NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS

Department of Economics

ECONOMETRICS I
Take Home Final Examination

Fall 2015

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

Today is Tuesday, December 8, 2015. This exam is due by 10AM, Tuesday, December 22, 2015.

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 test comprises 275 points. The allocation of points to the 10 questions is as follows:

1.	20
2.	20
3.	40
4.	20
5.	20
6.	25
7.	40
8.	25
9.	40
10.	25

1. 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.
b. What are the finite sample properties of this estimator? Make your assumptions explicit.
c. What are the asymptotic properties of the least squares estimator? Again, be explicit about all assumptions, and explain your answer carefully.
d. How would you compare the properties of the least absolute deviations (LAD) estimator to those of the ordinary least squares (OLS) estimator? Which is a preferable estimator?

2. The OLS regression results given below are based on the Baltagi-Griffin OECD gasoline market data. The LHS variable is log(per capita gasoline consumption). The RHS variables are logs of per capita income, the price index of gasoline and the per capita number of cars in the country. In the first set of results, the standard errors are computed using White's heteroscedasticity consistent, robust estimator of the covariance matrix. The second set use the conventional estimator, $s^2(\mathbf{X}'\mathbf{X})^{-1}$. The third set are "clustered" at the country level. (There are 18 countries and 19 years of data (1960-1978).)

a. Explain what is meant by the term "robust covariance estimator." Why would one report "robust standard errors?"
b. How is the White estimator computed?
c. Looking at these results, would you conclude that there is evidence of heteroscedasticity in these data?
d. How is the cluster corrected covariance matrix computed? Can you draw a conclusion about the data based on these results?

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LGASPCAR Mean                 =       4.29624
             Standard deviation   =        .54891
             Number of observs.   =           342
Model size   Parameters           =             4
             Degrees of freedom   =           338
Residuals    Sum of squares       =       14.9044
             Standard error of e  =        .20999
Fit          R-squared            =        .85494
             Adjusted R-squared   =        .85365
Model test   F[  3,    338] (prob) =   664.0(.0000)
White heteroscedasticity robust covariance matrix.
Br./Pagan LM Chi-sq [   3]  (prob) =     .83 (.8427)
--------+--------------------------------------------------------------------
        |              Standard            Prob.      95% Confidence
LGASPCAR|  Coefficient   Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|    2.39133***    .11795    20.27  .0000     2.16015   2.62250
LINCOMEP|     .88996***    .04429    20.09  .0000      .80315    .97677
   LRPMG|    -.89180***    .03891   -22.92  .0000     -.96806   -.81554
LCARPCAP|    -.76337***    .02153   -35.46  .0000     -.80557   -.72118
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
Conventional Standard Errors
--------+--------------------------------------------------------------------
Constant|    2.39133***    .11693    20.45  .0000     2.16214   2.62051
LINCOMEP|     .88996***    .03581    24.86  .0000      .81978    .96014
   LRPMG|    -.89180***    .03031   -29.42  .0000     -.95121   -.83238
LCARPCAP|    -.76337***    .01861   -41.02  .0000     -.79984   -.72690
--------+--------------------------------------------------------------------
```

```
+--------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.   |
| Sample of    342 observations contained     18 clusters defined by |
| variable COUNTRY  which identifies by a value a cluster ID.        |
+--------------------------------------------------------------------+
--------+------------------------------------------------------------------
Constant|    2.39133***      .44167      5.41  .0000      1.52567   3.25698
LINCOMEP|     .88996***      .17248      5.16  .0000       .55190   1.22803
   LRPMG|    -.89180***      .14578     -6.12  .0000     -1.17753   -.60607
LCARPCAP|    -.76337***      .06985    -10.93  .0000      -.90028   -.62647
--------+------------------------------------------------------------------
```

3. The regressions on the next page are computed using the gasoline data used in question 2 above. The estimated asymptotic covariance matrix is shown with each set of regression results. (Note the use of scientific notation in the covariance matrix). The first regression is the same as reported in question 2. I suspect that the effect of (log) income is nonlinear in the model, so in the second regression, I have added a quadratic term in log income to the model.

a. Show how each of the values in the box above the coefficient estimates in the first regression is computed, and interpret the value given.

b. Using the results given, form a confidence interval for the true value of the coefficient log price variable.

c. The second set of results given includes the quadratic term in log income. In the first regression, as we might have expected, log income is highly significant. Looking at the second regression, I might conclude that the quadratic model has revealed that income is not significant. Would this be the correct conclusion? Explain.

d. Test the null hypothesis of the log-linear model against the alternative of the log-quadratic model. Do the test in three ways: 1. Use a Wald test; 2. Use an F test. 3. Use a likelihood ratio test assuming that the disturbances are normally distributed.

e. I am interested in the partial of log Income. As such, the quantity

$$\delta = \partial E[\text{lGasPcar} \mid \mathbf{x}] / \partial\, \text{lincomep}$$

is of interest. Obtain the expression for this function based on the second regression. Estimate this at the average value of lincomep = -6.139425. Form a confidence interval for the estimate of δ.

f. Describe in detail how to use the method of Krinsky and Robb to obtain the standard error needed to compute the confidence interval in part f.

