NYU STERN
NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS

Department of Economics

ECONOMETRICS I
Take Home Final Examination

Fall 2017

Professor William Greene Office: KMC 7-90, Phone: 212.998.0876
Home page: http://people.stern.nyu.edu/wgreene
e-mail: wgreene@stern.nyu.edu
URL for course: people.stern.nyu.edu/wgreene/Econometrics/Econometrics.htm

Today is Tuesday, December 5, 2017. This exam is due by 10AM, Thursday, December 21, 2017.

Please do not include a copy of the exam questions with your submission; submit only your answers to the questions.

NOTE: In the empirical results below, a number of the form .nnnnnnE+aa means multiply the number .nnnnnn by 10 to the aa power. E-aa implies multiply 10 to the minus aa power. Thus, .123456E-04 is 0.0000123456. Note, as well, D+nn or D-nn or e+nn or e-nn all mean the same as E+nn or E-nn.

This exam involves some empirical work, theoretical support for some of the estimation, and one short piece of research. There are 8 sections. Each section constitutes 25 points, for a total of 200.

1. **Applied Econometrics**
2. **Properties of the least squares estimator**
3. **Interpreting Regression Results**
4. **World Health Organization, Modeling Health Attainment**
5. **Fixed and Random Effects**
6. **Endogeneity and Two Stage Least Squares**
7. **Binary Choice Modeling**
8. **Maximum Likelihood Estimation of a Loglinear Model**

1. Applied Econometrics

This question involves "library" research. Locate an empirical (applied) paper (study) in any field (political science, economics, finance, management, accounting, pharmacology, environment, energy, urban economics, etc.) in which a model that involved an endogenous variable on the right hand side is estimated. (This should be easy to find – most of the contemporary applied literature deals with such situations.) Report (a) what empirical issue the study was about; (b) what the model was; (c) what estimation technique the author used; (d) (briefly) what results they obtained. In part (d), describe the actual statistics that the author reported, and what conclusion they drew. This entire essay should not exceed one double spaced page.

2. Properties of the Least Squares Estimator

a. Show (algebraically) how the ordinary least squares coefficient estimator, **b**, and the estimated asymptotic covariance matrix are computed. Ignore any considerations of robustness. (We will approach this in Sections 3 and 4.)
b. What are the finite sample properties of the ordinary least squres estimator in the context of the linear regression model? Make your assumptions explicit.
c. What are the asymptotic properties of the ordinary least squares estimator? Again, be explicit about all assumptions, and explain your answer carefully.

3. Interpreting Regression Results

The results below are extracted from a paper that examined survey data about individuals' preferences for reliability of electricity service. The data used for the study are a three round panel data set with about 2,400 individuals. Each individual answered 3 surveys with 3 different offerings. A model was fit for which estimated parameters $\hat{\beta}$ are computed. *Willingness to pay* for reliability, $\widehat{WTP_i}$, for each individual i=1,...,2400, is computed using $\hat{\beta}$ and the three sets of data. The results below show the regression of $\ln(\widehat{WTP_i})$ on several variables given several specifications. *Informal user* and *Satisfaction* are dummy variables.

a. Notice in the footnotes the author states that "*Robust standard errors are in parentheses.*" How would these "robust standard errors" be computed? What assumption are the standard errors robust to? (Tip: The data for this part of the analysis are a cross section. Note: the "+2" that appears in the table is used to avoid taking logs of nonpositive values.)
b. Using the results of specification (5) estimate the effect of being an "*Informal user*" on ln(WTP). (Tip: The mean value of "*Satisfaction*" is 0.73.) Compare your result to the estimate based on specification (2).
c. The surveys were obtained using a few hundred observations from each of 8 provinces in the Dominican Republic. Theory 0 (null) states that the model (Specification (6)) for ln(WTP) in the population is the same for all 8 provinces. Theory 1 (alternative) states that the models that apply in the different provinces are different. How would you use the survey data to test the null hypothesis of the model against the more general alternative? (Tip: This question asks *how* you would carry out the test. The results in Annex 8 are not sufficient for the test.)

