Department of Economics

Econometric Analysis of Panel Data

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

Final Examination: Spring 2013

This is a 'take home' examination. Today is Tuesday, May 7, 2013. Your answers are due on Friday, May 17, 2013. You may use any resources you wish – textbooks, computer, the web, etc. – but please work alone and submit only your own answers to the questions.

The six parts of the exam are weighted as follows:

Part I.	**Literature**	**20**
Part II.	**The Hausman and Taylor Estimator**	**20**
Part III.	**Panel Data Regressions**	**20**
Part IV.	**Binary Choice Models**	**20**
Part V.	**A Loglinear Model**	**20**

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

Part I. Literature

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

Part II. The Hausman and Taylor Estimator

Write out a full statement of the procedure that Hausman and Taylor devised for estimation of the parameters in a panel data model in which some independent variables are correlated with the time invariant part of the disturbance in a random effects model. Now, show how the Arellano/Bond/Bover estimator uses the Hausman and Taylor result.

Part III. Panel Data Regressions

The course website contains Munnell's data set on statewide production, in ASCII text form,

http://people.stern.nyu.edu/wgreene/Econometrics/productivity.csv

and as an nlogit project,

http://people.stern.nyu.edu/wgreene/Econometrics/productivity.lpj

The data in the file are a panel on the following variables for the lower 48 states, 17 years (816 observations),

STATE	= state name
YR	= year, 1970,...,1986
P_CAP	= public capital
HWY	= highway capital
WATER	= water utility capital
UTIL	= utility capital
PC	= private capital
GSP	= gross state product
EMP	= employment
UNEMP	= unemployment rate

The basic model of interest is

$$Y_{it} = \alpha + \beta_1 X1_{it} + \beta_2 X2_{it} + \beta_3 X3_{it} + \beta_4 X4_{it} + \beta_5 X5_{it} + c_i + \varepsilon_{it}$$

Where Y is logGSP, X1 is logPC, X2 is logHWY, X3 is logWATER, X4 is logUTIL and X5 is logEMP.

This is a Cobb-Douglas production function.
a. Fit the "pooled" model and report your results
b. Fit a random effects model and a fixed effects model. Use your estimation results to decide which is the preferable model. If you find that neither panel data model is preferred to the pooled model, show how you reached that conclusion. As part of the analysis, test the hypothesis that there are no "state effects."
c. Assuming that there are "latent individual (state) effects," the asymptotic covariance matrix that is computed for the pooled estimator, $s^2(\mathbf{X}'\mathbf{X})^{-1}$, is inappropriate. What estimator can be computed for the covariance matrix of the pooled estimator that will give appropriate standard errors?
d. The hypothesis of constant returns would be that

$$H_0: \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1$$

Test this hypothesis in the context of the model in a. and in the context of your preferred model in part b. Do you reach the same conclusion in both cases?
e. A theory of production posed by Professor Axehandle claims that much of state production is committed years in advance, so that a model with lagged production,

$$Y_{it} = \alpha + \beta_1 X1_{it} + \beta_2 X2_{it} + \beta_3 X3_{it} + \beta_4 X4_{it} + \beta_5 X5_{it} + \gamma Y_{i,t-1} + c_i + \varepsilon_{it}$$

is more appropriate than the earlier model. What method should Professor A use to obtain consistent estimates of the parameters of her model? Explain your suggestion. (You need not actually estimate Professor A's model. Just advise her on how to do the estimation.)

Part IV. Binary Choice Models

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

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

or an nlogit project file,

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

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

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

I am interested in a binary choice model for the innovation variable, IP. You will fit your model using at least three of the independent variables in the data set. With respect to the model you specify,

A. THEORY

(a) If you fit a pooled **logit** model, there is the possibility that you might be ignoring unobserved heterogeneity (effects). Wooldridge argues that when one fits a probit model while ignoring unobserved heterogeneity, the raw coefficient estimator (MLE) is inconsistent, but the quantity of interest, the "Average Partial Effects" might well be estimated appropriately. Explain in detail what he has in mind here.

(b) Suppose we were to estimate a "fixed effects" probit model by "brute force," just by including the 1,270 dummy variables needed to create the empirical model. What would the properties of the resulting estimator likely be? What is "the incidental parameters problem?"

(c) How would I proceed to use Chamberlain's estimator to obtain a consistent slope estimator for the fixed effects logit model.

(d) Describe in detail how to fit a random effects logit model using quadrature and using simulation for the part of the computations where they would be necessary, under the assumption that the effects are uncorrelated with the other included exogenous variables.

(e) Using the random effects logit model that you described in part (d), describe how you would test the hypothesis that the same logit model applies to the four different sectors in the data set (CONSGOOD,FOOD,RAWMTL,INVGOOD).

B. PRACTICE

(a) Fit a pooled probit model using your specification. Provide all relevant estimation results. (Please condense and organize the results in a readable form.)

(b) Fit a random effects probit model.

(c) Use the Mundlak (correlated random effects) approach to approximate a fixed effects model. Recall this means adding the group means of the time varying variables to the model, then using a random effects model.

