMP, DALE, EDUC and HLTHEXP are time varying, but all other measured variables are time invariant.

The WHO model originally specified was

$$y_{it} = \alpha + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \gamma_{11} x_{1,it}^2 + \gamma_{22} x_{2,it}^2 + \gamma_{12} x_{1,it} x_{2,it} + \varepsilon_{it}$$

where

$$y = \text{logCOMP},\ x_1 = \text{logEDUC},\ x_2 = \text{logHEXP}.$$

Call this Model A. This is a translog production function. The authors found that the values of γ_{kl} implied a nonconcave production function, and fixed γ_{22} and γ_{12} both to zero in their final presentation. Call this restricted model Model B.

a. Fit the "pooled" model and report your results.

b. Using the pooled model, test the null hypothesis of Model B against the alternative Model A.

c. Using the formulation of Model B, fit a random effects model and a fixed effects model. Use your estimation results to decide which is the preferable model. If you find that neither panel data model is preferred to the pooled model, show how you reached that conclusion. As part of the analysis, test the hypothesis that there are no "country effects."

d. Using the Mundlak approach, determine which model, fixed or random effects is preferred.

e. Assuming that there are "latent individual (county) effects," the asymptotic covariance matrix that is computed for the pooled estimator, $s^2(\mathbf{X}'\mathbf{X})^{-1}$, is inappropriate. What estimator can be computed for the covariance matrix of the pooled estimator that will give appropriate standard errors?

f. The hypothesis of constant returns to scale in the translog model (Model A) would be

$$H_0: \beta_1 + \beta_2 = 1 \text{ and } \gamma_{11} + \gamma_{22} + 2\gamma_{12} = 0$$

Test this hypothesis in the context of Model A.

g. The 2004 Health Economics paper by Greene argued that WHO did not handle the obvious heterogeneity across countries appropriately. Variables GINI, TROPICS, logPOPDN, logGDPC, GEFF, VOICE, OECD all capture dimensions of this heterogeneity. Extend the random effects model to include some (or all) of these variables and test the hypothesis that they significantly add to the explanatory power of the model.

h. Are there "time effects" in the data. One approach find out would be to add the time variables (less one of them) to the preferred regression model and test for their joint significance. A second approach would be to use a CHOW test to test for homogeneity of the regression model over the 5 years. Test the homogeneity assumption using your preferred pooled model.

Part IV. Binary Choice Models

The course website describes the “German Manufacturing Innovation Data.” The actual data are not published on the course website. We will use them for purposes of this exercise, however. You can obtain them by downloading either a csv file,

http://people.stern.nyu.edu/wgreene/Econometrics/probit-panel.csv

or an nlogit project file,

http://people.stern.nyu.edu/wgreene/Econometrics/probit-panel.lpj

This data set contains 1,270 firms and 5 years of data for 6,350 observations in total – a balanced panel. The variables that you need for this exercise are described in the data sets area of the course home page,

http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm

(The csv file can easily be ported to other software such as R, SAS and Stata.) I am interested in a binary choice model for the innovation variable, IP. You will fit your model using at least three of the independent variables in the data set. With respect to the model you specify,

A. THEORY

(a) If you fit a pooled **logit** model, there is the possibility that you might be ignoring unobserved heterogeneity (effects). Wooldridge argues that when one fits a probit model while ignoring unobserved heterogeneity, the raw coefficient estimator (MLE) is inconsistent, but the quantity of interest, the “Average Partial Effects” might well be estimated appropriately. Explain in detail what he has in mind here.
(b) Suppose we were to estimate a “fixed effects” probit model by “brute force,” just by including the 1,270 dummy variables needed to create the empirical model. What would the properties of the resulting estimator likely be? What is “the incidental parameters problem?”
(c) How would I proceed to use Chamberlain’s estimator to obtain a consistent slope estimator for the fixed effects logit model.
(d) Describe in detail how to fit a random effects logit model using quadrature and using simulation for the part of the computations where they would be necessary, under the assumption that the effects are uncorrelated with the other included exogenous variables.
(e) Using the random effects logit model that you described in part (d), describe how you would test the hypothesis that the same logit model applies to the four different sectors in the data set (CONSGOOD,FOOD,RAWMTL,INVGOOD).