Annex 8: Determinants of WTP across Individuals

	Dependent: log(WtP for better electricity services)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(US$per capita hh income/100+2)	0.297***	0.347***	0.346***	0.290***	0.301***	0.240***
	(0.037)	(0.036)	(0.036)	(0.032)	(0.034)	(0.032)
Informal user		0.680***	0.724***	0.786***	0.805***	1.023***
		(0.021)	(0.035)	(0.035)	(0.037)	(0.037)
Satisfaction with service		0.159***	0.176***	0.006	0.009	0.037
		(0.025)	(0.032)	(0.033)	(0.032)	(0.030)
Informal*Satisfaction			-0.069	0.030	0.039	0.001
			(0.044)	(0.045)	(0.045)	(0.044)
refrigerator				0.146***	0.107**	0.057
				(0.037)	(0.037)	(0.037)

Notes: The willingness to pay (WTP) estimates are based on model 3. It assumes a restricted triangular distribution for random parameters. The dependent variable is $\ln(WTP_i+2)$ for each individual *i* as generated by model 5. Robust standard errors are in parentheses.

4. World Health Organization, Modeling Health Attainment

Sections 4 - 6 require you to download some data and compute several regressions:
You may use any software that you have used for the exercises during the semester. The data set http://people.stern.nyu.edu/wgreene/Econometrics/WHO-balanced-panel.csv (in csv format) http://people.stern.nyu.edu/wgreene/Econometrics/WHO-balanced-panel.lpj (in nlogit format) contains the data used in the 2000 World Health Organization study of efficiency of world health systems. The data set is a balanced panel with 5 years of data, 1993-1997. There were 191 countries in the original WHO data set. This subset of the data set includes the 140 countries for which all five years of data are available, for 700 country/year observations in total. The data set includes the following variables:

Country	= (text format) country name
LogComp	= log of composite measure of health care outcome in the economy
LogDALE	= log of disability adjusted life expectancy in the country in the given year
Meanlcmp	= country mean of *Logcomp*, repeated for each year
Loghexp	= log of health care expenditure
Meanexp	= country mean of *Loghexp*, repeated for each year
Logeduc	= log of average years of education
Meanlhc	= country mean of *Logeduc*, repeated for each year
Loghexp2	= square of *loghexp*
Logeduc2	= square of *logeduc*
Loged_ex	= *logeduc*loghexp*
Loggdpc	= log of per capita income
Gini	= estimated Gini coefficient for distribution of income
Geff	= World Bank measure of government effectiveness
Voice	= World Bank index of democratization of the political system
OECD	= 0, not a member of OECD; 1, in OECD; 2, not in OECD and 1997.
T93, T94, T95, T96, T97	= year dummy variables for years 1993 to 1997

Define the data matrices: $\mathbf{X}_1$ = *Logeduc,Loghexp* $= (z_1,z_2)$
$\mathbf{X}_2$ = *Logeduc2,Loghexp2,Loged_ex* $= (z_1^2, z_2^2, z_1z_2)$.
$\mathbf{T}$ = *T94, T95, T96, T97* (1993 ia the base year)

a. Obtain and report the overall, pooled sample means of the variables in ($\mathbf{X}_1$)

b. Obtain the results of linear regression of logComp on $(1, \mathbf{X}_1) = (1,z_1,z_2)$. Report all relevant results.

c. What is your estimate of δ, the expenditure elasticity of composite outcome with respect to health care expenditure? (Tip: This is the coefficient on Loghexp.) Provide point and interval estimates.

d. Since these data are a panel, it seems natural to expect correlation across the observations. As such, the conventional standard errors probably need correction. Obtain cluster robust standard errors, clustering on the countries, for the regression results in part b. Discuss the impact of the cluster correction on the estimated standard errors. Describe how the cluster corrected standard errors are computed.

