NYU STERN
NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS
Department of Economics

Econometric Analysis of Panel Data
Assignment 2

Part I. Interpreting Regression Results

The results below show OLS, fixed effects and random effects estimates for a reduced version of the model analyzed in class using the Cornwell and Rupert data.

1. Test the hypothesis of 'no effects' vs. 'some effects' using the results given below.

2. Explain in precise detail the difference between the fixed and random effects models.

3. Carry out the Hausman test for fixed effects against the null hypothesis of random effects and report your conclusion. Carefully explain what you are doing in this test.

4. 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 are given in the results.)

5. Using the fixed effects estimator, test the hypothesis that all of the coefficients in the model save for the constant terms are zero. Show all computations, and the appropriate degrees of freedom for *F*.

6. Discuss the impact of adding the individual dummy variables to the model – in terms of the substantive change (or lack of) in the estimated results.

Part II. Fixed Effects Normalization

Some researchers (such as your professor) prefer to fit the conventional fixed effects model (estimator) by having exactly one dummy variable in the model for each individual. In some other cases, the researchers prefer to have a single overall constant and a set of N-1 individual dummy variables, i.e., dropping one of the individual constants to avoid the collinearity problem. (This is the default setting in Stata, for example.) A third way to proceed is to include an overall constant and the full set of dummy variables, but constrain the dummy variable coefficients to sum to zero. How does this manipulation of the dummy variable coefficients affect the least squares estimates of the other coefficients in the model and the fit of the equation, i.e., R^2?

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LWAGE    Mean                 =        6.51800
             Standard deviation   =         .42809
----------   No. of observations  =             51  DegFreedom   Mean square
Regression   Sum of Squares       =        1.57896           4        .39474
Residual     Sum of Squares       =        7.58394          46        .16487
Total        Sum of Squares       =        9.16289          50        .18326
----------   Standard error of e  =         .40604  Root MSE          .38562
Fit          R-squared            =         .17232  R-bar squared     .10035
Model test   F[  4,    46]        =        2.39427  Prob F > F*       .06406
Diagnostic   Log likelihood       =      -23.76813  Akaike I.C.   = -1.70971
             Restricted (b=0)     =      -28.59094  Bayes  I.C.   = -1.52032
             Chi squared [  4]    =        9.64561  Prob C2 > C2* =   .04684
B-P test     Chi squared [  1]    =       39.30737  Prob C2 > C2* =   .00000
[High values of  LM favor FEM/REM over base model]
Baltagi-Li form of LM Statistic   =       26.75797  [= BP if balanced panel]
Moulton/Randolph form:SLM N[0,1]  =       12.91801
--------------------------------------------------
Panel Data Analysis of LWAGE             [ONE way]
               Unconditional ANOVA (No regressors)
Source         Variation  Deg. Free.   Mean Square
Between          5.96999          7.        .85286
Residual         3.19290         43.        .07425
Total            9.16289         50.        .18326
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
     OCC|      .25940         .28182      .92  .3621     -.30788    .82668
   UNION|     -.20027         .21010     -.95  .3455     -.62318    .22265
      MS|      .22821         .16103     1.42  .1632     -.09592    .55234
     EXP|      .00794         .01045      .76  .4513     -.01309    .02897
Constant|     6.03804***      .23289    25.93  .0000     5.56926   6.50682
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------


-----------------------------------------------------------------------------
LSDV         least squares with fixed effects ....
LHS=LWAGE    Mean                 =        6.51800
             Standard deviation   =         .42809
----------   No. of observations  =             51  DegFreedom   Mean square
Regression   Sum of Squares       =        8.58750          11        .78068
Residual     Sum of Squares       =        .575396          39        .01475
Total        Sum of Squares       =        9.16289          50        .18326
----------   Standard error of e  =         .12146  Root MSE          .10622
Fit          R-squared            =         .93720  R-bar squared     .91949
Model test   F[ 11,    39]        =       52.91412  Prob F > F*       .00000
Diagnostic   Log likelihood       =       41.98946  Akaike I.C.   = -4.01393
             Restricted (b=0)     =      -28.59094  Bayes  I.C.   = -3.55939
             Chi squared [ 11]    =      141.16079  Prob C2 > C2* =   .00000
Estd. Autocorrelation of e(i,t)   =       -.026845
--------------------------------------------------
Panel:Groups Empty      0,     Valid data        8
             Smallest   2,     Largest           7
             Average group size in panel      6.38
Variances    Effects a(i)         Residuals e(i,t)
                 1.186657                  .014754
Rho squared: Residual variation due to ai  .987720
Within groups variation in LWAGE            3.1929
R squared based on within group variation  .819789
Between group variation in LWAGE            5.9700
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
     OCC|     -.08928         .11812     -.76  .4536     -.32691    .14836
   UNION|     -.16358         .09779    -1.67  .1010     -.36030    .03315
      MS|     -.07189         .10711     -.67  .5054     -.28737    .14359
     EXP|      .10308***      .00967    10.66  .0000      .08363    .12253
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Test Statistics for the Fixed Effects Regression Model

