Department of Economics

Econometric Analysis of Panel Data

Professor William Greene **Phone: 212.998.0876**
Office: KMC 7-78 **Home page:www.stern.nyu.edu/~wgreene**
Email: wgreene@stern.nyu.edu
URL for course web page:
www.stern.nyu.edu/~wgreene/Econometrics/PanelDataEconometrics.htm

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 Assignment 1 (using the Cornwell and Rupert data).

(1) Test the hypothesis of 'no effects' vs. 'some effects' using the results given below.
Use the LM statistic. It strongly rejects the no effects model.
(2) Explain in precise detail the difference between the fixed and random effects models.
See Text and class notes
(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.
The H statistic is 2554.11 with 4 degrees of freedom. Strongly rejects REM
(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.)
F=[(.9054373-.2451121)/(595-1)] / [(1-.9054373)/(4165-595-4)]=41.921. The 95% critical value is 1.106 so the homogeneity hypothesis is rejected.
(5) Using the fixed effects estimator, test the hypothesis that all of the coefficients in the model save for the constant term are zero. Show all computations, and the appropriate degrees of freedom for *F*.
The sum of squares for the model with only the constant terms is given in the ANOVA table, 240.651. Sum of squares for the full model is 83.868.
F = [(240.651 – 83.868)/4] / [83.868/(4165 – 595 – 4)] =1666.572.
The 95% critical value is 2.374.
(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.
The fit goes up dramatically and the new estimates are completely different.

```
+----------------------------------------------------+
| OLS Without Group Dummy Variables                  |
| LHS=LWAGE    Mean                 =   6.676346     |
|              Standard deviation   =   .4615122     |
| Model size   Parameters           =          5     |
|              Degrees of freedom   =       4160     |
| Residuals    Sum of squares       =   669.5138     |
|              Standard error of e  =   .4011743     |
| Fit          R-squared            =   .2451121     |
|              Adjusted R-squared   =   .2443862     |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel Data Analysis of LWAGE      [ONE way]        |
|           Unconditional ANOVA (No regressors)      |
| Source      Variation   Deg. Free.     Mean Square |
| Between       646.254         594.     1.08797     |
| Residual      240.651        3570.     .674093E-01 |
| Total         886.905        4164.     .212994     |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
OCC	-.36608081	.01346550	-27.187	.0000	.51116447
UNION	.11154686	.01402315	7.954	.0000	.36398559
MS	.32218316	.01629572	19.771	.0000	.81440576
EXP	.00805812	.00057594	13.991	.0000	19.8537815
Constant	6.40050047	.01785232	358.525	.0000	

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Model size   Parameters           =        599     |
|              Degrees of freedom   =       3566     |
| Residuals    Sum of squares       =   83.86816     |
|              Standard error of e  =   .1533585     |
| Fit          R-squared            =   .9054373     |
|              Adjusted R-squared   =   .8895796     |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
OCC	-.02406298	.01384128	-1.738	.0821	.51116447
UNION	.03515301	.01502985	2.339	.0193	.36398559
MS	-.03226210	.01909579	-1.689	.0911	.81440576
EXP	.09672164	.00119030	81.258	.0000	19.8537815

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235188D-01  |
|             Var[u]              =   .137422D+00  |
|             Corr[v(i,t),v(i,s)] =   .853867      |
| Lagrange Multiplier Test vs. Model (3) = 4352.48 |
| ( 1 df, prob value =  .000000)                   |
| Fixed vs. Random Effects (Hausman)     = 2554.11 |
| ( 4 df, prob value =  .000000)                   |
+--------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
OCC	-.10630712	.01284206	-8.278	.0000	.51116447
UNION	.03971116	.01385277	2.867	.0041	.36398559
MS	-.02642760	.01737054	-1.521	.1282	.81440576
EXP	.05949249	.00091374	65.109	.0000	19.8537815
Constant	5.55660438	.02834081	196.064	.0000	

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. 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?

No effect on R^2 or sum of squares. The model is $\mathbf{y} = [\mathbf{X},\mathbf{D}]\,(\boldsymbol{\beta}',\boldsymbol{\alpha}')'+\boldsymbol{\varepsilon}$. The different normalizations amount to linear transformations of $\mathbf{D}$, say $\mathbf{D}^*=\mathbf{DP}$ where $\mathbf{P}$ is a nonsingular $N\times N$ matrix that mixes the columns of $\mathbf{D}$. The least squares coefficients will be $\mathbf{b}$ (the original one) and $\mathbf{a}^* = \mathbf{P}^{-1}\mathbf{a}$ where $\mathbf{a}$ is the original constants. Same residuals, same R^2

Part III. Estimating Variance Components

Greene (2008), Wooldridge (2000, page 26), 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 (on pages 203-205) based on the degrees of freedom corrected OLS and FE estimators. A different one is used by the TSP computer program (and *NLOGIT* after Bruno and DeBonis), namely using 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 261 of Wooldridge. Only one of these (the *TSP/LIMDEP* estimator) is guaranteed to produce a positive estimate of σ_u^2. Show this. (In fact, I have never seen the Wooldridge estimator implemented either in software or in any application.) 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? If you are not using *NLOGIT* (since the answer for that appears above), determine exactly how your software computes the variance components.

Note that the TSP/NLOGIT estimator is not, in fact, consistent. The estimator of σ_ε^2 converges to $\sigma_\varepsilon^2(T-1)/T$. What does this imply? The estimator of β 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.

$\mathbf{e}'\mathbf{e}/NT$ must be greater than $\mathbf{e}_{LSDV}'\mathbf{e}_{LSDV}//NT$ because the first is a restricted regression based on the second – equal constant terms. The sum of squares will never be smaller when restrictions are imposed. Whether degrees of freedom corrected or not, the OLS estimator estimates $\sigma_u^2+\sigma_\varepsilon^2$ while the LSDV estimator estimates σ_ε^2. Thus, σ_u^2 is estimated by subrtracting the second from the first. For the Wooldridge estimator, $\text{Cov}[\varepsilon_{it}+u_i,\varepsilon_{is}+u_i] = \sigma_u^2$. So, the sample covariance of the OLS residuals is estimating σ_u^2. There are NT(T-1)/2 unique pairs of residuals to be multiplied, so the sum divided by this provides another estimator. This is based on the method of moments.

The fact that the estimator of σ_ε^2 converges to something less than σ_ε^2 means that the two step GLS estimator, while still consistent, uses the "wrong" weights. As such, the estimated asymptotic covariance matrix for the estimator is incorrect. (Just like OLS, which also uses the wrong weights.)

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).