```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LGASPCAR Mean                 =        4.29624
             Standard deviation   =         .54891
----------   No. of observations  =            342  DegFreedom   Mean square
Regression   Sum of Squares       =        87.8386          3      29.27954
Residual     Sum of Squares       =        14.9044        338        .04410
Total        Sum of Squares       =        102.743        341        .30130
----------   Standard error of e  =         .20999  Root MSE           .20876
Fit          R-squared            =         .85494  R-bar squared      .85365
Model test   F[  3,    338]       =      663.99932  Prob F > F*        .00000
Diagnostic   Log likelihood       =       50.49289  Akaike I.C.    = -3.10977
             Restricted (b=0)     =     -279.63574  Bayes  I.C.    = -3.06491
             Chi squared [   3]   =      660.25726  Prob C2 > C2* =   .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
LGASPCAR|  Coefficient       Error       t     |t|>T*         Interval
--------+-------------------------------------------------------------------
Constant|    2.39133***      .11693    20.45  .0000     2.16214    2.62051
LINCOMEP|     .88996***      .03581    24.86  .0000      .81978     .96014
   LRPMG|    -.89180***      .03031   -29.42  .0000     -.95121    -.83238
LCARPCAP|    -.76337***      .01861   -41.02  .0000     -.79984    -.72690
--------+-------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
--------+-------------------------------------------------------
Cov.[b^]|          ONE      LINCOMEP         LRPMG      LCARPCAP
--------+-------------------------------------------------------
     ONE|      .0136736      .00236569   -.772904E-03  -.635905E-04
LINCOMEP|     .00236569      .00128206   -.867471E-03  -.558696E-03
   LRPMG|  -.772904E-03  -.867471E-03    .918984E-03   .450369E-03
LCARPCAP|  -.635905E-04  -.558696E-03    .450369E-03   .346269E-03

----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LGASPCAR Mean                 =        4.29624
             Standard deviation   =         .54891
----------   No. of observations  =            342  DegFreedom   Mean square
Regression   Sum of Squares       =        87.9772          4      21.99431
Residual     Sum of Squares       =        14.7657        337        .04382
Total        Sum of Squares       =        102.743        341        .30130
----------   Standard error of e  =         .20932  Root MSE           .20779
Fit          R-squared            =         .85628  R-bar squared      .85458
Model test   F[  4,    337]       =      501.97922  Prob F > F*        .00000
Diagnostic   Log likelihood       =       52.09097  Akaike I.C.    = -3.11326
             Restricted (b=0)     =     -279.63574  Bayes  I.C.    = -3.05720
             Chi squared [   4]   =      663.45343  Prob C2 > C2* =   .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
LGASPCAR|  Coefficient       Error       t     |t|>T*         Interval
--------+-------------------------------------------------------------------
Constant|     .28487       1.18991      .24  .8109    -2.04732    2.61705
LINCOMEP|     .24471        .36449      .67  .5024     -.46968     .95910
        |Constructed variable LINCOMEP^2.0
Intrct01|    -.05083*        .02857    -1.78  .0762     -.10683     .00518
   LRPMG|    -.92413***      .03526   -26.21  .0000     -.99324    -.85501
LCARPCAP|    -.77048***      .01897   -40.61  .0000     -.80767    -.73329
--------+---------------------------------------------------------------------
Cov.[b^]|          ONE      LINCOMEP      Intrct01         LRPMG      LCARPCAP
--------+---------------------------------------------------------------------
     ONE|      1.41589       .431902      .0338359      .0207538     .00466965
LINCOMEP|      .431902       .132854      .0103646     .00573056   .894615E-03
Intrct01|     .0338359      .0103646   .816422E-03   .519295E-03   .114198E-03
   LRPMG|     .0207538     .00573056   .519295E-03     .00124344   .520140E-03
LCARPCAP|    .00466965   .894615E-03   .114198E-03   .520140E-03   .360039E-03
```

4. In our discussion of the generalized regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0,\ E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\ |\mathbf{X}] = \sigma^2\boldsymbol{\Omega}$$

we arrived at the result of the Aitken theorem, that the optimal estimator of $\boldsymbol{\beta}$ was GLS,

$$\mathbf{b}_{GLS} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

The estimator is optimal in the sense that the Gauss Markov theorem applies – its variance is smaller than any other linear unbiased estimator, including OLS,

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

In the case of the random effects model for panel data,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i,\ i = 1,\ldots,n,\ t = 1,\ldots,T\ \text{(balanced panel)}$$

The covariance matrix for the disturbances is block diagonal with each block in the matrix equal to the T×T matrix

$$\Omega_i = \sigma^2\begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \rho & \ldots \\ \ldots & \rho & \ldots & \rho \\ \rho & \rho & \ldots & 1 \end{bmatrix},\ \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,\ \rho = \sigma_u^2 / \sigma^2.$$

The disturbances are homoscedastic but every pair of disturbances within the ith group has the same correlation.

a. Prove that the generalized least squares estimator is obtained by regressing $y_{it} - \theta\,\bar{y}_i$ on $x_{it} - \theta\,\bar{\mathbf{x}}_i$. (Hint, it suffices to prove that Ω_i^{-1} = the result given in your text, then multiply the ith group, y_i and X_i by the resulting matrix.

5. The three sets of results below show the least squares estimates of the model

$$\text{Health} = \beta_1 + \beta_2\,\text{Age} + \beta_3\,\text{Educ} + \beta_4\,\text{Income} + \varepsilon$$

(We are ignoring the problems with this equation that are discussed in question 8.) Results are given for male headed households (female=0), female headed households (female=1) and all households (female=*).

a. Theory 1 states that the coefficient vectors are the same for the two genders. Is there an optimal way that I could combine these two estimators to form a single efficient estimator of the model parameters? How should I do that? Describe the computations in detail.

b. Use a Chow test to test the hypothesis that the two coefficient vectors are the same. Explain the computations in full detail so that I know exactly how you obtained your result.

c. Show in detail how to use a Wald test to test the hypothesis that the coefficients are the same.

d. Is there any particular reason to use the Wald test or the Chow test – i.e., one and not the other? What assumptions would justify each. Do the regression results suggest that one or the other test might be appropriate? Explain.