e. The authors at WHO included $\mathbf{X}_2$ in the regression as well. Add $\mathbf{X}_2$ to the model in part b. Test the hypothesis that the coefficients on all three variables in $\mathbf{X}_2$ are zero. (Use the clustered standard errors.)

f. In part b, the model is

(1) $$y = \beta_1 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon$$

In part e, it is (2) $$y = \beta_1 + \beta_2 z_1 + \beta_3 z_2 + \beta_4 z_1^2 + \beta_5 z_2^2 + \beta_6 z_1 z_2 + \varepsilon$$

In part c, your estimate of the elasticity is the estimate of β_3 in (1). In the model in part e, your estimate of the elasticity would be $\delta = \partial y/\partial z_2 = \beta_3 + 2\beta_5 z_2 + \beta_6 z_1$. You have the means for z_1 and z_2 in part a. What is your new estimate of the elasticity from this expanded model at the means of the variables? How does this estimate compare to the one you obtained in part c?

g. In part c, you obtained the standard error for your simple estimate of δ. In part f, you obtained a new, more complex estimate of δ. How would you obtain an estimated standard error for your estimate of δ in part f.?

5. Fixed and Random Effects (FE and RE)

We continue to analyze the WHO data in Section 4. The data are a balanced panel, with 5 years of data. We consider fixed effects and random effects approaches.

a. What is the theoretical difference between the fixed and random effects linear regression models?

b. You obtained the results of the "pooled" regression in part b of Section 4. Report those results again (for convenience). (Tip: Use the simple specification in part b, not part e.)

c. Now estimate a fixed effects model. Report the relevant results. Compare the results to those of the pooled model.

d. Does the cluster correction of the standard errors of the fixed effects estimator produce a large difference? Comment on the results in this regression compared to those in part d in Section 4 (without the fixed effects).

e. Finally, estimate a random effects model and compare your results to the fixed effects results in part d. (Tip: the focus in this analysis is on the two slope coefficients. Discuss the estimates of these.)

f. It will be useful to determine whether the FE or RE is the preferred model. We will use the Wu, variable addition test. Add the two means variables to the random effects model in part e. The test consists of the joint test of the hypothesis that the coefficients on the two group means variables are both zero. Report your result. On the basis of the test, which is the preferred model, fixed or random?

g. To reinforce a surprising result we discussed in class, we will obtain the FE estimates two ways:

(1) Use your software's built in "fixed effects" estimator to obtain the coefficients on z_1 and z_2 and the estimated standard errors.

(2) Use a simple linear regression of logComp on $(1,z_1,z_2,\text{meanz}_1,\text{meanz}_2)$ Compare the coefficients on (z_1,z_2) and the estimated standard errors to the results in (1) above.

6. Endogeneity and Two Stage Least Squares

In the model

$$\text{logComp} = \beta_1 + \beta_2\text{logEduc} + \beta_3\text{logHexp} + \varepsilon,$$

it is unlikely that health expenditure is exogenous. We reconsider using least squares.

a. Explain the concept of endogeneity in the linear regression model.

b. For purposes of this exercises, we will use only the 1997 data. This becomes a cross section. Estimate the model using simple least squares.

b. We can use GEFF, GINI and logGDPC as instruments for two stage least squares. Explain how to compute the 2SLS estimator. Reestimate the model using 2SLS, and report your results. Does your estimate of the expenditure elasticity change very much? What is your finding?

c. It is customary to test the *relevance assumption* with respect to the instrumental variables. In this context, you would do this by regressing logHEXP on a contant, logEduc and the three instruments, then test the hypothesis that the coefficients on the three instruments are all zero. Carry out the test and report your conclusion.

d. A specification test for endogeneity can be carried out using a Wu (variable addition) test. Carry out the Wu test and comment on whether it appears to support the null hypothesis that logHexp is exogenous in this model.