```
name;x=q1,q2,q3,q4,q5$
regr;lhs=c;rhs=x,one;fixed;panel;pds=5$
matr;bf=b;vf=varb$
regr;lhs=c;rhs=x,one;random;panel;pds=5$
matr;br=b(1:5);vr=varb(1:5,1:5)$
matrix ; db=bf-br;dv=vf-vr ; list;h=db'<dv>db$
Matrix H        has  1 rows and  1 columns.
               1
        +--------------
       1|    1.10057
```

```
create;q1b=groupmean(q1,pds=5)$
create;q2b=groupmean(q2,pds=5)$
create;q3b=groupmean(q3,pds=5)$
create;q4b=groupmean(q4,pds=5)$
create;q5b=groupmean(q5,pds=5)$
regr;lhs=c;rhs=q1b,q2b,q3b,q4b,q5b,x,one;random;panel;pds=5$
matrix ; bm=b(1:5);vm=varb(1:5,1:5) ; list;wu=bm'<vm>bm$
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .158113D+00  |
|             Var[u]              =   .350217D-01  |
|             Corr[v(i,t),v(i,s)] =   .181333      |
| Lagrange Multiplier Test vs. Model (3) =    2.72 |
| ( 1 df, prob value =  .098958)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =           2.72 |
|             Sum of Squares          .482837D+03  |
|             R-squared               .859843D+00  |
+--------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Q1B	-.02822	.03963219	-.712	.4764	8.5876310
Q2B	.01988	.03555521	.559	.5760	10.093183
Q3B	-.04655	.05130903	-.907	.3643	9.7194921
Q4B	-.00345	.02161648	-.159	.8733	7.7829046
Q5B	.05098	.04908093	1.039	.2990	7.1371551
Q1	.08027***	.01293400	6.206	.0000	8.5876310
Q2	.38900***	.01217435	31.952	.0000	10.093183
Q3	.11681***	.01726313	6.766	.0000	9.7194921
Q4	.07838***	.00695211	11.274	.0000	7.7829046
Q5	.31522***	.01721915	18.306	.0000	7.1371551

```
|Constant|    3.00689***       .25552236    11.768   .0000              |
+--------+--------------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.                 |
+-----------------------------------------------------------------------+
```

```
--> matrix ; bm=b(1:5);vm=varb(1:5,1:5) ; list;wu=bm'<vm>bm$
Matrix WU       has  1 rows and  1 columns.
               1
        +--------------
       1|    3.51842
```