```
--> regr;for[female=*,0,1]; lhs=hsat;rhs=x;cov$
-----------------------------------------------------
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   27322
observations, the following counts were found:
Subsample   Observations    Subsample  Observations
FEMALE   =   0     14240    FEMALE  =   1     13082
FEMALE   =****     27322
----------------------------------------------------
-----------------------------------------------------------------
Subsample analyzed for this command is FEMALE   =       0
-----------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HSAT     Mean                 =        6.92428
             Standard deviation   =        2.25170
----------   No. of observations  =          14240  DegFreedom   Mean square
Regression   Sum of Squares       =        5386.20          3    1795.40049
Residual     Sum of Squares       =        66807.6      14236       4.69286
Total        Sum of Squares       =        72193.8      14239       5.07014
----------   Standard error of e  =        2.16630  Root MSE        2.16600
Fit          R-squared            =         .07461  R-bar squared     .07441
Model test   F[  3, 14236]        =      382.58133  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    7.90985***      .11779    67.15  .0000     7.67898   8.14071
     AGE|    -.04923***      .00162   -30.33  .0000     -.05241   -.04605
    EDUC|     .07188***      .00785     9.15  .0000      .05649    .08727
  HHNINC|     .75521***      .11001     6.87  .0000      .53960    .97082
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
--------+--------------------------------------------------------
Cov.[b^]|           ONE            AGE           EDUC         HHNINC
--------+--------------------------------------------------------
     ONE|      .0138745  -.123167E-03  -.688251E-03  -.610555E-03
     AGE|  -.123167E-03   .263389E-05   .139160E-05  -.153205E-04
    EDUC|  -.688251E-03   .139160E-05   .616681E-04  -.262788E-03
  HHNINC|  -.610555E-03  -.153205E-04  -.262788E-03      .0121018

-----------------------------------------------------------------
Subsample analyzed for this command is FEMALE   =       1
-----------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HSAT     Mean                 =        6.63407
             Standard deviation   =        2.32957
----------   No. of observations  =          13082  DegFreedom   Mean square
Regression   Sum of Squares       =        4241.14          3    1413.71331
Residual     Sum of Squares       =        66748.2      13078       5.10385
Total        Sum of Squares       =        70989.3      13081       5.42690
----------   Standard error of e  =        2.25917  Root MSE        2.25883
Fit          R-squared            =         .05974  R-bar squared     .05953
Model test   F[  3, 13078]        =      276.98943  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    7.31752***      .14966    48.89  .0000     7.02419   7.61085
     AGE|    -.04138***      .00180   -22.96  .0000     -.04491   -.03785
    EDUC|     .08589***      .00989     8.68  .0000      .06650    .10527
  HHNINC|     .64726***      .11231     5.76  .0000      .42714    .86739
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
--------+--------------------------------------------------------
Cov.[b^]|           ONE            AGE           EDUC         HHNINC
--------+--------------------------------------------------------
```