	Model	Log-Likelihood	Sum of Squares	R-squared
(1)	Constant term only	-28.59094	9.16289	.00000
(2)	Group effects only	-1.70802	3.19290	.65154
(3)	X - variables only	-23.76813	7.58394	.17232
(4)	X and group effects	41.98946	.57540	.93720

Hypothesis Tests

	Likelihood Ratio Test			F Tests			
	Chi-squared	d.f.	Prob	F	num	denom	P value
(2) vs (1)	53.77	7	.0000	11.49	7	43	.00000
(3) vs (1)	9.65	4	.0468	2.39	4	46	.06406
(4) vs (1)	141.16	11	.0000	52.91	11	39	.00000
(4) vs (2)	87.39	4	.0000	44.35	4	39	.00000
(4) vs (3)	131.52	7	.0000	67.86	7	39	.00000

```
Random Effects Model: v(i,t)    = e(i,t) + u(i)
Estimates:  Var[e]              =       .014754
            Var[u]              =       .150114
            Corr[v(i,t),v(i,s)] =       .910512
            Sum of Squares            30.108108
            R-squared                 -2.283534
Variances computed using OLS and LSDV with d.f.
-----------------------------------------------
Fixed vs. Random Effects (Hausman)  =     36.02
[ 4 degrees of freedom, prob. value =  .000000]
[High (low) values of H favor F.E.(R.E.) model]
```

LWAGE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
OCC	-.24613**	.11310	-2.18	.0295	-.46781	-.02446
UNION	-.11957	.09561	-1.25	.2111	-.30696	.06782
MS	-.10910	.10359	-1.05	.2922	-.31213	.09392
EXP	.07438***	.00804	9.25	.0000	.05861	.09014
Constant	5.31395***	.26855	19.79	.0000	4.78760	5.84031

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

Part III. Estimating Variance Components

Greene (2018), Wooldridge (2010, page 296), etc. suggest that in order to obtain the asymptotically efficient FGLS estimator of the coefficients in the random effects model, one only needs a consistent pair of estimators for σ_ε^2 and σ_u^2 – any consistent estimators will do. That is good, because there are quite a few available. One is suggested in Greene (2018 on pages 408-409) based on the degrees of freedom corrected OLS and FE estimators. A different one uses the pooled OLS estimate, $\mathbf{e}'\mathbf{e}/NT$ (note no degrees of freedom correction) and $\mathbf{e}_{LSDV}'\mathbf{e}_{LSDV}//NT$ (again, no correction). A third that is completely different is proposed on page 296 of Wooldridge. Only the second of these is guaranteed to produce a positive estimate of σ_u^2. Show this. For each estimator, show how the residuals are used to compute the two variance component estimators. The Wooldridge estimator appears to use cross observation products (covariances) to estimate a variance. Can you justify this computation? Determine exactly how your software computes the variance components.

Note that the estimator that does not make the degrees of freedom corrections is not, in fact, consistent. The estimator of σ_ε^2 converges to σ_ε^2(T-1)/T. What does this imply? The estimator of $\boldsymbol{\beta}$ based on this estimator is still consistent, since this is just weighted least squares with suboptimal weights as is, for example, OLS. But, it does raise an interesting question about the estimated standard errors. One hopes that T is large enough that the standard errors are nearly correct.

Part IV. The Hausman Test

We have considered two approaches to Hausman's test for random vs. fixed effects. A direct approach compares the random and fixed effects estimators using a Wald test and using Hausman's theoretical result on how to obtain the asymptotic covariance matrix for the difference. A second approach is a 'variable addition test,' in which the group means of the time varying variables are added to the regression (each group mean is repeated for each observation in the group), then an *F* (or Wald) test is used to test the significance of the coefficients on the means. A large *F* weighs against the random effects specification.

1. Using the bank cost data on the course website, carry out this test both ways with respect to the following model

$$\log C_{i,t} = \beta_1 \log Y1_{i,t} + \beta_2 \log Y2_{i,t} + \beta_3 \log Y3_{i,t} + \beta_4 \log Y4_{i,t} + \beta_5 \log Y5_{i,t} + \alpha_i + \varepsilon_{i,t}$$

(Note for the direct test, you use only the first 5 coefficients).

2. Using the preferred model based on part 1., now test the hypothesis of constant returns to scale, that $\beta_1+\beta_2+\beta_3+\beta_4+\beta_5 = 1$.

NOTE: In the data set, variable variable "*C*" is the log of the costs, and variables *Q*1,…,*Q*5 are the logs of the outputs. So, you need not transform the data after reading them into your program. The files are stored at http://people.stern.nyu.edu/wgreene/Econometrics/banks.csv and banks.lpj.

