CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TAs: Jelena Plazonja, Gega Todua
Due date: Monday 02/06/2014 (13:30)

ECONOMETRICS II

HOMEWORK #2

Problem 1. Introductory Wooldridge (2^{nd} edition) 14.2.

Problem 2. Advanced Wooldridge 10.12 (a) to (d) plus:

(e) After estimating FE model, test the hypothesis that the constant term is the same for all years.

(f) Carry out Hausman's specification test for the random effects versus the fixed effects model.

Problem 3. Suppose $y_{it} = \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + e_{it}$, and e_{it} is serially uncorrelated with constant variance, i.e. $\mathbb{E}(e_{it}e_{is}) = \sigma_e^2$ if $s = t$, and 0 otherwise. Show that first differencing induces negative serial correlation and find $corr(\Delta e_{it}, \Delta e_{i,t+1})$.

Problem 4. Consider the standard unobserved effects model

$$y_{it} = x_{it} \cdot \beta + \alpha_i + u_{it}$$

with two periods $t = 1, 2$. Let $\hat{\beta}^{FE}$ and $\hat{\beta}^{FD}$ denote the fixed effects and first difference estimators respectively.

(a) Show that the FE and FD estimates are numerically identical.

(b) Show that the error variance estimates from the FE and FD methods are numerically identical.

Problem 5. Introductory Wooldridge (2^{nd} edition) 14.6.

Problem 6. Use the data set *investment.dta* to estimate the effect of the real value of the firm (*value*) and real capital stock (*cstock*) on the real gross investment (*inv*). Data is available for 4 firms and for the period 1935-1954. You should report and briefly comment all your results obtained during estimations.

(a) Estimate the pooled OLS model. Comment the signs of the coefficients.

(b) Estimate the LSDV model. Is the constant term the same for each firm? Which model is better, the pooled OLS or LSDV? (Hint: use F-test or LR test)

(c) Estimate the model using the first difference estimator and the fixed effect (by demeaning) estimator and compare.

(d) Estimate the FE using standard command in STATA. Compare the results of the FE estimation obtained from (b), (c) and (d).

(e) Using the estimated residuals from (a) and (b), estimate the variances $\hat{\sigma}_{OLS} = \hat{\sigma}_u$ and $\hat{\sigma}_{LSDV} = \hat{\sigma}_\epsilon$ (Hint: be careful about the degrees of freedom).

(f) Compute λ and estimate the RE model by quasi-demeaning. Then estimate the RE model using standard command in STATA and compare.

(g) Compare pooled OLS and RE estimation results. Which model is better? (Hint: use Breusch-Pagan LM test)

(h) Compare FE and RE estimation results. Carry out Hausman's specification test to choose between the two models.

(i) Choose the best model among the ones estimated and justify your answer.

Problem 7. Consider the following panel data model

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

where $i = 1, \ldots, N$ denotes individuals and $t = 1, \ldots, T$ denotes time, and

$$u_{it} = \omega_i + e_{it} + \gamma e_{it-1}$$

$\forall i$: $\mathbb{E}(\omega_i) = 0, VAR(\omega_i) = \sigma^2_\omega$, and $\mathbb{E}(\omega_i \omega_j) = 0$, $\forall i \neq j$
$\forall t$: $\mathbb{E}(e_{it}) = 0, VAR(e_{it}) = \sigma^2_e$, and $\mathbb{E}(e_{it} e_{jt}) = 0$, $\forall i \neq j$
$\forall i$: $\mathbb{E}(e_{it} e_{is}) = 0$, $\forall t \neq s$, and ω_i and e_{it} are independent of each other and also independent of x.

(a) Write down the variance-covariance matrix of the composite error term u_{it} for $T = 3$ and $N = 2$.

(b) Is the pooled OLS estimate of βconsistent? Is it efficient?

Problem 8. Suppose that, for one semester, you can collect the following data on a random sample of college juniors and seniors for each class taken: a standardized final exam score, percentage of lectures attended, a dummy variable indicating whether the class is within the student's major, cumulative grade point average prior to the start of the semester, and SAT score.

(i) Why would you classify this data set as a cluster sample? Focus on sampling students rather than sampling their scores, and think of correlation that may arise for each student rather than for each class or any other cluster of that kind.

(ii) Write a model (unobserved effects model) that explains final exam performance in terms of attendance and the other characteristics. Use s to subscript student and c to subscript class. Which variables do not change within a student?

(iii) If you pool all of the data together and use OLS, what are you assuming about unobserved student characteristics that affect performance and attendance rate? What roles do SAT score and prior GPA play in this regard?

(iv) If you think SAT score and prior GPA do not adequately capture student ability, how would you estimate the effect of attendance on final exam performance?

Hint: look up the following commands in STATA: *xt*, *xtreg*, *xttest*0, *hausman*, and regression options *robust* and *cluster*.