```
     ONE|      .0223980  -.192149E-03    -.00117683    -.00192307
     AGE|  -.192149E-03   .324844E-05   .432051E-05   .197074E-05
    EDUC|    -.00117683   .432051E-05   .978490E-04  -.230813E-03
  HHNINC|    -.00192307   .197074E-05  -.230813E-03      .0126137
-----------------------------------------------------------------
Full pooled sample is used for this iteration.
-----------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HSAT     Mean                 =        6.78532
             Standard deviation   =        2.29386
----------   No. of observations  =          27322  DegFreedom   Mean square
Regression   Sum of Squares       =        10033.8           3    3344.60308
Residual     Sum of Squares       =        133724.       27318       4.89507
Total        Sum of Squares       =        143757.       27321       5.26179
----------   Standard error of e  =        2.21248  Root MSE         2.21232
Fit          R-squared            =         .06980  R-bar squared     .06969
Model test   F[  3, 27318]        =      683.25957  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |               Standard            Prob.      95% Confidence
    HSAT|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     7.61481***      .09226    82.54  .0000     7.43399   7.79563
     AGE|     -.04585***      .00120   -38.21  .0000     -.04821   -.04350
    EDUC|      .08170***      .00606    13.48  .0000      .06982    .09358
  HHNINC|      .68599***      .07850     8.74  .0000      .53213    .83986
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
--------+--------------------------------------------------------
Cov.[b^]|           ONE           AGE          EDUC        HHNINC
--------+--------------------------------------------------------
     ONE|     .00851141  -.757535E-04  -.427036E-03  -.570189E-03
     AGE|  -.757535E-04   .143995E-05   .127524E-05  -.386028E-05
    EDUC|  -.427036E-03   .127524E-05   .367549E-04  -.126496E-03
  HHNINC|  -.570189E-03  -.386028E-05  -.126496E-03     .00616291
```

6. The results below show OLS, fixed effects and random effects estimates of a loglinear cost function of the form

$$\log(C_{it} / w_5) = \alpha + \sum_{k=1}^{4} \beta_k \log(w_k / w_5) + \sum_{m=1}^{5} \gamma_m \log q_m + \varepsilon_{it} \; + \; u_i$$

Where C is total costs, there are 5 inputs with input prices w_k and 5 outputs denoted q_m, m = 1,…,5. The cost variable and the four inputs in the equation are divided by w_5 to enforce the linear homogeneity constraint $\beta_1+\beta_2+\beta_3+\beta_4+\beta_5 = 1$. The sample is a set of 500 banks observed for 5 years.

a. Test the hypothesis of 'no effects' vs. 'some effects' using the results given below.
b. Explain in precise detail the difference between the fixed and random effects models.
c. What is the result of the Hausman test for fixed vs. random effects? Report your conclusion. Carefully explain what you are doing in this test. Based on your result, which is the preferred model, fixed effects or random effects?
d. In the context of the fixed effects model, test the hypothesis that there are no effects – i.e., that all banks have the same constant term. (The statistics you need to carry out the test are given in the results.)
e. In the final regression below, I have added the bank (5 period) means of the log price ratios and log outputs to the model and reestimated the random effects model. I then used a Wald statistic to test the joint hypothesis that the coefficients on the 9 group means are jointly equal to zero. Are the results consistent or inconsistent with the results in part c? Explain.

```
-----------------------------------------------------------------
OLS Without Group Dummy Variables.................
Ordinary     least squares regression ............
LHS=C        Mean                   =      11.46039
             Standard deviation     =       1.17411
             Number of observs.     =          2500
Model size   Parameters             =            10
             Degrees of freedom     =          2490
Residuals    Sum of squares         =     156.19259
             Standard error of e    =        .25046
Fit          R-squared              =        .95466
             Adjusted R-squared     =        .95450
Model test   F[  9,  2490] (prob) =  5825.4(.0000)
Diagnostic   Log likelihood         =     -81.15108
             Restricted(b=0)        =   -3948.12242
             Chi-sq [   9]  (prob) =  7733.9(.0000)
Panel Data Analysis of C                 [ONE way]
               Unconditional ANOVA (No regressors)
Source         Variation  Deg. Free.   Mean Square
Between        668.14287       499.       1.33896
Residual      2776.81463      2000.       1.38841
Total         3444.95749      2499.       1.37853
--------+--------------------------------------------------------
        |                  Standard            Prob.       Mean
C       | Coefficient       Error       z    z>|Z|       of X
--------+--------------------------------------------------------
      W1|     .42321***      .01771    23.90  .0000    6.73864
      W2|     .03656***      .00809     4.52  .0000    1.88257
      W3|     .17770***      .01498    11.86  .0000    -.23288
      W4|     .10602***      .01174     9.03  .0000    -.68155
      Q1|     .10335***      .00745    13.87  .0000    8.58763
      Q2|     .37493***      .00709    52.89  .0000    10.0932
      Q3|     .09658***      .00965    10.01  .0000    9.71949
      Q4|     .05624***      .00400    14.05  .0000    7.78290
      Q5|     .28603***      .00969    29.51  .0000    7.13716
Constant|     .56364***      .12344     4.57  .0000
--------+--------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------
```

```
-----------------------------------------------------------------
Least Squares with Group Dummy Variables..........
Ordinary     least squares regression ............
LHS=C        Mean                 =       11.46039
             Standard deviation   =        1.17411
             Number of observs.   =           2500
Model size   Parameters           =            509
             Degrees of freedom   =           1991
Residuals    Sum of squares       =      120.89104
             Standard error of e  =         .24641
Fit          R-squared            =         .96491
             Adjusted R-squared   =         .95595
Model test   F[508,  1991] (prob) =   107.8(.0000)
Diagnostic   Log likelihood       =      239.09913
             Restricted(b=0)      =    -3948.12242
             Chi-sq [508]  (prob) =  8374.4(.0000)
Estd. Autocorrelation of e(i,t)   =       -.233637
Panel:Groups Empty      0,     Valid data      500
             Smallest   5,     Largest           5
             Average group size in panel      5.00
--------+--------------------------------------------------------
        |                  Standard          Prob.       Mean
C       | Coefficient        Error       z    z>|Z|       of X
--------+--------------------------------------------------------
      W1|      .40903***      .01951    20.97  .0000    6.73864
      W2|      .04421***      .00891     4.96  .0000    1.88257
      W3|      .18063***      .01652    10.93  .0000    -.23288
      W4|      .11294***      .01289     8.77  .0000    -.68155
      Q1|      .10693***      .00820    13.04  .0000    8.58763
      Q2|      .37667***      .00785    47.99  .0000    10.0932
      Q3|      .10037***      .01070     9.38  .0000    9.71949
      Q4|      .05536***      .00440    12.57  .0000    7.78290
      Q5|      .27849***      .01086    25.64  .0000    7.13716
--------+--------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------