B. PRACTICE

(a) Fit a pooled probit model using your specification. Provide all relevant estimation results. (Please condense and organize the results in a readable form.)
(b) Fit a random effects probit model.
(c) Use the Mundlak (correlated random effects) approach to approximate a fixed effects model. Recall this means adding the group means of the time varying variables to the model, then using a random effects model.
(d) Note the difference between the estimates in (b) and (c). Which do you think is appropriate? Explain.

Tip for nlogit users: You can use
CREATE ; new variable = GroupMean(variable,pds=5)$
To obtain the group means you need for a variable.

Part V. A Loglinear Model

This semester, we have examined several 'loglinear models,' including the logit model for binary choice, Poisson and negative binomial models for counts and the exponential model for a continuous nonnegative random variable. We will now examine one more loglinear model. The nonnegative, continuous random variable y|**x** has a Weibull distribution:

$$f(y\,|\,\mathbf{x}) = \lambda_i P\, y_i^{P-1} \exp\left(-\lambda_i y_i^{\,P}\right), y \geq 0, P > 0,$$
$$\lambda_i \quad = \exp(\alpha + \boldsymbol{\beta}'\mathbf{x}_i).$$

(We examined a version of this model in Assignment 5.) Estimation and analysis is based on a sample of N observations on y_i,$\mathbf{x}_i$. The conditional mean function is

$$\mathrm{E}[y_i|\mathbf{x}_i] = \frac{1}{\lambda_i}\Gamma\left(\frac{P+1}{P}\right) = \exp(-\alpha - \boldsymbol{\beta}'\mathbf{x}_i)\Gamma\left(\frac{P+1}{P}\right) \text{(Note the minus sign.)}$$

The variables used in the regressions are described below.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
INCOME	.352135	.176857	.001500	3.067100	27326	0
logINCOM	-1.157442	.491452	-6.502290	1.120732	27326	0
AGE	43.52569	11.33025	25.0	64.0	27326	0
EDUC	11.32063	2.324885	7.0	18.0	27326	0
HSAT	6.785662	2.293725	0.0	10.0	27326	0
MARRIED	.758618	.427929	0.0	1.0	27326	0
HHKIDS	.402730	.490456	0.0	1.0	27326	0

The data set is a panel. There are 7,293 groups with group sizes ranging from 1 to 7. This exercise will examine a variety of regression formulations. I have done the estimation for you; the results appear below. Some of the questions will involve a small amount of ancillary computation.

A. I propose to estimate the parameters (P,α,**β**) by maximum likelihood. The results are shown in regression 1 below. Derive the log likelihood function, likelihood equations and Hessian. Show precisely how to use Newton's method to estimate the parameters. How will you obtain asymptotic standard errors for your estimator? Test the hypothesis of 'the regression model.' That is, test the hypothesis that all of the coefficients in **β** are equal to zero using the likelihood ratio test.

B. There are several interesting special cases of the Weibull model. If P = 1, the model reduces to the exponential model discussed in class. We considered three different ways to test a parametric restriction such as this, Wald, Likelihood ratio and LM tests. Using the results of regressions 1, 2 and 3 below, carry out the three tests. Do the results of the three tests agree?

C. The conditional mean function shown above suggests a nonlinear least squares approach. Note that the conditional mean function can be written

$$E[y\,|\,\mathbf{x}] = \exp\left[\log\Gamma\left(\frac{P+1}{P}\right) - \alpha - \boldsymbol{\beta}'\mathbf{x}\right] = \exp(\delta - \boldsymbol{\beta}'\mathbf{x})$$

Thus, the constant term in the conditional mean function is not $-\alpha$. The nonlinear least squares results are shown in regression 4. How do the two results compare to the MLE? We now have two possible estimators of $\boldsymbol{\beta}$. In theoretical terms, which is better, MLE or NLS? Why? Do the empirical results support your argument?

