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

Department of Economics

ECONOMETRICS I
Take Home Final Examination

Fall 2016

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

Today is Thursday, December 8, 2016. This exam is due by 3PM, Friday, December 16, 2012. You may submit your answers to me electronically as an attachment to an e-mail if you wish. Please do not include a copy of the exam questions with your submission; submit only your answers to the questions. Your submission for this examination is to be a single authored project. You are assumed to be working alone.

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.

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 paper, Farsi, M, M. Filippini, and W. Greene, "Efficiency Measurement in Network Industries, Applicationto the Swiss Railroads," *Journal of Regulatory Economics*, 28, 1, 2005, pp. 69-90 is an analysis of an unbalanced panel of data on 50 railroads for 13 years, 605 observations in total. The variables in the data set are

ct	= total cost
q	= total output, sum of freight, passenger and mail
pe	= price of electricity
pk	= price of capital
pl	= price of labor
narrow	= dummy for narrow gauge track
tunnel	= dummy variable for long tunnels on routes
rack	= dummy variable for a certain track configuration

I propose first to analyze the cost data with a loglinear model. My first model is

$$\begin{aligned} \text{lnc}_{it} = \; & \beta_1 + \beta_2\text{lnq}_{it} + \beta_3\text{lnpe}_{it} + \beta_4\text{lnpl}_{it} + \beta_5\text{lnpk}_{it} + \beta_6\text{narrow}_i + \\ & \beta_7\text{tunnel}_i + \beta_8\text{rack}_i + \varepsilon_{it}, \;\; \varepsilon_{it} \sim N[0, \sigma^2], \end{aligned}$$

where "i" indicates the railroad and "t" indicates the year. Note that some variables are time invariant. For this application, I intend to ignore any panel data aspects of the data set, and treat the whole thing as a cross section of 605 observations. The ordinary least squares results are shown as Regression 1 on page 4.

a. Show how each of the values in the box above the coefficient estimates is computed, and interpret the value given.
b. Using the results given, form a confidence interval for the true value of the coefficient on the RACK dummy variable.
c. An expanded, now nonlinear model appears as follows:

$$\begin{aligned} \text{lnc}_{it} = \; & \beta_1 + \beta_2 \,\text{lnq}_{it} + \beta_3 \ln^2\text{q}_{it} + \beta_4\text{lnpe}_{it} + \beta_5\text{lnpl}_{it} + \beta_6\text{lnpk}_{it} + \beta_7\text{narrow}_i + \\ & \beta_8\text{tunnel}_i + \beta_9\text{rack}_i + \beta_{10}\,\text{lnq}\times\text{tunnel} + \varepsilon_{it}, \;\; \varepsilon_{it} \sim N[0, \sigma^2]. \end{aligned}$$

The second set of results given includes the quadratic specification (β_3) and an interaction of log output with tunnel (β_{10}). Test the hypothesis of the linear model as a restriction on the nonlinear model. Do the test in three ways: 1. Use a Wald test to test the hypothesis that the two coefficients in the quadratic terms (β_3 and β_{10})are zero. 2. Use an F test. 3. Use a likelihood ratio test assuming that the disturbances are normally distributed. The estimated least squares regression for this model is shown on page 5 with the estimated asymptotic covariance matrix.

d. I am interested economies of scale for railroads. In the loglinear equation (regression 1), that quantity is

$$\Delta = \partial \text{lnct}/\partial \text{lnq} = \beta_2$$

In the second model (regression 2), the measure is a linear function of lnq and tunnel;