+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model           Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only    -3948.12238        3444.95749     .00000 |
|(2)  Group effects only    -3678.61349        2776.81463     .19395 |
|(3)  X - variables only      -81.15104         156.19259     .95466 |
|(4)  X and group effects     239.09916         120.89104     .96491 |
+--------------------------------------------------------------------+
|                        Hypothesis Tests                            |
|         Likelihood Ratio Test          F Tests                     |
|         Chi-squared   d.f.   Prob         F    num   denom  P value |
|(2) vs (1)     539.02    499  .1042       .96    499    2000   .70821 |
|(3) vs (1)    7733.94      9  .0000   5825.45      9    2490   .00000 |
|(4) vs (1)    8374.44    508  .0000    107.77    508    1991   .00000 |
|(4) vs (2)    7835.43      9  .0000   4860.16      9    1991   .00000 |
|(4) vs (3)     640.50    499  .0000      1.17    499    1991   .00625 |
+--------------------------------------------------------------------+
```

```
----------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .060719
            Var[u]              =       .002009
            Corr[v(i,t),v(i,s)] =       .032030
Lagrange Multiplier Test vs. Model (3) =   4.93
( 1 degrees of freedom, prob. value =  .026436)
(High values of LM favor FEM/REM over CR model)
Fixed vs. Random Effects (Hausman)     =  10.17
( 9 degrees of freedom, prob. value =  .337051)
(High (low) values of H favor F.E.(R.E.) model)
            Sum of Squares           156.194972
            R-squared                   .954660
--------+-------------------------------------------------------------
        |                Standard            Prob.       Mean
C       | Coefficient       Error       z    z>|Z|       of X
--------+-------------------------------------------------------------
      W1|     .42157***      .01768    23.85  .0000    6.73864
      W2|     .03744***      .00807     4.64  .0000    1.88257
      W3|     .17802***      .01495    11.90  .0000    -.23288
      W4|     .10687***      .01172     9.12  .0000    -.68155
      Q1|     .10376***      .00744    13.95  .0000    8.58763
      Q2|     .37513***      .00708    52.98  .0000    10.0932
      Q3|     .09700***      .00964    10.07  .0000    9.71949
      Q4|     .05614***      .00400    14.05  .0000    7.78290
      Q5|     .28516***      .00969    29.42  .0000    7.13716
Constant|     .57111***      .12316     4.64  .0000
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
----------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .060994
            Var[u]              =       .001679
            Corr[v(i,t),v(i,s)] =       .026794
Lagrange Multiplier Test vs. Model (3) =   3.97
( 1 degrees of freedom, prob. value =  .046403)
(High values of LM favor FEM/REM over CR model)
--------+-------------------------------------------------------------
        |                Standard            Prob.       Mean
C       | Coefficient       Error       z    z>|Z|       of X
--------+-------------------------------------------------------------
     W1B|     .06958         .04640     1.50  .1337    6.73864
     W2B|    -.03553*        .02113    -1.68  .0927    1.88257
     W3B|    -.01122         .03875     -.29  .7721    -.23288
     W4B|    -.04204         .03102    -1.36  .1753    -.68155
     Q1B|    -.01657         .01932     -.86  .3912    8.58763
     Q2B|    -.00843         .01821     -.46  .6435    10.0932
     Q3B|    -.01524         .02463     -.62  .5362    9.71949
     Q4B|     .00441         .01042      .42  .6724    7.78290
     Q5B|     .03734         .02409     1.55  .1212    7.13716
      W1|     .40903***      .01955    20.92  .0000    6.73864
      W2|     .04421***      .00893     4.95  .0000    1.88257
      W3|     .18063***      .01656    10.91  .0000    -.23288
      W4|     .11294***      .01291     8.75  .0000    -.68155
      Q1|     .10693***      .00822    13.01  .0000    8.58763
      Q2|     .37667***      .00787    47.88  .0000    10.0932
      Q3|     .10037***      .01073     9.36  .0000    9.71949
      Q4|     .05536***      .00441    12.55  .0000    7.78290
      Q5|     .27849***      .01089    25.58  .0000    7.13716
Constant|     .26717         .29913      .89  .3718
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
--> matrix ; wub=b(1:9);vwub=varb(1:9,1:9)$
--> matrix ; list;wustat = wub'<vwub>wub$
Matrix WUSTAT   has  1 rows and  1 columns.
               1
        +-------------+
       1|   10.32455
        +-------------+
```

7. The data for this exercise are parked on the web on my home page in 4 formats:

http://people.stern.nyu.edu/wgreene/labor.lpj (limdep or nlogit project file)
http://people.stern.nyu.edu/wgreene/labor.dta (stata data file)
http://people.stern.nyu.edu/wgreene/labor.csv (Excel comma separated values)
http://people.stern.nyu.edu/wgreene/labor.txt (Text format. (Same as csv))

All three files contain 753 observations on 5 variables relating to a study of labor market behavior of married women.

LFP = the dependent variable, labor force participation coded 0 and 1
AGE = age
WE=wife's education in years
FAMINC=family income
KIDS = dummy variable for whether there are kids in the home.

The raw data look like these, which are the first few observations.

LFP,AGE,WE,FAMINC,KIDS