(d) Note the difference between the estimates in (b) and (c). Which do you think is appropriate? Explain.

Tip for nlogit users: You can use

SETPANEL ; Group=firm ; Pds=ti $

To define the panel, then

CREATE ; new variable = GroupMean(variable,pds=ti)$

To obtain the group means you need for a variable.

Part V. A Loglinear Model

The random variable y|**x** has a gamma distribution:

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

Estimation and analysis is based on a sample of N observations on y_i,$\mathbf{x}_i$.

(a) I propose to estimate the parameters P,α,**β** by maximum likelihood. Derive the log likelihood function, likelihood equations and Hessian. Show precisely how to use Newton's method to estimate the parameters. How will you obtain asymptotic standard errors for your estimator? Note, the derivatives of the gamma function, Γ(P) are $d\log\Gamma(P)/dP = \psi(P)$ and $d^2\log\Gamma(P)/dP^2 = \Psi'(P)$. Just use these symbols in your results.

(b) The conditional mean function is

$$E[y_i|\mathbf{x}_i] = P/\lambda_i = P\times\exp(-\alpha - \boldsymbol{\beta}'\mathbf{x}_i) = \exp[\ (\log P-\alpha) - \boldsymbol{\beta}'\mathbf{x}_i] = \exp(\theta - \boldsymbol{\beta}'\mathbf{x}_i)$$

Show how this can be used to construct a nonlinear least squares estimator of **β**. Note that it is not possible to obtain separate estimates of P and α by this method – the constant term will be α – logP for the reason shown in the equation above.

(c) Another conditional mean is

$$E[\log y_i|x_i] = \psi(P) - \log\lambda_i = -[\alpha- \psi(P)] - \beta' x_i = -(\delta + \beta' x_i)$$

How can you use this result to obtain an estimator of **β**.

(d) Which of the three estimators suggested above will be the most efficient? Explain.

(e) Maximum likelihood estimates of the model using our German health care data, in which the dependent variable is household income, HHNINC, appear on the next page.

(1) The Erlang form of the gamma model has an integer value of P. The second set of results below are obtained with P = 3. Test the hypothesis of the Erlang form vs. the more general form using both Wald and Likelihood ratio tests.

(2) Test the hypothesis that **β**=0 using a likelihood ratio test.

(3) Show how to use a Wald test to test the hypothesis that **β** = 0.

(f) Derive the functional form of the partial effects in the model, $\partial E[y|x]/\partial x$. Then, using the general model with unrestricted P, compute these partial effects at the means of the variables. (The means are given in the final set of results after the least squares results.)

(g) Nonlinear least squares, OLS, and OLS results for the log of income are shown . What is the relationship between thes results, the parameters computed in (g) and the partial effects? Note, compared to the MLEs, the OLS results seem to have thw wrong signs. Explain.

```
--------------------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                INCOME
Log likelihood function      14279.41554
Restricted log likelihood     1195.24508
Chi squared [   5 d.f.]      26168.34092
Estimation based on N =  27326, K =   6
Inf.Cr.AIC  = -28546.8 AIC/N =    -1.045
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     3.53885***      .02233   158.48  .0000      3.49509   3.58262
     AGE|      .00040         .00026     1.55  .1213     -.00011    .00091
    EDUC|     -.05435***      .00119   -45.81  .0000     -.05667   -.05202
 MARRIED|     -.23530***      .00646   -36.44  .0000     -.24796   -.22264
    HSAT|     -.01074***      .00122    -8.82  .0000     -.01313   -.00836
        |Scale parameter for gamma model
 P_scale|     5.11950***      .04245   120.61  .0000      5.03630   5.20270
--------+-----------------------------------------------------------------------
```

Estimated covariance matrix for coefficient vector

[6, 6] Cell: 0.000498604

	Constant	AGE	EDUC	MARRIED	HSAT	P_scale
Constant	0.000498604	-3.41396e-006	-1.65021e-005	-1.62231e-005	-1.10514e-005	0.000351957
AGE	-3.41396e-006	6.71176e-008	3.12955e-008	-4.04454e-007	6.56027e-008	-4.67543e-014
EDUC	-1.65021e-005	3.12955e-008	1.40767e-006	3.09253e-007	-1.51836e-007	-6.21876e-013
MARRIED	-1.62231e-005	-4.04454e-007	3.09253e-007	4.17027e-005	-1.9308e-007	2.24721e-012
HSAT	-1.10514e-005	6.56027e-008	-1.51836e-007	-1.9308e-007	1.48273e-006	1.98232e-013
P_scale	0.000351957	-4.67543e-014	-6.21876e-013	2.24721e-012	1.98232e-013	0.00180184

```
--------------------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                INCOME
Log likelihood function      12505.22870
Restricted log likelihood     1195.24508
Chi squared [   5 d.f.]      22619.96724
Estimation based on N =  27326, K =   5
Inf.Cr.AIC  = -25000.5 AIC/N =     -.915
--------+-----------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+-----------------------------------------------------------------------
        |Parameters in conditional mean function
Constant|     3.00441***      .02708   110.93  .0000      2.95132   3.05749
     AGE|      .00040         .00034     1.19  .2356     -.00026    .00106
    EDUC|     -.05435***      .00155   -35.07  .0000     -.05739   -.05131
 MARRIED|     -.23530***      .00844   -27.89  .0000     -.25183   -.21877
    HSAT|     -.01074***      .00159    -6.75  .0000     -.01386   -.00763
        |Scale parameter for gamma model
 P_scale|         3.0     .....(Fixed Parameter).....
--------+-----------------------------------------------------------------------
```