$$\Delta = \beta_2 + 2\beta_3 \text{ lnq} + \beta_{10}\text{Tunnel}.$$

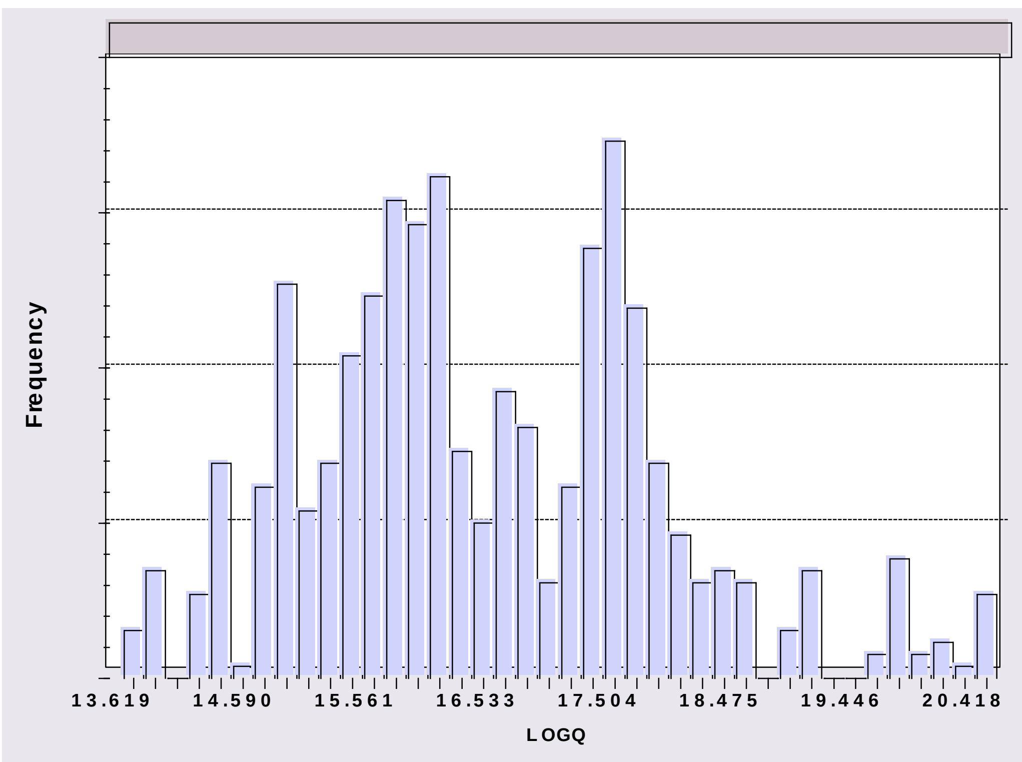
Estimate this value for the average sized railroad with long tunnels (tunnel = 1). (The average of lnq is shown in the regression results.) Form a confidence interval for

$$\Delta \text{ (given average lnq = 16.4616 and tunnel = 1)}.$$

e. The efficient scale for a production model is that point where economies of scale equal 1. Assuming that tunnel = 0, the scale elasticity is

$$\Delta = \beta_2 + 2\beta_3 \text{ lnq}.$$

Solving for $\Delta = 1$, I obtain lnq = $(1 - \beta_2)/(2\beta_3)$. Using the results of regression 2 (and the delta method), form a confidence interval for this function of the parameters. The histogram below shows the distribution of lnq in the sample. Locate the point of constant returns to scale in the graph and comment on the efficient size of firm compared to the values found in the sample.



3. The third set of results (on page 6) is computed using White's heteroscedasticity consistent, robust estimator of the covariance matrix.

a. How is the White estimator computed?

b. Looking at these results, would you conclude that there is evidence of heteroscedasticity in these data?

4. The fourth regression is reported (on page 6) with a correction of the standard errors to accommodate the "clustering" in the data – these data are a panel.

a. How is the cluster estimator computed?

b. Why is it computed; what problem is it intended to solve?

c. Compare the results to regression 2 (which is the same model with conventional standard errors). What do you conclude about the effect of "clustering" in these data?