D. The likelihood equations for estimation of (P,α,β) imply that $E[y^P|x] = 1/\lambda$. Prove this result.

E. Derive the partial effects for the Weibull conditional mean function, $\partial E[y|\mathbf{x}]/\partial \mathbf{x}$. Compute the partial effects at the means of the data. Hint: $\Gamma((P+1)/P)$ for the P in regression 1 equals .88562. How would you obtain standard errors for your estimated partial effects? Explain in detail.

F. Regression 5 presents *linear* least squares results for the regression of $-y$ on $(1,\mathbf{x})$. (The minus sign on y changes the sign of the coefficients so they will be comparable to the earlier results.) How do these results compare to the MLEs in part A? How do they compare to the results in part E? Why would they resemble the results in part E?

G. The log of a Weibull distributed variable has a type 1 extreme value distribution. The expected value of logy is $-(\alpha+\boldsymbol{\beta}'\mathbf{x}) + \gamma$, where γ is the Euler-Mascheroni constant, 0.57721566…. Regression 6 presents the results of linear regression of –logy on x. Which other result should these resemble? Do they?

H. Since these are panel data, it is appropriate to rebuild the model to accommodate the unobserved heterogeneity. Explain the difference between fixed and random effects models. How would they appear in the loglinear model formulated here?

I. Regressions 7 and 8 show FEM and REM.
(1) What is the incidental parameters problem? Would the result apply to the model shown in (7)?
(2) Show how the parameters of the random effects model in regression 8 are computed. I.e., describe how the maximum simulated likelihood estimator is computed.
(3) Regression 9 presents estimates of a random effects model that also contains the group means of the regressors. As noted earlier, this Mundlak style treatment helps to distinguish the FE and RE specifications. Based on the results given, which appears to be the preferable model, FE or RE?

J. Some have argued that marital status might be endogenous in an income equation when there are households that have two working people. (You probably thought people married for love.) To investigate in the present model, I will use a control function approach. Regression 10 presents a probit eqution for marital status based on age, education, gender and whether the household head has a white collar job. The variable GENRES is the generalized residual from this model,
GENRES = $q\phi(\beta'x)/\Phi(q\beta'x)$ where q = 2Married – 1. The expected value of GENRES is zero, and since it is the derivative of logL with respect to the constant term, it will sum to zero in the sample. I am going to use GENRES as a control function. What is a control function, and why will I use it in the INCOME model?

K. Regression 11 presents estimates of the Weibull INCOME model that includes the control function. Regression 12 is similar to 11, but regression 12 includes normal heterogeneity in the model in the form of what appears to be a random effect – a random constant. But, this is not a panel data model look closely at the results and note that the 'panel' has one period. The implied two equation model underlying 12 is

$MARRIED_i^* = \gamma' z + u_i$, $MARRIED_i = 1[MARRIED_i^* > 0]$, $u_i \sim N[0,1]$.

$INCOME_i^* \sim Weibull(\lambda_i, P)$ where $\lambda_i = \exp(\beta' x_i + \varepsilon_i)$

where (ε_I, u_i) have a bivariate normal distribution with means (0,0), standard deviations $(\sigma_\varepsilon, 1)$ and correlation ρ. The endogeneity issue turns on ρ. The coefficient on GENRES in the model in regression 12 will approximate $\sigma_\varepsilon \rho$. So, based on the estimated model, marry for money (endogenous, ρ not equal to zero) or marry for love (exogenous, ρ equal to zero)?

L. In this model, the argument in parts J and K about MARRIED could also be made about health satisfaction, HSAT. But, HSAT is an ordered outcome, coded 0,1,2 (bad, middling, good) in our data. How would you proceed to deal with endogeneity of HSAT in this model?