(2) Using the preferred model based on the outcome of part (1), now test the hypothesis that $\beta_1+\beta_2+\beta_3\beta_4+\beta_5 = 1$.

Using the random effects model and a wald test

```
regr;lhs=c;rhs=x,one;random;panel;pds=5$
matrix ; r = [1,1,1,1,1,0];q=[1];
d=r*b-q ; vd = r*varb*r' ; list; w = d'<vd>d $
Matrix W        has  1 rows and  1 columns.
               1
        +--------------
       1|    6.48397
```

Critical value is 3.84, so the hypothesis is rejected.

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.

Part V. Heteroscedasticity in the Fixed Effects Model

As we saw in class, there is ample evidence of heteroscedasticity in the Baltagi/Griffin gasoline demand data. Consider the model

$\log G_{i,t} = \beta_1 + \beta_2 \log Y_{i,t} + \beta_3 \log P_{i,t} + \beta_4 \log C_{i,t} + \alpha_i + \varepsilon_{i,t}$
$E[\alpha_i|\mathbf{x}_i] = g(\mathbf{x}_i)$ (i.e., a 'fixed effects' model)
$E[\varepsilon_{i,t}|\mathbf{x}_i] = 0$
$Var[\varepsilon_{i,t}|\mathbf{x}_i] = \sigma_i^2$

Note that "i" above refers to the country, not the period. There are 18 countries in the data set.

(1) How would you estimate the parameters of this model? Suggest a two step, efficient estimator.

(2) Using the Baltagi/Griffin data, compute the estimates of the model and report your results. Does the assumption of heteroscedasticity substantially change the model results compared to the simpler fixed effects model?

(1) You can use a two step FGLS estimator. At the first step, use OLS (with dummy variables) to obtain a set of residuals. Then, using country specific residual vectors, estimate the country specific disturbamnce variance. Finally, use weighted least squares at the second step.

(2)

```
sample ; all $
regress;lhs=lgaspcar;rhs=one,lincomep,lrpmg,lcarpcap;
panel;fixed;pds=19;res=e$
matrix ; bfe=b $
create ; e2 = e*e $
create ; vi  = group mean(e2,pds=19)$
create ; wi = 1/vi $
regr ; lhs= lgaspcar
     ; rhs=one,lincomep,lrpmg,lcarpcap,d*
     ; wts = wi $
```

Unweighted

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
LINCOMEP	.66225***	.07338604	9.024	.0000	-6.1394254
LRPMG	-.32170***	.04409925	-7.295	.0000	-.5231032
LCARPCAP	-.64048***	.02967885	-21.580	.0000	-9.0418047

Weighted

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	3.21519***	.11916619	26.981	.0000	
LINCOMEP	.57507***	.02926687	19.649	.0000	-5.8479021
LRPMG	-.27967***	.03518536	-7.949	.0000	-.8773696
LCARPCAP	-.56540***	.01613491	-35.042	.0000	-8.3474219

Seems not to make very much difference. The estimates appear to become somewhat more precise. Of course, if there really is heteroscedasticity, then the OLS standard errors are inappropriate.
The corrected results follow.

```
create ; dli=groupdevs(lincomep,pds=19)$
create ; dlr=groupdevs(lrpmg,pds=19)$
create ; dlc=groupdevs(lcarpcap,pds=19)$
name;dx=dli,dlr,dlc$
matr;vols=<dx'dx> * dx'[vi]dx * <dx'dx> $
matr;stat(bfe,vols,dx)$
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
|DLI     |     .66225***       .06238100    10.616   .0000 |
|DLR     |    -.32170***       .05197389    -6.190   .0000 |
|DLC     |    -.64048***       .03035538   -21.099   .0000 |
+--------+---------------------------------------------------+
```

Part VI. Algebra for the Two Period Model

(1) [This is Wooldridge's problem 10.2, page 292.] Suppose you have T=2 years of data 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; $d2_t = 1$ if t=2 and 0 if t=1. In what follows: assume that

$$E[\varepsilon_{it}|\text{female}_i,\mathbf{z}_{i1},\mathbf{z}_{i2},c_i] = 0,\ t = 1,2.$$

a. Without further assumptions, what parameters in the log wage equation can be consistently estimated?