REGRESSION 1

```
-----------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LNCT     Mean                 =      11.30622
             Standard deviation   =       1.10169
             Number of observs.   =           605
Model size   Parameters           =             8
             Degrees of freedom   =           597
Residuals    Sum of squares       =      48.49299
             Standard error of e  =        .28500
Fit          R-squared            =        .93385
             Adjusted R-squared   =        .93308
Model test   F[  7,   597] (prob) =  1204.0(.0000)
Diagnostic   Log likelihood       =     -95.00555
             Restricted(b=0)      =    -916.54939
             Chi-sq [  7]  (prob) =  1643.1(.0000)
Info criter. Akaike Info. Criter. =      -2.49736
--------+--------------------------------------------------------
        |                  Standard            Prob.       Mean
    LNCT| Coefficient        Error       z    z>|Z|       of X
--------+--------------------------------------------------------
Constant|   -7.78982***     1.98172    -3.93  .0001
     LNQ|     .77200***      .01125    68.64  .0000    16.4616
    LNPL|     .15615         .17391      .90  .3692    13.2194
    LNPE|    -.41790**       .18157    -2.30  .0214   -1.85956
    LNPK|     .34916***      .02891    12.08  .0000    10.1795
  NARROW|    -.13716***      .02699    -5.08  .0000     .67603
    RACK|     .48192***      .03098    15.56  .0000     .23471
  TUNNEL|    -.14985***      .03790    -3.95  .0001     .18843
--------+--------------------------------------------------------
```

REGRESSION 2

```
-----------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LNCT     Mean                 =      11.30622
             Standard deviation   =       1.10169
             Number of observs.   =           605
Model size   Parameters           =            10
             Degrees of freedom   =           595
Residuals    Sum of squares       =      38.53713
             Standard error of e  =        .25450
Fit          R-squared            =        .94743
             Adjusted R-squared   =        .94664
Model test   F[  9,   595] (prob) =  1191.5(.0000)
Diagnostic   Log likelihood       =     -25.49196
             Restricted(b=0)      =    -916.54939
             Chi-sq [  9]  (prob) =  1782.1(.0000)
Info criter. LogAmemiya Prd. Crt. =      -2.72055
             Akaike Info. Criter. =      -2.72055
             Bayes Info. Criter.  =      -2.64773
--------+--------------------------------------------------------
        |                 Standard           Prob.       Mean
    LNCT| Coefficient       Error       z    z>|Z|       of X
--------+--------------------------------------------------------
Constant|    14.1048***     3.08775     4.57  .0000
     LNQ|   -1.81538***      .28753    -6.31  .0000    16.4616
   LNQSQ|     .07876***      .00890     8.85  .0000    272.853
    LNPL|     .14406         .15695      .92  .3587    13.2194
    LNPE|    -.30467*        .16469    -1.85  .0643   -1.85956
    LNPK|     .30383***      .02608    11.65  .0000    10.1795
NARROW_T|     .00122         .02783      .04  .9651     .67603
    RACK|     .42526***      .02867    14.83  .0000     .23471
  TUNNEL|    2.05124***      .75377     2.72  .0065     .18843
 LNQTUNL|    -.13334***      .04345    -3.07  .0021    3.40727
--------+--------------------------------------------------------
 9.53418
-0.723063   0.0826737
 0.0225023 -0.00255569   7.91249e-005
-0.313888   0.00430309  -0.000147886   0.0246349
-0.282299   0.00320699  -0.000112702   0.023335     0.0271221
-0.0147959  0.000883942 -2.83063e-005  0.000107397  0.000194245  0.00068034
 0.0345858 -0.00400616   0.000122773  -5.22415e-005 0.000303053 -0.000125345   0.0007747
-0.0276479  0.00216755  -6.61934e-005  0.000655545  0.000603249  0.000244045  -0.00031242
                                                                               0.00082181
 1.65435   -0.181774     0.00568441   -0.0178937   -0.0196452  - 0.00127023    0.00820813
                                                                -0.00533895    0.568164
-0.0952809  0.0105601   -0.000330446   0.000968367  0.00105588   7.36731e-005 -0.00047073
                                                    0.000300395 -0.0327144     0.00188776
```