```
   1.00000       32.0000       12.0000       16310.0       1.00000
   1.00000       30.0000       12.0000       21800.0       1.00000
   1.00000       35.0000       12.0000       21040.0       1.00000
   1.00000       34.0000       12.0000       7300.00       1.00000
   1.00000       31.0000       14.0000       27300.0       1.00000
   1.00000       54.0000       12.0000       19495.0       .000000
   1.00000       37.0000       16.0000       21152.0       1.00000
   1.00000       54.0000       12.0000       18900.0       .000000
```

You will need a statistical package to do this part of the exam. The .csv file can be read directly into Excel without conversion. The .lpj file is an nlogit or limdep project file. The .txt and .dta files are suitable for export to Stata. The .csv or .txt file can be exported to R.

a. Your assigment is to estimate a binary choice model using these data. Your model should explain LFP using age, education and family income. (You may fit a probit model or a logit model – your choice. Indicate in your report which form you used.) As part of your analysis, compute the partial effect on the probability of participation in the formal labor market of an additional year of education and of an additional thousand dollars in family income.. Report your result, and interpret it. Use least squares to fit a "linear probability model," and compare your results to your maximum likelihood based results for the probit or logit model.

b. One might think that the presence of children in the household would completely change the labor force participation decision. Split the sample based on the KIDS dummy variable, and compute the two probit (or logit) models for the subsamples. Carry out a likelihood ratio test of the hypothesis that pooling is valid versus the alternative that separate models apply to the two subsamples.

8. I propose to estimate the model

$$E[\text{Income}|\text{Age},\text{Education},\text{Health}] = \exp(\beta_1 + \beta_2\text{Age} + \beta_3\text{Educ} + \beta_4\text{Health})$$

However, I have been convinced by my colleagues that in a model of income determination, Health would be endogenous. So, added to the complication that my model is nonlinear is the complication that one of the right hand side variables is endogenous.

a. Explain what is meant by 'endogenous' in this discussion. Why would endogeneity be a problem?

b. I have data on income (hhninc = household income), Age, Education, and HSAT = health satisfaction. I will take HSAT, the individuals assessment of their health, as a reliable proxy for their actual health. At least for purposes of this exercise, I will take Age and Educ to be exogenous. I also have a set of instrumental variables, Z = (married, hhkids, working, female). (All are dummy variables, in fact.) Two sets of results are given below. This first are nonlinear least squares estimates that ignore the endogeneity question. The second are nonlinear instrumental variables estimates.

(i) Explain how the nonlinear least squares estimates are computed.

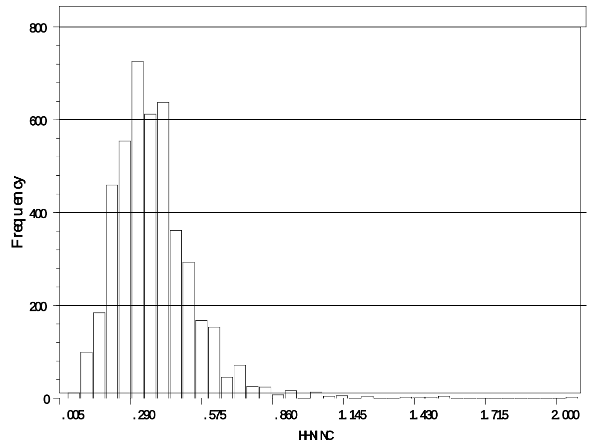
(ii) Explain how the nonlinear instrumental variables are computed.

(iii) I might have improved my estimator by using GMM. Explain how I would use GMM to estimate the parameters of this model.

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ...........
LHS=HHNINC    Mean                  =      .35214
              Standard deviation    =      .17687
              Number of observs.    =       27322
Model size    Parameters            =           4
              Degrees of freedom    =       27318
Residuals     Sum of squares        =     792.505
              Standard error of e   =      .17031
Fit           R-squared             =      .07275
--------+--------------------------------------------------------------------
        |              Standard            Prob.      95% Confidence
UserFunc|  Coefficient   Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -1.82573***   .02135   -85.50  .0000    -1.86759  -1.78388
     Age|     .00260***   .00027     9.72  .0000      .00208    .00312
    Educ|     .05108***   .00106    48.26  .0000      .04900    .05315
    HSAT|     .01219***   .00135     9.05  .0000      .00955    .01483
--------+--------------------------------------------------------------------
Instrumental Variables (NLIV)....................
Nonlinear     least squares regression ...........
Residuals     Sum of squares        =     2144.34
              Standard error of e   =      .28015
Fit           R-squared             =    -1.50894
--------+--------------------------------------------------------------------
        |              Standard            Prob.      95% Confidence
UserFunc|  Coefficient   Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -4.61690***   .70523    -6.55  .0000    -5.99913  -3.23467
     Age|     .01437***   .00225     6.38  .0000      .00995    .01878
    Educ|     .03105***   .00336     9.25  .0000      .02447    .03763
    HSAT|     .34524***   .07902     4.37  .0000      .19037    .50011
--------+--------------------------------------------------------------------
```