1. Weibull, MLE

```
-----------------------------------------------------------------------------
Weibull (Loglinear) Regression Model
Dependent variable                 INCOME
Log likelihood function       12133.14495
Restricted log likelihood      1195.24508  (Log likelihood when β = 0)
Chi squared [   7](P= .000)   21875.79975
Significance level                 .00000
McFadden Pseudo R-squared      -9.1511775
Estimation based on N =   27326, K =    8
Inf.Cr.AIC  = -24250.3 AIC/N =     -.887
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  INCOME|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.67075***     .01433   116.62  .0000     1.64267   1.69883
     AGE|     .00086***     .00022     3.91  .0001      .00043    .00130
    EDUC|    -.05084***     .00073   -69.23  .0000     -.05228   -.04940
    HSAT|    -.01233***     .00077   -15.96  .0000     -.01385   -.01082
 MARRIED|    -.16990***     .00371   -45.79  .0000     -.17717   -.16262
  FEMALE|    -.02041***     .00334    -6.11  .0000     -.02696   -.01386
  HHKIDS|     .06403***     .00375    17.07  .0000      .05668    .07139
        |Scale parameter for Weibull model
 P_scale|    2.13722***     .00495   431.40  .0000     2.12751   2.14693
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

	Constant	AGE	EDUC	HSAT	MARRIED	HHKIDS	P_scale
Constant	0.000203282	-2.18842e-006	-5.43762e-006	-4.95755e-006	1.02554e-005	-1.15906e-005	-2.77586e-005
AGE	-2.18842e-006	4.7703e-008	-7.89778e-009	4.31199e-008	-3.07207e-007	1.75484e-007	1.31682e-007
EDUC	-5.43762e-006	-7.89778e-009	5.32837e-007	-5.61965e-008	-3.2693e-007	2.5195e-007	9.33686e-007
HSAT	-4.95755e-006	4.31199e-008	-5.61965e-008	5.90819e-007	-2.22175e-008	-3.58706e-007	-5.27993e-007
MARRIED	1.02554e-005	-3.07207e-007	-3.2693e-007	-2.22175e-008	1.33275e-005	-4.0085e-006	-6.66408e-007
HHKIDS	-1.15906e-005	1.75484e-007	2.5195e-007	-3.58706e-007	-4.0085e-006	1.35887e-005	3.46969e-006
P_scale	-2.77586e-005	1.31682e-007	9.33686e-007	-5.27993e-007	-6.66408e-007	3.46969e-006	2.40568e-005

2. Exponential, MLE

```
-----------------------------------------------------------------------------
Exponential (Loglinear) Regression Model
Dependent variable                 INCOME
Log likelihood function        1558.04494
Restricted log likelihood      1195.24508
Chi squared [   5](P= .000)     725.59973
Significance level                 .00000
McFadden Pseudo R-squared       -.3035360
Estimation based on N =   27326, K =    6
Inf.Cr.AIC  =  -3104.1 AIC/N =     -.114
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  INCOME|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.85106***     .04834    38.29  .0000     1.75632    1.94580
     AGE|     .00158**      .00064     2.48  .0133      .00033     .00283
    EDUC|    -.05438***     .00268   -20.27  .0000     -.05963    -.04912
    HSAT|    -.01101***     .00275    -4.00  .0001     -.01641    -.00561
 MARRIED|    -.26249***     .01568   -16.75  .0000     -.29322    -.23177
  HHKIDS|     .06619***     .01399     4.73  .0000      .03877     .09360
--------+--------------------------------------------------------------------
```

	Constant	AGE	EDUC	HSAT	MARRIED	HHKIDS
Constant	0.00233667	-2.04546e-005	-8.41197e-005	-5.58263e-005	-1.67755e-005	-0.000163745
AGE	-2.04546e-005	4.06572e-007	1.57594e-007	3.21119e-007	-3.44467e-006	3.49705e-006
EDUC	-8.41197e-005	1.57594e-007	7.19294e-006	-7.79848e-007	1.52206e-006	-7.72587e-008
HSAT	-5.58263e-005	3.21119e-007	-7.79848e-007	7.58913e-006	-6.64156e-007	-7.8368e-007
MARRIED	-1.67755e-005	-3.44467e-006	1.52206e-006	-6.64156e-007	0.00024571	-8.04925e-005
HHKIDS	-0.000163745	3.49705e-006	-7.72587e-008	-7.8368e-007	-8.04925e-005	0.000195638

3. Constrained Weibull, MLE