REGRESSION 3

```
-----------------------------------------------------------------
Ordinary     least squares regression ...........
LHS=LNCT     Mean                  =      11.30622
             Standard deviation    =       1.10169
             Number of observs.    =           605
Model size   Parameters            =            10
             Degrees of freedom    =           595
Residuals    Sum of squares        =      38.53713
             Standard error of e   =        .25450
Fit          R-squared             =        .94743
             Adjusted R-squared    =        .94664
Model test   F[  9,    595] (prob) =  1191.5(.0000)
White heteroscedasticity robust covariance matrix.
Br./Pagan LM Chi-sq [  9]  (prob) =  58.95 (.0000)
--------+--------------------------------------------------------
        |                  Standard            Prob.       Mean
    LNCT| Coefficient        Error       z    z>|Z|        of X
--------+--------------------------------------------------------
Constant|    14.1048***     2.94550     4.79  .0000
     LNQ|   -1.81538***      .24881    -7.30  .0000    16.4616
   LNQSQ|     .07876***      .00775    10.16  .0000    272.853
    LNPL|     .14406         .17091      .84  .3993    13.2194
    LNPE|    -.30467*        .18467    -1.65  .0990   -1.85956
    LNPK|     .30383***      .02510    12.10  .0000    10.1795
NARROW_T|     .00122         .02850      .04  .9659     .67603
    RACK|     .42526***      .02564    16.59  .0000     .23471
  TUNNEL|    2.05124***      .57610     3.56  .0004     .18843
LOGQTUNL|    -.13334***      .03355    -3.97  .0001    3.40727
--------+--------------------------------------------------------
```

REGRESSION 4

```
+----------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.     |
| Sample of    605 observations contained     50 clusters defined by   |
| variable ID       which identifies groups by a cluster ID.          |
+----------------------------------------------------------------------+
-----------------------------------------------------------------
Ordinary     least squares regression ...........
LHS=LNCT     Mean                  =      11.30622
Model size   Parameters            =            10
--------+--------------------------------------------------------
        |                  Standard            Prob.       Mean
    LNCT| Coefficient        Error       z    z>|Z|        of X
--------+--------------------------------------------------------
Constant|    14.1048*       7.65472     1.84  .0654
     LNQ|   -1.81538**       .79951    -2.27  .0232    16.4616
   LNQSQ|     .07876***      .02490     3.16  .0016    272.853
    LNPL|     .14406         .33121      .43  .6636    13.2194
    LNPE|    -.30467         .45559     -.67  .5037   -1.85956
    LNPK|     .30383***      .07350     4.13  .0000    10.1795
NARROW_T|     .00122         .09298      .01  .9895     .67603
    RACK|     .42526***      .08414     5.05  .0000     .23471
  TUNNEL|    2.05124        1.78394     1.15  .2502     .18843
 LNQTUNL|    -.13334         .10411    -1.28  .2003    3.40727
--------+--------------------------------------------------------
```

5. Munnell, A., "Why has Productivity Declined? Productivity and Public Investment," New England Economic Review, 1990, pp. 3-22, examined the productivity of public capital in a panel of data using the lower 48 states and 17 years. These data are examined at length in Chapter 10 of the 7^{th} edition of your text. In this exercise, we will use a very simple version of her model,

$$\log GSP_{it} = \beta_1 \ \beta_2 \log PublicK_{it} + \beta_3 \log PrivateK_{it} + \beta_4 \log Labor_{it} + \varepsilon_{it}$$

where GSP is gross state product.Ordinary least squares regression results appear below. KP is public capital; PC is private capital.

a. Test the hypothesis that the marginal products of (coefficients on) private and public capital are the same.
b. Test the hypothesis of constant returns to scale (that is, the hypothesis that the three coefficients sum to 1.0)
c. Test the two hypotheses simultaneously.