9. 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*. That is a model in which the conditional density for *Income* is

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

The model is 'loglinear' in that $E[Income_i \mid \mathbf{x}_i] = P/\lambda_i$, so that the log of the mean is

$$\log E[Income_i \mid \mathbf{x}_i] = \log P - \log\lambda_i = \alpha - \boldsymbol{\beta}'\mathbf{x}_i.$$

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 $\gamma = \log P - \beta_0$ and the other elements are $-\beta_k$. (Note the sign change.) 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 $\log\Gamma(P)$. 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 $E[Income_i \mid \mathbf{x}_i] = P/\lambda_i$ and $E[\log Income_i \mid \mathbf{x}_i] = \Psi(P) - \log\lambda_i$ where $\Psi(P)$ (which is called the 'psi function' or the 'digamma function') is $d\log\Gamma(P)/dP$.

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 β. (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 $d^2\log\Gamma(P)/dP^2 = \Psi'(P)$. 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 provides unrestricted estimates of the gamma model using the full sample (less the four observations for which hhninc = 0). Using these results, test the hypothesis that AGE is not a significant determinant of income.

g. Your colleague who is skeptical of nonlinear models to begin with points out that in your first set of results, the reported value for the 'Pseudo-R^2' is -10.9599753. There is obviously something drastically wrong here – proportions of variance explained are between 0 and 1. On this basis, they dismiss your nonlinear model as obviously wrong (and implore you to use ordinary least squares). How would you answer this criticism?

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

i. 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. Using both the first and third set of results (and a different type of test from that used in part g), test the hypothesis of the exponential model.

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

$$\partial E[\text{Income}|x]/\partial x = -E[\text{Income}|x] \times \beta$$

That is, the slopes of the mean are equal to the negative of beta times the mean.

k. Means of the variables in the model are given at the beginning of the results. 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.)

l. 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.)

m. 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*.

n. 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 regression 1 – in fact, the signs are all opposite and the magnitudes are different. Your critical colleagues is by now really upset – something is obviously drastically wrong. OLS is always robust, and your coefficients all have the wrong signs! Can you suggest what might explain this semingly contradictory finding?

```
Descriptive Statistics for   5 variables
--------+---------------------------------------------------------------------
Variable|       Mean        Std.Dev.      Minimum       Maximum     Cases Missing
--------+---------------------------------------------------------------------
     AGE|     43.52719      11.33032         25.0          64.0     27322        0
 Age*Age|     2022.988      1004.087        625.0        4096.0     27322        0
    EDUC|     11.32018      2.324347          7.0          18.0     27322        0
 MARRIED|      .758693       .427884          0.0           1.0     27322        0
  FEMALE|      .478808       .499560          0.0           1.0     27322        0
  HHKIDS|      .402716       .490453          0.0           1.0     27322        0
--------+---------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
1.   Gamma (Loglinear) Regression Model
Dependent variable                 HHNINC
Log likelihood function       14293.00214
Restricted log likelihood(β=0)1195.06953
Chi squared [   6 d.f.]       26195.86522
Significance level                 .00000
McFadden Pseudo R-squared     -10.9599753
Estimation based on N =  27322,  K =   7
Inf.Cr.AIC  = -28572.0  AIC/N =   -1.046
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     3.40841***      .02154   158.21  .0000     3.36618    3.45063
     AGE|      .00205***      .00028     7.41  .0000      .00151     .00260
    EDUC|     -.05572***      .00120   -46.50  .0000     -.05807    -.05337
 MARRIED|     -.26341***      .00692   -38.04  .0000     -.27698    -.24984
  FEMALE|     -.00542         .00545     -.99  .3198     -.01611     .00526
  HHKIDS|      .06512***      .00618    10.54  .0000      .05302     .07723
        |Scale parameter for gamma model
 P_shape|     5.12486***      .04250   120.59  .0000     5.04157    5.20815
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Matrix - VARB

[7, 7] Cell:

	1	2	3	4	5	6	7
1	0.000464108	-3.46813e-006	-1.8497e-005	-4.27673e-006	-2.39817e-005	-3.26011e-005	0.000352398
2	-3.46813e-006	7.68273e-008	3.42455e-008	-6.66724e-007	-7.38979e-008	6.89342e-007	-4.19768e-013
3	-1.8497e-005	3.42455e-008	1.43625e-006	2.8321e-007	1.15748e-006	-5.28885e-008	-3.62485e-012
4	-4.27673e-006	-6.66724e-007	2.8321e-007	4.79481e-005	9.71058e-008	-1.57257e-005	1.76474e-011
5	-2.39817e-005	-7.38979e-008	1.15748e-006	9.71058e-008	2.97137e-005	-5.10296e-007	1.09561e-011
6	-3.26011e-005	6.89342e-007	-5.28885e-008	-1.57257e-005	-5.10296e-007	3.81658e-005	1.90277e-011
7	0.000352398	-4.19768e-013	-3.62485e-012	1.76474e-011	1.09561e-011	1.90277e-011	0.00180599

