ECON 861
Empirical MicroEconomics II
Winter Term 2012/13

Assignment 2
Due: In my office (DH 306) on ***February 14th*** *2013*

Rules of the 'game':

1) you can work in groups (actually, I encourage you to do so), but you MUST submit your own set of answers

2) you MUST list the people you worked with in the front page

3) late submissions will NOT be accepted. If you cannot make it to the class, you can pop into my office (DH306) and submit the assignment whenever you are ready, before the 14th

4) I will keep the set of answers you are going to hand in, so please make copies if you think you are going to need them

5) you have to dowload the datasets for the exercises from the webpage of the course:
http://qed.econ.queensu.ca/pub/faculty/cozzi/Webpage/Cozzi_ECON861.htm

Devote some time to give the graphs, plots and tables a format easy to understand. Also the way you present your answers matter for the final grade. Even if a question is mainly analytical, briefly explain what you are doing, stressing the economic meaning of the various steps.

Ex.1: The Lindberg–Levy Central Limit Theorem.

Consider random sampling from the uniform $[-5, 5]$ distribution.

Draw 500 samples of 3, 6, 12 and 100 observations from this population (hint: you will need to write a do-loop in STATA, use the built-in random number generator from the Uniform $[0, 1]$ distribution, and rescale your draws). Compute the sample means for each. For each mean, compute the normalized variable $z_{i,n} = \sqrt{n}\,(\overline{x}_{i,n} - \mu)$, where $i = 1, ..., 500$, n is 3, 6, 12 or 100, $\overline{x}$ is the sample mean and μ is the population mean.

Present and discuss your findings by plotting the four sampling distributions and their kernel densities. How far are the results of each sample from normality (hint: check higher order moments, such as the skewness and the kurtosis, and rely on a formal test based on them)?

Ex.2: Estimation of Static Panel Data models.

Download the first dataset: PS2_1.dta. It includes a set of variables on a sample of airline companies and their fares. We are interested in estimating the model

$$\log(fare_{it}) = \alpha_t + \beta_1 concen_{it} + \beta_2 \log(dist_{it}) + \beta_3 \left[\log(dist_{it})\right]^2 + \eta_i + u_{it}$$

where α_t means that we allow for different year intercepts.

a) Estimate the above equation by pooled OLS. What is the estimated effect of *concen*? Give an economic explanation.

b) What is the usual OLS 95% confidence interval for β_1? Why is it probably not reliable? Find the robust 95% confidence interval for β_1. Comment.

c) Now estimate the equation using Random Effects. How does the estimate of β_1 change?

d) Now estimate the equation using Fixed Effects. What is the FE estimate of β_1? Why is it fairly similar to the RE estimate?

e) Test formally the RE vs. FE specification.

f) Name at least one characteristic of a route that is captured by η_i. Might this be correlated with $concen_{it}$?

g) Comment on your findings, and on the relationship between concentration on a route and increases in airfare in particular.

Ex.3: Dynamic Panel Data.

Assume you have a sample with $T = 5$ and $N = 28456$. Consider the simple AR(1) panel data model, $y_{it} = \alpha y_{it-1} + \varepsilon_{it}$, with the error term being $\varepsilon_{it} = \eta_i + u_{it}$. Assume that u_{it} is not *iid* but it's an MA(1) process instead.

a) What happens to the structure of the error term for the model in first differences?

b) How can you get a consistent estimate of the parameter α?

c) What problems arise in this set-up? How can you fix them?

d) Write the explicit formula for α, when you consider a model which is exactly identified, i.e. you rely on only one moment condition.

Ex.4: Estimation of Dynamic Panel Data models.

Download the second dataset: PS2_2.dta. Source: The Panel Study of Income Dynamics, taken from Cornwell and Rupert (1988).

Description: Balanced Panel Data, 595 individuals over 7 years, 1976-1982.

Variables:

(1) EXP = Years of potential full time work experience (i.e. EXP=age-6).
(2) WKS = Weeks worked.
(3) OCC = (OCC=1, if the individual is in a blue collar occupation).
(4) IND = (IND=1, if the individual works in a manufacturing industry).
(5) SOUTH = (SOUTH=1, if the individual resides in the South).
(6) SMSA = (SMSA=1, if the individual resides in a standard metropolitan statistical area).
(7) MS = (MS=1, if the individual is married).
(8) FEM = (FEM=1, if the individual is female).
(9) UNION = (UNION=1, if the individual's wage is set by a union contract).
(10) ED = Years of education.
(11) BLK = (BLK=1, if the individual is black).
(12) LWAGE = Logarithm of wage.

Part a)

- Estimate by OLS a dynamic wage equation, using as explanatory variables lagged wages, education, work experience, work experience squared, and time dummies.

Part b)

- Estimate the model using the within groups estimator (use the **egen** function in STATA to construct individual means of the data, and perform the within groups transformation on wages.)
- Will you include education? Will you include time dummies? Will you include experience and experience squared?
- Explain what the coefficients on the time dummies mean.

Part c)

- Create the first difference of wages. Estimate the model in first differences, using both OLS and an Instrumental Variable estimator of your choice.
- Will you include time dummies? Will you include experience and experience squared?
- What instruments did you use? Are the results very different from the ones in part a and part b?
- If you used more moments than parameters to be estimated, test for overidentifying restrictions.
- Test for first-order serial correlation of the errors.