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LOGGSP   Mean                 =   10.50885     |
|              Standard deviation   =   1.021132     |
| WTS=none     Number of observs.   =        816     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =        812     |
| Residuals    Sum of squares       =   6.469532     |
|              Standard error of e  =   .8926031E-01 |
| Fit          R-squared            =   .9923871     |
|              Adjusted R-squared   =   .9923589     |
| Model test   F[  3,   812] (prob) =****** (.0000) |
| Diagnostic   Log likelihood       =   815.7689     |
|              Restricted(b=0)      =  -1174.417     |
|              Chi-sq [  3]  (prob) =3980.37(.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.64886431       .05833603    28.265   .0000
 LOGKP   |     .15078348       .01735707     8.687   .0000   9.67920583
 LOGPC   |     .30553817       .01037855    29.439   .0000   10.5594618
 LOGEMP  |     .59815198       .01390006    43.032   .0000   6.97849785

Asymptotic Covariance Matrix
              1         2          3             4
       1|     .00340
       2|    -.00059   .00030
       3|    -.00020  -.00009078   .00011
       4|     .00064  -.00020   -.000008636   .00019
```

6. The three sets of results below show the least squares estimates for two of the states, then the results for these two states combined. (Presumably, these two are representative of the 48 in the data set.)

a. Theory 1 states that the coefficient vectors are the same for the two states. 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 thecomputations in full detail so that I know exactly how you obtained your result.
c. Use a Wald test to test the hypothesis that the coefficients are the same. Again, document your computations.

```
| LHS=LOGGSP   Mean                 =   10.53753     |
|              Standard deviation   =   .1584103     |
| WTS=none     Number of observs.   =         17     |
| Model size   Parameters           =          4     |
| Residuals    Sum of squares       =   .9063141E-02 |
|              Standard error of e  =   .2640388E-01 |
| Fit          R-squared            =   .9774269     |
|              Adjusted R-squared   =   .9722177     |
| Model test   F[  3,    13] (prob) = 187.64 (.0000) |
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    6.20005924      3.26548334     1.899   .0800
 LOGKP   |   -1.11530796       .69214743    -1.611   .1311   9.79136043
 LOGPC   |     .49706712       .25806357     1.926   .0762   10.8133466
 LOGEMP  |    1.38609118       .30315422     4.572   .0005   7.13004557
             1            2             3             4
        +--------------------------------------------------------
      1|   10.66338     -2.24221       .68603       .54315
      2|   -2.24221       .47907      -.14467      -.12400
      3|     .68603     -.14467       .06660       .00145
      4|     .54315     -.12400       .00145       .09190

| LHS=LOGGSP   Mean                 =   11.54882     |
| Residuals    Sum of squares       =   .2267268E-02 |
|              Standard error of e  =   .1320626E-01 |
| Fit          R-squared            =   .9972928     |
|              Adjusted R-squared   =   .9966681     |
| Model test   F[  3,    13] (prob) =1596.37 (.0000) |
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.81099728      1.00540552     1.801   .0949
 LOGKP   |     .48580792       .28675566     1.694   .1140   10.5943058
 LOGPC   |    -.24199469       .13905321    -1.740   .1054   11.3937222
 LOGEMP  |     .91031345       .09078688    10.027   .0000   8.07221529
             1            2             3             4
        +--------------------------------------------------------
      1|    1.01084      -.28472       .12393       .07353
      2|    -.28472       .08223      -.03730      -.02000
      3|     .12393      -.03730       .01934       .00631
      4|     .07353      -.02000       .00631       .00824