7. Binary Choice Modeling

In this exercise, you will estimate a binary choice model. The analysis is based on a "live" data set of credit card applications. The data file http://people.stern.nyu.edu/wgreene/Econometrics/amex.csv (and also .lpj) contains a cross section of 13,444 actual records on applications for a credit card. The variables in the data set include

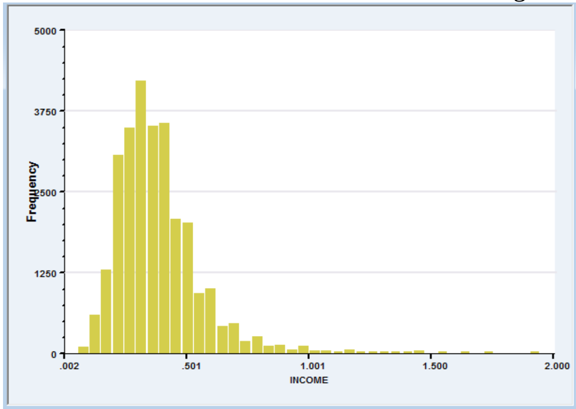
CARDHLDR	= 0 for application rejected, 1 if application accepted
AGE	= age in years and 12ths.
ACADMOS	= number of months living at current address
ADEPCNT	= number of dependents in home
MAJORDRG	= number of major derogatory reports (60 days late) currently in credit history
MINORDRG	= number of minor derogatory reports (30 days late) in credit history
OWNRENT	= 0 rents home, 1 owns home
INCOME	= base income (another variable, ADDLINCM = additional income, is omitted)
SELFEMPL	= 0 employed by others (e.g., corp.), 1 = self employed

Credit evaluation agencies such as Fair Isaacs, use data such as these to to help credit card vendors decide whether to accept or reject an application for credit. The preceding are real data from such a process. There is an actual formula, known only to Fair Isaacs, that produces the CARDHLDR variable. You are going to use the data above to approximate the rule.

a. Produce descriptive statistics for the data.
b. Use these data to develop a binary choice model (probit or logit) for the dependent variable CARDHLDR. Report all relevant results and describe your findings to your reader.
c. Compute and report the partial effects for the variables in your model. As part of your presentation, describe how partial effects are computed for your model (theoretically).
d. Standard thinking in this industry suggests that the two variables OWNRENT and SELFEMPL are very important in any acceptance/rejection model. Do your results appear to support this idea? Explain.
e. Compute estimates of a "linear probability model" and compare your least squares results to the partial effects computed in part c. Report all relevant results.

8. Maximum Likelihood Estimation of a Loglinear Model

In analyzing skewed income data such as those shown in the histogram below,



it is customary to analyze logs of income with conventional regression methods. Suppose, in an attempt to impress my colleagues with my facility with 'loglinear models,' I propose, instead to analyze *Income*, not log*Income*, in the context of a *gamma regression model*. In this model, the conditional density for *Income* is

$$f(Income_i \mid \mathbf{x}_i) = \frac{\lambda_i^P Income_i^{P-1} \exp(-\lambda_i Income_i)}{\Gamma(P)},\ \lambda_i = \exp(-\boldsymbol{\beta}'\mathbf{x}_i),\ Income_i > 0,\ P > 0.$$

(Note the reversal of the sign of $\boldsymbol{\beta}$ in λ_i.) The model is 'loglinear' in that E[$Income_i$ |$\mathbf{x}_i$] = P/λ_i, so that the log of the mean is

$$\ln E[Income_i \mid \mathbf{x}_i] = \ln P - \ln\lambda_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i.$$ (Note, **x** does contain a constant term.)

It will be assumed that $\mathbf{x}_i$ contains a constant term as well as covariates such as *age*, *education* and *gender*, so that the log of the mean is $\boldsymbol{\delta}'\mathbf{x}_i$ where the element in $\boldsymbol{\delta}$ that corresponds to a constant is $\gamma = \ln P + \beta_1$ and the other elements are β_k. The parameters to be estimated are *P* and the elements of $\boldsymbol{\beta}$. (P is known as the 'shape parameter.' If *P* is less than or equal to 1, then the distribution looks like the exponential while if it is greater than one, it looks like chi squared.)