```
-----------------------------------------------------------------------------
Weibull (Loglinear) Regression Model
Dependent variable                 INCOME
LM Stat. at start values      21526.22099
LM statistic kept as scalar        LMSTAT
Log likelihood function        1558.04494
Restricted log likelihood      1195.24508
Chi squared [   6](P= .000)     725.59973
Significance level                 .00000
McFadden Pseudo R-squared       -.3035360
Estimation based on N =   27326, K =    7
Inf.Cr.AIC  =  -3102.1 AIC/N =     -.114
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  INCOME|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.85106***     .09976    18.56  .0000     1.65553    2.04658
     AGE|     .00158        .00130     1.22  .2233     -.00096     .00412
    EDUC|    -.05438***     .00574    -9.47  .0000     -.06563    -.04312
    HSAT|    -.01101**      .00556    -1.98  .0477     -.02190    -.00011
 MARRIED|    -.26249***     .02881    -9.11  .0000     -.31896    -.20603
  HHKIDS|     .06619**      .02827     2.34  .0192      .01078     .12160
        |Scale parameter for Weibull model
 P_scale|        1.0***     .00672   148.81  .0000      .98683  .10132D+01
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

4. Nonlinear Least Squares, y on exp(-b'x)

Nonlinear	least squares regression		
LHS=INCOME	Mean	=	.35214
	Standard deviation	=	.17686
Fit	R-squared	=	.11070
	Adjusted R-squared	=	.11073
Model test	F[5, 27320] (prob)	=	680.2(.0000)

UserFunc	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
B_ONE	1.92270***	.02202	87.33	.0000	1.87955	1.96585
B_AGE	-.00022	.00030	-.75	.4535	-.00081	.00036
B_EDUC	-.05378***	.00103	-52.09	.0000	-.05580	-.05175
B_HSAT	-.01072***	.00132	-8.12	.0000	-.01330	-.00813
B_MARR	-.25986***	.00821	-31.66	.0000	-.27594	-.24377
B_KIDS	.05581***	.00664	8.40	.0000	.04279	.06883

5. Linear Least Squares, -y on b'x

Ordinary	least squares regression				
LHS=MINCOME	Mean	=	-.35214		
	Standard deviation	=	.17686		
----------	No. of observations	=	27326	DegFreedom	Mean square
Regression	Sum of Squares	=	93.8115	5	18.76231
Residual	Sum of Squares	=	760.870	27320	.02785
Total	Sum of Squares	=	854.682	27325	.03128
----------	Standard error of e	=	.16688	Root MSE	.16687
Fit	R-squared	=	.10976	R-bar squared	.10960
Model test	F[5, 27320]	=	673.68410	Prob F > F*	.00000

MINCOME	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	-.03873***	.00815	-4.75	.0000	-.05470	-.02275
AGE	.00012	.00010	1.14	.2535	-.00008	.00032
EDUC	-.02088***	.00044	-47.07	.0000	-.02175	-.02001
HSAT	-.00366***	.00046	-8.03	.0000	-.00455	-.00277
MARRIED	-.08630***	.00260	-33.21	.0000	-.09140	-.08121
HHKIDS	.02024***	.00238	8.51	.0000	.01558	.02489

6. Linear Least Squares, -logy on b'x

Ordinary	least squares regression				
LHS=MLINCOME	Mean	=	1.15744		
	Standard deviation	=	.49145		
----------	No. of observations	=	27326	DegFreedom	Mean square
Regression	Sum of Squares	=	968.991	5	193.79827
Residual	Sum of Squares	=	5630.67	27320	.20610
Total	Sum of Squares	=	6599.66	27325	.24152
----------	Standard error of e	=	.45398	Root MSE	.45393
Fit	R-squared	=	.14682	R-bar squared	.14667
Model test	F[5, 27320]	=	940.30851	Prob F > F*	.00000