| LHS=LOGGSP   Mean                 =   11.04318     |
| WTS=none     Number of observs.   =         34     |
| Residuals    Sum of squares       =   .3297875E-01 |
|              Standard error of e  =   .3315557E-01 |
| Fit          R-squared            =   .9966796     |
|              Adjusted R-squared   =   .9963475     |
| Model test   F[  3,    30] (prob) =3001.65 (.0000) |
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.63395245       .82374079     3.198   .0033
 LOGKP   |    -.02655970       .19357706     -.137   .8918   10.1928331
 LOGPC   |     .05983913       .04814774     1.243   .2236   11.1035344
 LOGEMP  |    1.05451622       .17031130     6.192   .0000   7.60113043
             1            2             3             4
        +--------------------------------------------------------
      1|     .67855      -.14905      -.01781       .13662
      2|    -.14905       .03747       .00111      -.03227
      3|    -.01781       .00111       .00232      -.00254
      4|     .13662      -.03227      -.00254       .02901
```

7. We now return to the panel data set examined in question 2. The results below show OLS, fixed effects and random effects estimates.

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 model.

d. In the context of the fixed effects model, test the hypothesis that there are no effects – i.e., that all individuals have the same constant term. (The statistics you need to carry out the test aregiven in the results.)

e. The variables narrow_t, rack and tunnel are time invariant. Explain why it is necessary to omit these variables from the equation to compute the fixed effects regression..

f. Since there are time invariant variables in the model, the Hausman statistic cannot be computed. What I did instead was compute the group means of the time varying variables (logq, lnpe, lnpl, lnpk) and add them to the model. I then used this regression to compute the Wu statistic to test the hypothesis of fixed versus random effects. The value of the statistic is 106.6.

(1) How is the statistic computed?

(2) What should I conclude on the basis of the test?

```
-----------------------------------------------------------------
OLS Without Group Dummy Variables................
Ordinary     least squares regression ...........
LHS=LNCT     Mean                 =      11.30622
             Standard deviation   =       1.10169
             Number of observs.   =           605
Model size   Parameters           =             8
             Degrees of freedom   =           597
Residuals    Sum of squares       =      48.49299
             Standard error of e  =        .28500
Fit          R-squared            =        .93385
             Adjusted R-squared   =        .93308
Model test   F[  7,   597] (prob) =  1204.0(.0000)
Diagnostic   Log likelihood       =     -95.00555
             Restricted(b=0)      =    -916.54939
             Chi-sq [  7]  (prob) =  1643.1(.0000)
Info criter. LogAmemiya Prd. Crt. =      -2.49736
             Akaike Info. Criter. =      -2.49736
             Bayes Info. Criter.  =      -2.43911
--------+--------------------------------------------------------
        |                  Standard           Prob.       Mean
    LNCT| Coefficient       Error       z    z>|Z|       of X
--------+--------------------------------------------------------
    LOGQ|     .77200***      .01125    68.64  .0000    16.4616
    LNPL|     .15615         .17391      .90  .3692    13.2194
    LNPE|    -.41790**       .18157    -2.30  .0214   -1.85956
    LNPK|     .34916***      .02891    12.08  .0000    10.1795
NARROW_T|    -.13716***      .02699    -5.08  .0000     .67603
    RACK|     .48192***      .03098    15.56  .0000     .23471
  TUNNEL|    -.14985***      .03790    -3.95  .0001     .18843
Constant|   -7.78982***     1.98172    -3.93  .0001
--------+--------------------------------------------------------
```