a. Derive the log likelihood function for maximum likelihood estimation of *P* and $\boldsymbol{\beta}$. (Note, the log likelihood involves the function $\ln\Gamma(P)$. You can just leave it in this form.)
b. Obtain the likelihood equations for estimation of *P* and $\boldsymbol{\beta}$. (Hint: Use the chain rule. Obtain the derivative with respect to λ_i. Then, the derivative of λ_i with respect to $\boldsymbol{\beta}$ is $\lambda_i\mathbf{x}_i$.)
c. Use the likelihood equations to show that E[$Income_i$ |$\mathbf{x}_i$] = P/λ_i and E[ln$Income_i$|$\mathbf{x}_i$] = $\Psi(P) - \ln\lambda_i$ where $\Psi(P)$ (which is called the 'psi function' or the 'digamma function') is $d\ln\Gamma(P)/dP$. (Hint: $E[\partial\ln L/\partial\lambda_i|x_i] = 0$ and $E[\partial\ln L/\partial P|\mathbf{x}_i] = 0$.)
d. Contining to manipulate the first order conditions, show that the solution for *P* is $P = (1/n)\Sigma_i\, \lambda_i\, Income_i$. Insert this solution for *P* into the log likelihood function to obtain the *concentrated log likelihood* which is only a function of the data and the unknown $\boldsymbol{\beta}$. (Note that if λ_i were a constant, the solution would be $P/\lambda = \overline{Income}$, which makes sense.
e. Derive an estimator for the asymptotic covariance matrix of the MLE of (P,$\boldsymbol{\beta}$). Hint: this will involve $d^2\ln\Gamma(P)/dP^2 = \Psi'(P)$. This is called the 'trigamma function.' Just leave the function in this form.

Several sets of results are given below, where the estimated models are based on the German health care data that we have discussed in class. The dependent variable is *hhninc* = household income. The first two are maximum likelihood estimates of the gamma loglinear model. Use the results provided to answer the following questions:

f. The first set of results (Results [1]) provides unrestricted estimates of the gamma model using the full sample. Using these results, test the hypothesis that AGE is not a significant determinant of income.

g. Using the first set of results, test the hypothesis that all five slope coefficients in the model are jointly equal to zero.

h. Using the first set of results, test the hypothesis of the exponential model as a restriction on the gamma model. The restriction is $P = 1$. The third set of results is the maximum likelihood estimates of the exponential model – that is, the restriction $P = 1$ is imposed. You can use a Wald test based on Results 1 and a likelihood ratio test using Results [1]. and Results [3] to test the hypothesis. Carry out the test both ways. Do you get the same answer? Report your results.

i. Show that the set of partial effects in this gamma regression model are

$$\partial E[\text{Income}|\mathbf{x}]/\partial \mathbf{x} = \partial(P/\lambda_i)/\partial \mathbf{x} = (-P/\lambda_i^2)\lambda_i(-\boldsymbol{\beta}) = E[\text{Income}|\mathbf{x}] \times \boldsymbol{\beta}$$

That is, the slopes of the mean are equal to $\boldsymbol{\beta}$ times the mean.

j. Means of the variables in the model are given below. Using the estimated parameters, compute the partial effects for AGE, EDUC and MARRIED at the means of the data. (Hint: MARRIED is a dummy variable.)

k. Show how you would compute standard errors for the partial effects in part j. (You don't actually have to do the computations. Just show precisely how it would be done.)

l. The second set of results below adds a quadratic term in AGE to the gamma model. I am interested in the age profile of incomes. At what age does Income reach its maximum? (Hint: the log function is a monotonic function of Income, so you can answer this by finding the AGE at which the log of expected Income reaches its maximum. The expression given earlier for log E[$Income_i$ |$\mathbf{x}_i$] will be extremely useful. Now that you have found AGE*, the AGE at which income is maximized, use the delta method to compute an asymptotic standard error for your estimator of AGE*. (Tip: It is sufficient to list the specific computations (with values) – you need not do the full calculation.)

m. The final set of results given below shows the linear regression of *Income* on the constant and the same variables used in the first model. The coefficient estimates in regression 5 are completely different from those in Results [1] Your critical colleagues is by now really upset – something is obviously drastically wrong. OLS is always robust! Can you suggest what might explain this semingly contradictory finding?