Take first differences. θ_1 falls out. $d2_2\text{-}d2_1 = 1$, so it becomes the constant, θ_2 is estimable. As long as z_{it} varies over the periods, γ is estimable. δ_1 is not estimable as female is time invariant. d2×female becomes "female" so δ_2 is estimable.

b. Interpret the coefficients θ_1 and θ_2.

Intercept in first period and shift in intercept from first period to second. Second period intercept is $\theta_1+\theta_2$.

c. Write the log wage explicitly for the two time periods. Show that the differenced equation can we written as

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

where $\Delta\log(\text{wage}_{it}) = \log(\text{wage}_{i2}) - \log(\text{wage}_{i1})$.

See part a.

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

The first differenced model in part c. can be used to estimate the variance of ε_{it} (times 2). Simple OLS will estimate $\sigma_\varepsilon^2 + \sigma_c^2$, so the variances are estimable. After that, random effects estimation is done in the usual fashion.

Part VII. The Random Effects Model

(1) [Based on Wooldridge, problem 10.5, page 293.]
(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?
If the true variance of u_i is heterogeneous, then the covariance matrix for the full disturbance vector is still block diagonal, with each block having its own common variance component. Note, as discussed in class, this model is not estimable as such.

(b) How would this change the behavior (asymptotic properties) of the OLS estimator and the GLS estimator
It would change the asymptotic variances, but consistency and asymptotic normality would be unchanged.

(c) Given this modification of the model, how would you modify your estimation and inference procedures?
You would have to know the variances in advance. They are not estimable. If they were, it is a small change in the estimator of the RE model to have different variances for the different groups.

(2) The problem has a couple (fatal) typos in it. APOLOGIES. [Based on Wooldridge, problem 10.14, page 297] 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

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

a. Obtain $\sigma_c^2 = \text{Var}[c_i|\mathbf{z}_i]$. Show that this is at least as large as σ_h^2.

The correct question requests $\text{Var}[c_i]$ not $\text{Var}[c|z_i]$. $\text{Var}[c_i] = \text{Var}[\mathbf{z}_i'\boldsymbol{\gamma}] + \sigma_h^2 \geq \sigma_h^2$.

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.

The fixed effects residual variance estimator of c_i will estimate $\text{Var}[c_i]$ while the random effects estimator will be estimating σ_h^2. Note, “variance of the unobserved effect” means variance of c_i or h_i for the two cases, not ε_{it}.

(3). The Lagrange multiplier statistic for testing the hypothesis that $\sigma_u^2 = 0$ in the model $y_{it} = x_{it}'\beta + u_i + \varepsilon_{it}$ appears on slide 27 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 $\frac{1}{N}LM = \frac{T}{2(T\text{-}1)}\left[\frac{\Sigma_{i=1}^N(\Sigma_{t=1}^T e_{it})^2}{\Sigma_{i=1}^N\Sigma_{t=1}^T e_{it}^2} - 1\right]^2$. You can obtain the probability limits of the two sums and use the Slutsky theorem to obtain the end result.

Manipulate it

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

$$= \frac{T}{2(T\text{-}1)}\left[\frac{\frac{1}{N}\Sigma_{i=1}^N(\Sigma_{t=1}^T e_{it})^2}{\frac{1}{N}\Sigma_{i=1}^N\Sigma_{t=1}^T e_{it}^2} - 1\right]^2 \quad \text{(divide n and d by N)}$$

$$= \frac{T}{2(T\text{-}1)}\left[\frac{T\frac{1}{N}\Sigma_{i=1}^N(\frac{\Sigma_{t=1}^T e_{it}}{T})^2}{\frac{1}{NT}\Sigma_{i=1}^N\Sigma_{t=1}^T e_{it}^2} - 1\right]^2 \quad \text{(multiply n by T/T}^2\text{ and d by 1/T)}$$