```
-----------------------------------------------------------------
Least Squares with Group Dummy Variables..........
Ordinary     least squares regression ............
LHS=LNCT     Mean                 =      11.30622
             Standard deviation   =       1.10169
             Number of observs.   =           605
Model size   Parameters           =            57
             Degrees of freedom   =           548
Residuals    Sum of squares       =       3.35773
             Standard error of e  =        .07828
Fit          R-squared            =        .99542
             Adjusted R-squared   =        .99495
Model test   F[ 56,   548] (prob) =  2126.7(.0000)
Diagnostic   Log likelihood       =     712.71615
             Restricted(b=0)      =    -916.54939
             Chi-sq [ 56]  (prob) =  3258.5(.0000)
Info criter. LogAmemiya Prd. Crt. =      -5.00497
             Akaike Info. Criter. =      -5.00553
             Bayes Info. Criter.  =      -4.59050
Estd. Autocorrelation of e(i,t)   =       .682787
Panel:Groups Empty      0,     Valid data       50
             Smallest   1,     Largest          13
             Average group size in panel     12.10
These  3 variables have no within group variation.
NARROW_T RACK     TUNNEL
F.E. estimates are based on a generalized inverse.
--------+--------------------------------------------------------
        |                  Standard           Prob.       Mean
    LNCT| Coefficient        Error       z    z>|Z|       of X
--------+--------------------------------------------------------
    LOGQ|     .31912***      .02940    10.85  .0000    16.4616
    LNPL|     .45701***      .06676     6.85  .0000    13.2194
    LNPE|   -.30126***       .08677    -3.47  .0005   -1.85956
    LNPK|     .30843***      .01886    16.35  .0000    10.1795
NARROW_T|       .000    .....(Fixed Parameter).....     .67603
    RACK|       .000    .....(Fixed Parameter).....     .23471
  TUNNEL|       .000    .....(Fixed Parameter).....     .18843
--------+--------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
-----------------------------------------------------------------

+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model             Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only     -916.54938         733.08869     .00000 |
|(2)  Group effects only      306.82066          12.84646     .98248 |
|(3)  X - variables only      -95.00554          48.49299     .93385 |
|(4)  X and group effects     712.71616           3.35773     .99542 |
+--------------------------------------------------------------------+
|                        Hypothesis Tests                            |
|         Likelihood Ratio Test           F Tests                    |
|         Chi-squared   d.f.   Prob          F    num    denom  P value |
|(2) vs (1)   2446.74     49  .0000    635.03    49      555   .00000 |
|(3) vs (1)   1643.09      7  .0000   1204.01     7      597   .00000 |
|(4) vs (1)   3258.53     56  .0000   2126.72    56      548   .00000 |
|(4) vs (2)    811.79      7  .0000    221.23     7      548   .00000 |
|(4) vs (3)   1615.44     49  .0000    150.33    49      548   .00000 |
+--------------------------------------------------------------------+
```

```
------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .006127
            Var[u]              =       .075101
            Corr[v(i,t),v(i,s)] =       .924567
Lagrange Multiplier Test vs. Model (3) =2638.43
( 1 degrees of freedom, prob. value =  .000000)
(High values of LM favor FEM/REM over CR model)
Fixed vs. Random Effects (Hausman)     =    .00
( 7 degrees of freedom, prob. value = 1.000000)
(High (low) values of H favor F.E.(R.E.) model)
            Sum of Squares           97.926721
            R-squared                  .866509
--------+---------------------------------------------------------
        |                Standard           Prob.       Mean
    LNCT| Coefficient       Error       z    z>|Z|       of X
--------+---------------------------------------------------------
    LOGQ|     .50994***      .02267    22.50  .0000    16.4616
    LNPL|     .33111***      .06530     5.07  .0000    13.2194
    LNPE|    -.38440***      .08547    -4.50  .0000   -1.85956
    LNPK|     .29346***      .01849    15.87  .0000    10.1795
NARROW_T|    -.08171         .08643     -.95  .3444     .67603
    RACK|     .30482***      .09890     3.08  .0021     .23471
  TUNNEL|     .36262***      .11075     3.27  .0011     .18843
Constant|   -5.24152***      .72709    -7.21  .0000
--------+---------------------------------------------------------