Variable	Mean	Standard Deviation	Minimum	Maximum	Cases	Missing Values
INCOME	.352135	.176857	.0015	3.0671	27326	0
AGE	43.52569	11.33025	25.0	64.0	27326	0
EDUC	11.32063	2.324885	7.0	18.0	27326	0
FEMALE	.478775	.499558	0.0	1.0	27326	0
MARRIED	.758618	.427929	0.0	1.0	27326	0
HHKIDS	.40273	.490456	0.0	1.0	27326	0
AGE	* AGE					
Intrct01	2022.855	1004.078	625.0	4096.0	27326	0

Results [1]. Gamma (Loglinear) Regression Model

Gamma (Loglinear) Regression Model

Dependent variable	INCOME
Log likelihood function	14296.44965
Restricted log likelihood	1195.24508
Chi squared [6](P= .000)	26202.40916
Significance level	.00000
McFadden Pseudo R-squared	-10.9611032
Estimation based on N = 27326, K = 7	
Inf.Cr.AIC = -28578.9 AIC/N = -1.046	

INCOME	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Parameters in conditional mean function					
Constant	-3.40833***	.02154	-158.24	.0000	-3.45055	-3.36612
AGE	-.00205***	.00028	-7.41	.0000	-.00260	-.00151
EDUC	.05571***	.00120	46.50	.0000	.05336	.05805
FEMALE	.00539	.00545	.99	.3223	-.00529	.01608
MARRIED	.26343***	.00692	38.05	.0000	.24986	.27700
HHKIDS	-.06514***	.00618	-10.55	.0000	-.07725	-.05303
	Scale parameter for gamma model					
P_scale	5.12550***	.04250	120.60	.0000	5.04220	5.20880

***, **, * ==> Significance at 1%, 5%, 10% level.

Matrix - VARB

[6, 6] Cell:

	1	2	3	4	5	6
1	0.000463945	-3.46745e-006	-1.8487e-005	-2.39605e-005	-4.28454e-006	-3.25856e-005
2	-3.46745e-006	7.68145e-008	3.42467e-008	-7.39586e-008	-6.66741e-007	6.89189e-007
3	-1.8487e-005	3.42467e-008	1.43525e-006	1.15622e-006	2.85161e-007	-5.34441e-008
4	-2.39605e-005	-7.39586e-008	1.15622e-006	2.97049e-005	9.88186e-008	-5.12724e-007
5	-4.28454e-006	-6.66741e-007	2.85161e-007	9.88186e-008	4.79312e-005	-1.57229e-005
6	-3.25856e-005	6.89189e-007	-5.34441e-008	-5.12724e-007	-1.57229e-005	3.81554e-005

Partial Effects for Gamma Regression Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable

(Delta Method)	Partial Effect	Standard Error	|t|	95% Confidence Interval	
AGE	-.00072	.00010	7.41	-.00092	-.00053
EDUC	.01962	.00043	45.37	.01877	.02047
* FEMALE	.00190	.00192	.99	-.00186	.00566
* MARRIED	.08663	.00215	40.22	.08241	.09086
* HHKIDS	-.02282	.00215	10.59	-.02704	-.01860