```
-----------------------------------------------------------------------------
2.  Gamma (Loglinear) Regression Model
Dependent variable                HHNINC
Log likelihood function      14709.58448
Restricted log likelihood     1195.06953
Chi squared [   7 d.f.]      27029.02989
Significance level                .00000
McFadden Pseudo R-squared    -11.3085595
Estimation based on N =  27322, K =    8
Inf.Cr.AIC  = -29403.2 AIC/N =    -1.076
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    4.59487***      .04470   102.79  .0000     4.50726   4.68249
     AGE|    -.05827***      .00208   -28.04  .0000     -.06234   -.05419
 AGE*AGE|     .00069***   .2360D-04    29.29  .0000      .00065    .00074
    EDUC|    -.05323***      .00119   -44.92  .0000     -.05556   -.05091
 MARRIED|    -.22994***      .00694   -33.13  .0000     -.24354   -.21634
  FEMALE|    -.00068         .00538     -.13  .8993     -.01122    .00986
  HHKIDS|     .10563***      .00627    16.85  .0000      .09334    .11792
        |Scale parameter for gamma model
 P_shape|    5.27392***      .04377   120.49  .0000     5.18813   5.35970
--------+--------------------------------------------------------------------
Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Matrix - VARB

[7, 7] Cell:

	1	2	3	4	5	6	7
1	0.00199829	-8.44201e-005	9.30409e-007	-1.44653e-005	4.20678e-005	-1.70356e-005	2.45288e-005
2	-8.44201e-005	4.31772e-006	-4.86415e-008	-1.49224e-007	-3.064e-006	-3.97192e-007	-2.26723e-006
3	9.30409e-007	-4.86415e-008	5.57059e-010	2.08149e-009	2.77308e-008	3.70895e-009	3.35675e-008
4	-1.44653e-005	-1.49224e-007	2.08149e-009	1.40424e-006	3.57818e-007	1.13815e-006	9.20511e-008
5	4.20678e-005	-3.064e-006	2.77308e-008	3.57818e-007	4.81601e-005	3.25504e-007	-1.3768e-005
6	-1.70356e-005	-3.97192e-007	3.70895e-009	1.13815e-006	3.25504e-007	2.89066e-005	-3.74216e-007
7	2.45288e-005	-2.26723e-006	3.35675e-008	9.20511e-008	-1.3768e-005	-3.74216e-007	3.93173e-005

```
-----------------------------------------------------------------------------
3.  Exponential (Loglinear) Regression Model
Dependent variable                HHNINC
Log likelihood function       1550.07536
Restricted log likelihood     1195.06953
Chi squared [   5 d.f.]        710.01166
Significance level                .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.77430***      .04501    39.42  .0000     1.68608   1.86253
     AGE|     .00205***      .00063     3.27  .0011      .00082    .00328
    EDUC|    -.05572***      .00271   -20.54  .0000     -.06104   -.05040
 MARRIED|    -.26341***      .01568   -16.80  .0000     -.29413   -.23269
  FEMALE|    -.00542         .01234     -.44  .6603     -.02961    .01876
  HHKIDS|     .06512***      .01399     4.66  .0000      .03771    .09254
--------+--------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
4.  Gamma (Loglinear) Regression Model
Gamma (Loglinear) Regression Model
Dependent variable                HHNINC
LM Stat. at start values      9618.48555
LM statistic kept as scalar     LMSTAT
Log likelihood function       1550.07538
Restricted log likelihood     1195.06953
Chi squared [   6 d.f.]        710.01169
Significance level                .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     1.77430***      .04564    38.88  .0000     1.68485   1.86375
     AGE|      .00205***      .00063     3.27  .0011      .00082    .00328
    EDUC|     -.05572***      .00271   -20.54  .0000     -.06104   -.05040
 MARRIED|     -.26341***      .01568   -16.80  .0000     -.29413   -.23269
  FEMALE|     -.00542         .01234     -.44  .6603     -.02961    .01876
  HHKIDS|      .06512***      .01399     4.66  .0000      .03771    .09254
        |Scale parameter for gamma model
 P_shape|         1.0***      .00753   132.74  .0000  .98523D+00  .10148D+01
--------+--------------------------------------------------------------------

-----------------------------------------------------------------------------
5.  Ordinary     least squares regression ............
LHS=HHNINC    Mean                 =          .35214
              Standard deviation   =          .17687
----------    No. of observations  =           27322  DegFreedom   Mean square
Regression    Sum of Squares       =         92.1956          5       18.43911
Residual      Sum of Squares       =         762.486      27316         .02791
Total         Sum of Squares       =         854.682      27321         .03128
----------    Standard error of e  =          .16707  Root MSE          .16706
Fit           R-squared            =          .10787  R-bar squared     .10771
Model test    F[  5, 27316]        =       660.57945  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .06198***      .00760     8.16  .0000      .04709    .07687
     AGE|     -.00030***      .00010    -2.90  .0037     -.00050   -.00010
    EDUC|      .02144***      .00045    47.75  .0000      .02056    .02232
 MARRIED|      .08705***      .00260    33.43  .0000      .08194    .09215
  FEMALE|      .00497**       .00206     2.41  .0160      .00093    .00901
  HHKIDS|     -.01993***      .00238    -8.37  .0000     -.02460   -.01527
--------+--------------------------------------------------------------------
```

10. This question involves a small amount of "library" research. (You can do it on the web, of course.) 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.