MLINCOME	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	2.03085***	.02217	91.60	.0000	1.98740	2.07430
AGE	.00190***	.00028	6.73	.0000	.00135	.00246
EDUC	-.05651***	.00121	-46.82	.0000	-.05887	-.05414
HSAT	-.01175***	.00124	-9.47	.0000	-.01418	-.00932
MARRIED	-.34733***	.00707	-49.14	.0000	-.36118	-.33348
HHKIDS	.06628***	.00647	10.25	.0000	.05361	.07896

7. Fixed Effects Weibull, MLE

```
-----------------------------------------------------------------------------
FIXED EFFECTS Weibul Model
Dependent variable               INCOME
Log likelihood function     34910.40335
Estimation based on N =  27326, K =7299
Inf.Cr.AIC  = -55222.8 AIC/N =   -2.021
Unbalanced panel has   7293 individuals
Skipped    0 groups with inestimable ai
Weibull loglinear regression model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
     AGE|    -.04322***       .00055   -78.85  .0000     -.04429   -.04214
    EDUC|    -.07959***       .00616   -12.91  .0000     -.09167   -.06750
    HSAT|    -.00339***       .00088    -3.85  .0001     -.00511   -.00166
 MARRIED|    -.18215***       .00836   -21.80  .0000     -.19853   -.16578
  HHKIDS|     .07732***       .00550    14.06  .0000      .06654    .08810
        |Scale parameter for Weibull distribution
 P_scale|    5.77115***       .02935   196.61  .0000     5.71362   5.82868
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

8. Random Effects Weibull, Maximum Simulated Likelihood

```
-----------------------------------------------------------------------------
Random Coefficients  WeiblReg Model
Dependent variable               INCOME
Log likelihood function     19489.51857
Restricted log likelihood    1558.04494
Chi squared [  1](P= .000)  35862.94726
Significance level               .00000
McFadden Pseudo R-squared   -11.5089579
Estimation based on N =  27326, K =   8
Inf.Cr.AIC  = -38963.0 AIC/N =   -1.426
Unbalanced panel has   7293 individuals
Simulation  based on   100 Halton draws
Weibull loglinear regression model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|    -.01369***       .00015   -91.51  .0000     -.01398   -.01339
    EDUC|    -.06413***       .00057  -111.59  .0000     -.06525   -.06300
    HSAT|    -.00478***       .00059    -8.05  .0000     -.00594   -.00361
 MARRIED|    -.19181***       .00307   -62.53  .0000     -.19782   -.18580
  HHKIDS|     .08751***       .00271    32.26  .0000      .08219    .09282
        |Means for random parameters
Constant|    2.43436***       .01030   236.40  .0000     2.41418   2.45455
        |Scale parameters for dists. of random parameters
Constant|     .50166***       .00138   364.84  .0000      .49896    .50435
        |Scale parameter for Weibull distribution
 P_scale|    4.15999***       .01130   368.03  .0000     4.13783   4.18214
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