Results [2]. Gamma (Loglinear) Regression Model

```
-----------------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                INCOME
Log likelihood function      14713.16610
Restricted log likelihood     1195.24508
Chi squared [   7](P= .000)  27035.84205
Significance level                .00000
McFadden Pseudo R-squared    -11.3097483
Estimation based on N =   27326, K =   8
Inf.Cr.AIC  = -29410.3 AIC/N =    -1.076
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  INCOME|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function.......................
Constant|   -4.59494***     .04470  -102.80  .0000    -4.68255   -4.50733
     AGE|     .05827***     .00208    28.05  .0000      .05420     .06234
 AGE*AGE|    -.00069***   .2360D-04  -29.30  .0000     -.00074    -.00065
    EDUC|     .05322***     .00118    44.93  .0000      .05090     .05554
  FEMALE|     .00067        .00538      .12  .9014     -.00987     .01120
 MARRIED|     .22995***     .00694    33.14  .0000      .21635     .24355
  HHKIDS|    -.10564***     .00627   -16.85  .0000     -.11793    -.09335
        |Scale parameter for gamma model...............................
 P_scale|    5.27460***     .04377   120.50  .0000     5.18881    5.36039
--------+--------------------------------------------------------------------
nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Matrix - VARB

[7, 7] Cell:

	1	2	3	4	5	6	7
1	0.00199799	-8.44082e-005	9.30278e-007	-1.44601e-005	-1.70351e-005	4.20546e-005	2.45283e-005
2	-8.44082e-005	4.31694e-006	-4.86325e-008	-1.4896e-007	-3.96158e-007	-3.06371e-006	-2.26639e-006
3	9.30278e-007	-4.86325e-008	5.56953e-010	2.07844e-009	3.69646e-009	2.77275e-008	3.3556e-008
4	-1.44601e-005	-1.4896e-007	2.07844e-009	1.40323e-006	1.13681e-006	3.59697e-007	9.12161e-008
5	-1.70351e-005	-3.96158e-007	3.69646e-009	1.13681e-006	2.88976e-005	3.26851e-007	-3.77517e-007
6	4.20546e-005	-3.06371e-006	2.77275e-008	3.59697e-007	3.26851e-007	4.8143e-005	-1.37651e-005
7	2.45283e-005	-2.26639e-006	3.3556e-008	9.12161e-008	-3.77517e-007	-1.37651e-005	3.93058e-005

Results [3]. Exponential (Loglinear) Regression Model

```
-----------------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                INCOME
Log likelihood function       1550.20104
Restricted log likelihood     1195.24508
Chi squared [   6](P= .000)    709.91193
Significance level               .00000
McFadden Pseudo R-squared      -.2969734
Estimation based on N =  27326, K =    6
Inf.Cr.AIC  =  -3088.4 AIC/N =    -.113
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function.........................
Constant|   -1.77410***      .04501   -39.42  .0000    -1.86231  -1.68589
     AGE|    -.00205***      .00063    -3.27  .0011     -.00328   -.00083
    EDUC|     .05571***      .00271    20.54  .0000      .05039    .06102
  FEMALE|     .00539         .01234      .44  .6620     -.01879    .02958
 MARRIED|     .26343***      .01567    16.81  .0000      .23271    .29415
  HHKIDS|    -.06514***      .01398    -4.66  .0000     -.09255   -.03773
        |Scale parameter for gamma model.................................
 P_scale|        1.0      .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

Results [4]. Ordinary least squares regression ....

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=INCOME  Mean                 =        .35214
            Standard deviation   =        .17686
----------  No. of observations  =         27326  DegFreedom   Mean square
Regression  Sum of Squares       =       92.1787           5      18.43574
Residual    Sum of Squares       =       762.503       27320        .02791
Total       Sum of Squares       =       854.682       27325        .03128
----------  Standard error of e  =        .16706  Root MSE           .16704
Fit         R-squared            =        .10785  R-bar squared      .10769
Model test  F[  5, 27320]        =     660.54085  Prob F > F*        .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
Constant|     .06206***      .00760     8.17  .0000      .04717    .07695
     AGE|    -.00030***      .00010    -2.90  .0037     -.00050   -.00010
    EDUC|     .02143***      .00045    47.75  .0000      .02055    .02231
  FEMALE|     .00496**       .00206     2.40  .0162      .00092    .00900
 MARRIED|     .08706***      .00260    33.44  .0000      .08196    .09216
  HHKIDS|    -.01994***      .00238    -8.38  .0000     -.02460   -.01527
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```