Inside the brackets, the denominator converges to σ_ε^2 (obviously). The numerator is T times an estimator of $E[\overline{e}_i^2]$ which is T times σ_ε^2/T. Thus, the fraction in brackets is converging to 1 and 1-1 is converging to zero, as is $(1\text{-}1)^2$. Therefore, 1/N × LM converges to zero. It makes sense, since the limiting distribution of LM, itself, is a chi-squared with one degree of freedom, and a finite variance, 2. As discussed in class, in the alternative case, 1/N×LM converges to $[T(T\text{-}1)/2]\rho^2$ where $\rho = \sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2)$.

Part VIII. The Fixed Effects Model

[Based on Wooldridge, problem 10.3, page 292.] 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.
a. Show that the two estimators are numerically identical.
b. Show that the estimates of the disturbance variance from the two estimators are also identical.

Deviations from group means are x1 – (x1+x2)/2 = (x1-x2)/2
First differences are x1-x2. For the FE estimator, the variables on both sides of the equals sign are exactly ½ those of those for the first difference estimator. Since we are using OLS after the transformation, this leaves the estimator unchanged. With the same estimator, the same residuals result, as does the variance estimator.

Part IX. Homogeneity

Alicia Munnell's statewide productivity data appear on the course website. These data are for the 48 states (excluding Alaska and Hawaii) and the 17 years 1970-1986. Consider the model,

$$\ln GSP_{it} = \beta_1 + \beta_2 \ln EMP_{it} + \beta_3 \ln P\_Cap_{it} + \varepsilon_{it}.$$

1. Compute the estimates of the parameters, using least squares.

```
--> create ; y=log(gsp);x1=log(emp);x2=log(p_cap)$
--> regress ; lhs=y;rhs=one,x1,x2$
```

```
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=Y        Mean                 =   10.50885      |
|              Standard deviation   =   1.021132      |
| WTS=none     Number of observs.   =        816      |
| Model size   Parameters           =          3      |
|              Degrees of freedom   =        813      |
| Residuals    Sum of squares       =   13.37471      |
|              Standard error of e  =   .1282617      |
| Fit          R-squared            =   .9842615      |
|              Adjusted R-squared   =   .9842228      |
| Model test   F[  2,   813] (prob) =******* (.0000)  |
| Diagnostic   Log likelihood       =   519.4541      |
|              Restricted(b=0)      =  -1174.417      |
|              Chi-sq [  2]  (prob) =3387.74 (.0000)  |
| Info criter. LogAmemiya Prd. Crt. =  -4.103696      |
|              Akaike Info. Criter. =  -4.103696      |
|              Bayes Info. Criter.  =  -4.086400      |
| Autocorrel   Durbin-Watson Stat.  =   .1732736      |
|              Rho = cor[e,e(-1)]   =   .9133632      |
| Model was estimated Feb 19, 2009 at 11:22:15AM      |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	2.21180***	.07919412	27.929	.0000	
X1	.62265***	.01993772	31.230	.0000	6.9784979
X2	.40829***	.02154257	18.953	.0000	9.6792058

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

2. We are interested in determining if the coefficients of the model changed over time – we wish to test this statistically. To do this, compute the regression separately for the 17 years, then test the hypothesis of homogeneity. (Note, there are three ways to do this. The most straightforward will be to use the Chow test, based on the sums of sqaured residuals.)

```
create ; y=log(gsp);x1=log(emp);x2=log(p_cap)$
regress ; lhs=y;rhs=one,x1,x2$
calc; sspool=sumsqdev; sschow=0$
create ; t=yr-1969$
proc=chowtest(i)$
regress ; for[t=i] ; quietly ; lhs=y;rhs=one,x1,x2 $
calc ; sschow=sschow+sumsqdev$
endproc$
exec ; i = 1,17$
calc ; list ; ftest = ((sspool - sschow)/(16*3))/(sschow/(n-17*3))$
calc ; list ; ftb(.95,48,(n-51))$
--> calc ; list ; ftest = ((sspool - sschow)/(16*3))/(sschow/(n-17*3))$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 FTEST   =       1.482332
--> calc ; list ; ftb(.95,48,(n-51))$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =       1.374645