Part V. Algebra for the Two Period Model

1. [This is Wooldridge's problem 10.2, page 334.] Suppose you have T=2 years of data, year 0 and year 1, on the same group of N working individuals. Consider the following model of wage determination:

$$\text{Log}(\text{wage}_{it}) = \theta_1 + \theta_2\, d2_t + \mathbf{z}_{it}'\boldsymbol{\gamma} + \delta_1\, \text{female}_i + \delta_2\, d2_t \times \text{female}_i + c_i + \varepsilon_{it}.$$

The unobserved effect, c_i is allowed to be correlated with z_{it} and with female_i. The variable d2 is a time period indicator; $d2_t = 0$ if t = 0 and $d2_t = 1$ if t = 1. In what follows: assume that

$$E[\varepsilon_{it}|\text{female}_i,\mathbf{z}_{i0},\mathbf{z}_{i1},c_i] = 0,\ t = 0,1.$$

   a. Without further assumptions, what parameters in the log wage equation can be consistently estimated?
   b. Interpret the coefficients θ_1 and θ_2.
   c. Write the log wage explicitly for the two time periods. Show that the differenced equation can be written as

$$\Delta\log(\text{wage}_{it}) = \theta_2 + \Delta\mathbf{z}_{it}'\boldsymbol{\gamma} + \delta_2\, \text{female}_i + \Delta\varepsilon_{it}$$

   where $\Delta\log(\text{wage}_{it}) = \log(\text{wage}_{i1}) - \log(\text{wage}_{i0})$.

2. Continuing part 1, discuss estimation of the model under the 'random' effects assumption. How would you proceed? Can it be done?

3. Note that without the "confounding" effects, $\mathbf{z}_{it}'\boldsymbol{\gamma}$, this is a difference in differences model,
$\delta_2 = \left(\Delta \log(wage) \mid female = 1\right) - \left(\Delta \log(wage) \mid female = 0\right)$

Part VI. The Random Effects Model

1. [Based on Wooldridge, problem 10.5, page 336.]
   a. Consider an extension of the random effects model in which the variance of u_i differs across individuals. How does the covariance matrix of the disturbance vector in the RE model change if the individual component is heteroscedastic?
   b. How would this change the behavior (asymptotic properties) of the OLS estimator and the GLS estimator
   c. Given this modification of the model, how would you modify your estimation and inference procedures?

2. [Based on Wooldridge, problem 10.14, page 340] Suppose we specify the unobserved effects model

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + h_i + \varepsilon_{it},\ i = 1,\ldots,N;\ t = 1,\ldots,T.$$

   $\mathbf{x}_{it}$ is a set of time varying variables while $\mathbf{z}_i$ is a set of time invariant variables. We assume that

   $E[\varepsilon_{it}|\mathbf{x}_i, \mathbf{z}_i, h_i] = 0$, I.e., ε_{it} is uncorrelated with $\mathbf{x}_{is}$ for all periods, as well as with $\mathbf{z}_i$ and h_i.

   $E[h_i \mid \mathbf{x}_i, \mathbf{z}_i] = 0$. (The random effects specification).

   If we use the fixed effects estimator to estimate this random effects model, we are implicitly estimating the parameters of the equation

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it} \text{ where } c_i = \alpha + \mathbf{z}_i'\boldsymbol{\gamma} + u_i.$$

   a. Obtain $\sigma_c^2 = \text{Var}[c_i|\mathbf{z}_i]$. Show that this is at least as large as σ_h^2.
   b. Explain why estimation of the model by fixed effects will lead to a larger estimated variance of the unobserved effect (the disturbance) than if the model is estimated by the random effects procedure.

3. The Lagrange multiplier statistic for testing the hypothesis that $\sigma_u^2 = 0$ in the model $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it}$ appears on Slide 23 of PanelDataNotes-5 (Random Effects). Derive the probability limit of (1/N)LM under the null hypothesis that σ_u^2 is actually zero. Hint: a simpler form to work with is

$$\frac{1}{N}LM = \frac{T}{2(T\text{-}1)}\left[\frac{T\left\{\frac{1}{N}\Sigma_{i=1}^{N}\left(\frac{1}{T}\Sigma_{t=1}^{T}e_{it}\right)^2\right\}}{\frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}e_{it}^2}-1\right]^2.$$

   You can obtain the probability limits of the two sums and use the Slutsky theorem to obtain the end result.

Part VII. The Fixed Effects Model

[Based on Wooldridge, problem 10.3, page 335.] For T = 2, consider the standard unobserved effects model,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it},\ i = 1,\ldots,N;\ t = 1,2.$$

Let $\mathbf{b}_{FE}$ and $\mathbf{b}_D$ denote the fixed effects (within) and first difference estimators, respectively.

1. Show that the two estimators are numerically identical.
2. Show that the estimates of the disturbance variance from the two estimators are also identical.