Estimated covariance matrix for coefficient vector

[6, 6] Cell: 0.000733548

	Constant	AGE	EDUC	MARRIED	HSAT	P_scale
Constant	0.000733548	-5.82592e-006	-2.81609e-005	-2.76847e-005	-1.88592e-005	0
AGE	-5.82592e-006	1.14536e-007	5.34058e-008	-6.90201e-007	1.11951e-007	0
EDUC	-2.81609e-005	5.34058e-008	2.40218e-006	5.2774e-007	-2.59109e-007	0
MARRIED	-2.76847e-005	-6.90201e-007	5.2774e-007	7.11657e-005	-3.29491e-007	0
HSAT	-1.88592e-005	1.11951e-007	-2.59109e-007	-3.29491e-007	2.53028e-006	0
P_scale	0	0	0	0	0	0

```
-----------------------------------------------------------------------------
Nonlinear     least squares regression ............
LHS=INCOME    Mean                 =          .35214
              Standard deviation   =          .17686
              Number of observs.   =           27326
Model size    Parameters           =               5
              Degrees of freedom   =           27321
Residuals     Sum of squares       =         761.973
              Standard error of e  =          .16699
Fit           R-squared            =          .10847
              Adjusted R-squared   =          .10850
Model test    F[  4, 27321] (prob) =    831.0(.0000)
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      B1|   -1.97224***      .02139   -92.21  .0000    -2.01416  -1.93032
      B2|     .00132***      .00027     4.88  .0000      .00079    .00185
      B3|     .05381***      .00103    52.16  .0000      .05179    .05584
      B4|     .23520***      .00764    30.77  .0000      .22022    .25018
      B5|     .01043***      .00132     7.89  .0000      .00784    .01301
--------+--------------------------------------------------------------------
Ordinary      least squares regression ............
----------    No. of observations  =           27326  DegFreedom   Mean square
Regression    Sum of Squares       =         91.7939          4     22.94847
Residual      Sum of Squares       =         762.888      27321       .02792
Total         Sum of Squares       =         854.682      27325       .03128
----------    Standard error of e  =          .16710  Root MSE          .16709
Fit           R-squared            =          .10740  R-bar squared     .10727
Model test    F[  4, 27321]        =       821.84443  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  INCOME|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .02071***      .00788     2.63  .0086      .00527    .03616
     AGE|     .00025***   .9450D-04     2.65  .0081      .00007    .00044
    EDUC|     .02094***      .00044    47.15  .0000      .02007    .02181
 MARRIED|     .07801***      .00241    32.34  .0000      .07328    .08274
    HSAT|     .00358***      .00046     7.84  .0000      .00268    .00447
--------+--------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=logINCOM Mean                  =        -1.15744
              Standard deviation   =          .49145
----------    No. of observations  =           27326  DegFreedom   Mean square
Regression    Sum of Squares       =         947.342          4    236.83541
Residual      Sum of Squares       =         5652.32      27321       .20689
Total         Sum of Squares       =         6599.66      27325       .24152
----------    Standard error of e  =          .45485  Root MSE          .45481
Fit           R-squared            =          .14354  R-bar squared     .14342
Model test    F[  4, 27321]        =      1144.76509  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
logINCOM|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|   -2.08985***      .02145   -97.43  .0000    -2.13190  -2.04781
     AGE|    -.00069***      .00026    -2.70  .0070     -.00120   -.00019
    EDUC|     .05670***      .00121    46.90  .0000      .05433    .05907
 MARRIED|     .32016***      .00657    48.77  .0000      .30729    .33303
    HSAT|     .01148***      .00124     9.24  .0000      .00904    .01391
--------+--------------------------------------------------------------------
Descriptive Statistics for   6 variables
--------+---------------------------------------------------------------------
Variable|       Mean        Std.Dev.      Minimum      Maximum     Cases Missing
--------+---------------------------------------------------------------------
  INCOME|      .352135      .176857      .001500     3.067100    27326       0
logINCOM|    -1.157442      .491452    -6.502290     1.120732    27326       0
     AGE|     43.52569     11.33025         25.0         64.0    27326       0
    EDUC|     11.32063     2.324885          7.0         18.0    27326       0
 MARRIED|      .758618      .427929          0.0          1.0    27326       0
    HSAT|     6.785662     2.293725          0.0         10.0    27326       0
--------+---------------------------------------------------------------------
```