--> calc ; list; pvalue = 1- fds(ftest,48,(n-51))$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 PVALUE  =        .020554
```

3. Compute fixed and random (state) effects models and determine which of the two would be the preferred specification.

Note: For NLOGIT's purposes, you have to rearrange the data. They are arranged by state in blocks of 17 observations, but for this exercise, you need them arranged by year in blocks of 48 states. To obtain this form, just use the commands

SAMPLE ; All $
CREATE ; Obs = Trn(1,1) $
SORT ; LHS= YR ; RHS = * $

(Note, the states may end up out of order, but that does not matter.) Then, you can analyze the data as a panel with ;PDS=48. If you need to restore the original form of the data, the instruction

SORT ; Lhs = Obs ; Rhs = * $

will put the data back the way they were when you started.

It looks like the data slightly favor the fixed effects model.

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=Y        Mean                 =   10.50885     |
|              Standard deviation   =   1.021132     |
| WTS=none     Number of observs.   =        816     |
| Model size   Parameters           =         19     |
|              Degrees of freedom   =        797     |
| Residuals    Sum of squares       =   12.64587     |
|              Standard error of e  =   .1259636     |
| Fit          R-squared            =   .9851192     |
|              Adjusted R-squared   =   .9847831     |
| Model test   F[ 18,   797] (prob) =2931.21 (.0000) |
| Diagnostic   Log likelihood       =   542.3163     |
|              Restricted(b=0)      =  -1174.417     |
|              Chi-sq [ 18]  (prob) =3433.47 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -4.120507     |
|              Akaike Info. Criter. =  -4.120515     |
|              Bayes Info. Criter.  =  -4.010976     |
| Model was estimated Feb 19, 2009 at 11:29:19AM     |
| Estd. Autocorrelation of e(i,t)    -.105547        |
+----------------------------------------------------+

+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      17 |
|                Smallest   48,   Largest         48 |
|                Average group size            48.00 |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|X1      |     .60654***      .01979576    30.640   .0000   6.9784979|
|X2      |     .42319***      .02133388    19.836   .0000   9.6792058|
+--------+------------------------------------------------------------+
| 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    -1174.41747  .8498088814D+03    .0000000 |
|(2)  Group effects only    -1166.97921  .8344563655D+03    .0180658 |
|(3)  X - variables only      519.45415  .1337470662D+02    .9842615 |
|(4)  X and group effects     542.31631  .1264586955D+02    .9851192 |
+------------------------------------------------------------------+
|                        Hypothesis Tests                          |
|         Likelihood Ratio Test           F Tests                  |
|         Chi-squared   d.f.   Prob.        F    num. denom.   P value |
|(2) vs (1)     14.877     16  .5337        .92    16     799   .54717 |
|(3) vs (1)  3387.743      2  .0000  25421.90      2     813   .00000 |
|(4) vs (1)  3433.468     18  .0000   2931.21     18     797   .00000 |
|(4) vs (2)  3418.591      2  .0000  25897.11      2     797   .00000 |
|(4) vs (3)     45.724     16  .0001      2.87     16     797   .00014 |
+------------------------------------------------------------------+

+--------------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)           |
| Estimates:  Var[e]              =   .158668D-01        |
|             Var[u]              =   .584216D-03        |
|             Corr[v(i,t),v(i,s)] =   .035512            |
| Lagrange Multiplier Test vs. Model (3) =   21.32       |
| ( 1 df, prob value =  .000004)                         |
| (High values of LM favor FEM/REM over CR model.)       |
| Baltagi-Li form of LM Statistic =          21.32       |
| Fixed vs. Random Effects (Hausman)     =   11.40       |
| ( 2 df, prob value =  .003352)                         |
| (High (low) values of H favor FEM (REM).)              |
|             Sum of Squares            .133803D+02      |
|             R-squared                 .984255D+00      |
+--------------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|X1      |     .61246***      .01971690    31.063   .0000   6.9784979|
|X2      |     .41771***      .02126894    19.639   .0000   9.6792058|
|Constant|    2.19170***      .07818382    28.033   .0000            |
+--------+------------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.               |
+---------------------------------------------------------------------+
```