9. Random Effects Weibull, Maximum Simulated Likelihood with Group Means

```
-----------------------------------------------------------------------------
Random Coefficients  WeiblReg Model
Dependent variable              INCOME
Log likelihood function    21443.98658
Restricted log likelihood   1735.72267
Chi squared [  1](P= .000)  39416.52782
Significance level              .00000
McFadden Pseudo R-squared   -11.3545005
Estimation based on N =  27326, K =  13
Inf.Cr.AIC  = -42862.0 AIC/N =   -1.569
Unbalanced panel has   7293 individuals
Simulation  based on   100 Halton draws
Weibull loglinear regression model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  INCOME|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Nonrandom parameters
     AGE|    -.04140***      .00041  -100.59  .0000     -.04221   -.04060
    EDUC|    -.07436***      .00382   -19.47  .0000     -.08185   -.06687
    HSAT|    -.00775***      .00076   -10.19  .0000     -.00924   -.00626
 MARRIED|    -.20973***      .00473   -44.37  .0000     -.21899   -.20046
  HHKIDS|     .08242***      .00454    18.17  .0000      .07353    .09132
  gmnAGE|     .04656***      .00043   107.94  .0000      .04571    .04741
 gmnEDUC|     .03803***      .00384     9.90  .0000      .03050    .04556
 gmnHSAT|    -.01136***      .00083   -13.69  .0000     -.01299   -.00974
gmnMARRI|     .01497**       .00587     2.55  .0107      .00347    .02646
gmnHHKID|    -.01393**       .00603    -2.31  .0210     -.02575   -.00210
        |Means for random parameters
Constant|    1.41275***      .00930   151.96  .0000     1.39453   1.43097
        |Scale parameters for dists. of random parameters
Constant|     .46708***      .00118   395.64  .0000      .46477    .46940
        |Scale parameter for Weibull distribution
 P_scale|    4.38591***      .01205   363.95  .0000     4.36229   4.40953
--------+--------------------------------------------------------------------
```

10. Probit Model for Marital Status, MLE

Binomial Probit Model
Dependent variable MARRIED

MARRIED	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Index function for probability					
Constant	.20370***	.05761	3.54	.0004	.09079	.31662
AGE	.02234***	.00076	29.28	.0000	.02084	.02383
EDUC	-.03308***	.00367	-9.02	.0000	-.04027	-.02589
FEMALE	-.12946***	.01727	-7.50	.0000	-.16330	-.09562
WHITEC	-.03858**	.01861	-2.07	.0382	-.07506	-.00210

11. Weibull with Control Function, MLE

Weibull (Loglinear) Regression Model
Dependent variable INCOME
Log likelihood function 12160.11190
Restricted log likelihood 1195.24508
Chi squared [7](P= .000) 21929.73366
Significance level .00000

INCOME	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters in conditional mean function					
Constant	1.19668***	.03721	32.16	.0000	1.12374	1.26962
AGE	-.00528***	.00051	-10.39	.0000	-.00627	-.00428
EDUC	-.04215***	.00107	-39.51	.0000	-.04425	-.04006
HSAT	-.01251***	.00077	-16.34	.0000	-.01401	-.01101
MARRIED	.67313***	.06091	11.05	.0000	.55375	.79251
HHKIDS	.05201***	.00371	14.03	.0000	.04475	.05927
GENRES	-.49197***	.03508	-14.02	.0000	-.56073	-.42321
	Scale parameter for Weibull model					
P_scale	2.13826***	.00492	434.41	.0000	2.12862	2.14791

12. Weibull with Normal Heterogeneity and Control Function, Maximum Simulated Likelihood

Random Coefficients WeiblReg Model
Dependent variable INCOME
Log likelihood function 13158.01422
Restricted log likelihood 1563.62291
Sample is 1 pds and 27326 individuals
Simulation based on 10 Halton draws

INCOME	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
AGE	-.00331***	.00046	-7.24	.0000	-.00421	-.00241
EDUC	-.04512***	.00097	-46.49	.0000	-.04702	-.04321
HSAT	-.01211***	.00072	-16.84	.0000	-.01352	-.01070
MARRIED	.45855***	.05526	8.30	.0000	.35024	.56686
HHKIDS	.06500***	.00369	17.60	.0000	.05776	.07224
GENRES	-.39467***	.03169	-12.45	.0000	-.45678	-.33256
	Means for random parameters					
Constant	1.31805***	.03473	37.95	.0000	1.24999	1.38612
	Scale parameters for dists. of random parameters					
Constant	.25368***	.00194	130.76	.0000	.24988	.25748
	Scale parameter for Weibull distribution					
P_scale	2.64003***	.00733	359.95	.0000	2.62566	2.65441