------------------------------------------------------------------
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .006172
            Var[u]              =       .072825
            Corr[v(i,t),v(i,s)] =       .921867
Lagrange Multiplier Test vs. Model (3) =2838.51
( 1 degrees of freedom, prob. value =  .000000)
(High values of LM favor FEM/REM over CR model)
Baltagi-Li form of LM Statistic =       1739.15
            Sum of Squares           47.156449
            R-squared                  .935674
--------+---------------------------------------------------------
        |                Standard           Prob.       Mean
    LNCT| Coefficient       Error       z    z>|Z|       of X
--------+---------------------------------------------------------
    LOGQ|     .31912***      .02951    10.81  .0000    16.4616
    LNPL|     .45701***      .06700     6.82  .0000    13.2194
    LNPE|    -.30126***      .08709    -3.46  .0005   -1.85956
    LNPK|     .30843***      .01893    16.29  .0000    10.1795
NARROW_T|    -.16251*        .08937    -1.82  .0690     .67603
    RACK|     .50776***      .10847     4.68  .0000     .23471
  TUNNEL|    -.20105         .12896    -1.56  .1190     .18843
   LOGQB|     .45832***      .04880     9.39  .0000    16.4616
   LNPLB|     .16415         .95577      .17  .8636    13.2194
   LNPEB|     .32844         .87197      .38  .7064   -1.85956
   LNPKB|     .05097         .11268      .45  .6511    10.1795
Constant|   -13.2818       11.06214    -1.20  .2299
--------+---------------------------------------------------------
```

8. You may use any data set you wish for this exercise. You are going to fit a binary choice model, so choose one that contains an interesting binary variable that you will explain with your binary choice model. The data set may be one of the ones that we used during the semester, or one of the data sets on the data page for your text, or any other data set that you find interesting.

You will need a statistical package to do this part of the exam. You may use NLOGIT, Stata, R, MatLab, Gauss, or any other package you are familiar with.

Your assignment is to estimate a binary choice model using your data. Your report should document the data set, lay out the model you are estimating, and present the results in a professional looking format in a table of results. The entire writeup need not be more than a page or two, but it should document your empirical work as if you were submitting for review by a referee or a colleague in your department.

Your may fit a logit or probit model. (We are not interested in the 'linear probability model.') In addition to the main assignment above, also do the following:

a. In a 'technical appendix,' write down the log likelihood for your model. Derive the likelihood equations (first order conditions) for estimation of the model parameters. Derive an asymptotic covariance matrix for your estimator.

b. Using an alternative specification for your equation, carry out a hypothesis test using a likelihood ratio test.

c. Report and interpret the partial effects for your model. Indicate whether you have used the partial effects at the means of the data or the average partial effects.

9. The probit model below examines the probability that an individual reports Health Satisfaction greater than 6 in the 0 – 10 scale for HSAT in the GSOEP. Age10 = AGE/10. RICH is a dummy variable that equals one if the individual's income is in the top 20% of the incomes in the sample. The results agree with my expectations (with one exception; I would not expect HEALTHY to increase with age, as it . However, I am concerned that RICH may be endogenous in this model – the unobservables that influence health are likely to influence the ability to obtain a high income as well. How can I consistently estimate the model in this case of an endogenous variable in a binary choice model? Once I do, how do I estimate the interesting impact of income on health (i.e., the"treatment effect")?

```
Binomial Probit Model
Dependent variable                HEALTHY
Log likelihood function    -17476.18255
Restricted log likelihood  -18279.94994
Chi squared [   5](P= .000)   1607.53478
Significance level                .00000
McFadden Pseudo R-squared       .0439699
Estimation based on N =   27326, K =    6
Inf.Cr.AIC  =   34964.4 AIC/N =    1.280
--------+------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
 HEALTHY|  Coefficient     Error       z     |z|>Z*        Interval
--------+------------------------------------------------------------
        |Index function for probability.................................
Constant|    1.73753***     .12904    13.46  .0000     1.48461   1.99044
    RICH|     .21776***     .02357     9.24  .0000      .17156    .26395
  FEMALE|   -.14399***      .01572    -9.16  .0000    -.17479   -.11319
 MARRIED|     .04220**      .01923     2.20  .0282      .00452    .07989
   AGE10|   -.40707***      .06182    -6.58  .0000    -.52824   -.28591
AGESQ100|     .01652**      .00690     2.40  .0166      .00300    .03004
--------+------------------------------------------------------------
```