CERGE-EI, Summer 2014
Econometrics II

Midterm Exam

Please answer all questions below in a clear and precise way. Make sure your handwriting is legible. The exam has 3 pages; make sure you have all of them. There are 9 questions summing to 100 points. This is a closed book exam and no student interactions are allowed. *Good luck!*

1. (5 points) Using econometric model of your choice explain the case of non-linear relationship between control and dependent variables.

2. (5 points) Heteroscedasticity in errors leads to biased estimates of the regression coefficients and their standard errors. True or False? Briefly explain.

3. (10 points) Write three econometric models capturing 1) *overall*; 2) *within*; and 3) *between* variation.

4. (20 points) Consider the following earnings equation:

$$\begin{aligned} log(wage)_i &= \beta_1 + \beta_2 educ_i + \beta_3 exp_i + \gamma abil_i + v_i \\ &\equiv x\beta + \gamma abil_i + v_i \end{aligned} \tag{1}$$

where, *educ* is education; *exp* is experience and, *abil* is ability.

Assume that ability can not be observed, and that we know the following:
$E(x|v) = 0$; $E(abil|v) = 0$; $Cov(educ, abil) \neq 0$ and, none of the variables are perfectly correlated.

a) What happens if we estimate (1) without the ability variable? Are $\hat{\beta}_2$, and $\hat{\beta}_3$ consistent?

b) Now suppose we are interested in estimating the effect of experience on earnings. Somebody tells you that since eduction is correlated with ability it would be better to omit education from the regression, and estimate it with just experience. Is this a good idea? Briefly explain.

c) Suppose you can find two variables z_1 and z_2, and you think that they may be valid instruments for education. Think of (z_1, z_2) as for example mother's and father's education. What are the requirements to be satisfied by (z_1, z_2) so that the instruments are valid?

d) Suppose your classmate estimated (1) with 2SLS procedure using the following procedure: In the first stage your classmate regressed education on z_1 and z_2, and saved the fitted values $\widehat{educ}$. In the second stage, he or she estimated equation by OLS:

$$log(wage)_i = \beta_1 + \beta_2 \widehat{educ}_i = \beta_x exp_i + \gamma abil_i + v_i$$

Do you agree with your classmate's steps?

5. (10 points) Suppose we have

$$y = \alpha + \beta x + u$$

Where x is measured with error. Now, let's say we have two different measurements of x:

$$x_1 = x + e_1$$
$$x_2 = x + e_2$$

$E(e_1 u) = E(e_2 u) = E(e_1 e_2) = 0;\ cov(x, u) = 0$

How do we solve for the measurement error in the case when we have two different measurements of x?

6. (20 points) Consider the following model: $y_{it} = \beta_0 + \beta_1 x_{it} + e_{it}$ for panel data with N individuals observed over T time periods.

   (a) Under what conditions are the OLS estimates of the two parameters consistent? under what conditions are they efficient?

   (b) Assume that the error term e_{it} consists of two parts, a zero mean time invariant individual effect u_i and a time varying individal specific error term ϵ_{it}, which has a zero mean, constant variance and is independent across time and uncorrelated with both u_i and x_i.Are the pooled OLS estimates of β_0 and β_1 consistent? Are they efficient?

   (c) Consider the case when $cov(u_i, x_{it}) = 0$. Propose a consistent and an efficient estimator of β_1 and describe the estimation steps. Also, consider the case when $cov(u_i, x_{it}) = \sigma_{ux} \neq 0$. Propose a consistent estimator of β_1 and describe the estimation steps.

   Consider the following modification of the random effects model by the addition of a time lagged disturbance. thus,

$$e_{it} = u_i + \epsilon_{it} + \gamma \epsilon_{i,t-1}$$

   where
   $\mathbb{E}(\epsilon_{it}) = \mathbb{E}(u_i) = 0$
   $\mathbb{E}(\epsilon_{it} u_j) = 0$ for all i, j, t, s
   $VAR(\epsilon_{it}) = \sigma_\epsilon^2,\ COV(\epsilon_{it}, \epsilon_{is}) = 0,\ COV(\epsilon_{it}, \epsilon_{jt}) = 0$ for all i, j, t, s
   $VAR(u_i) = \sigma_u^2,\ COV(u_i, u_j) = 0$ for all i, j

   (d) Write the full covariance matrix for a data set with $N = 2$ and $T = 3$. Is the pooled OLS estimate of β consistent? Is it efficient? How would you estimate β consistently and efficiently? Provide an exact formula in terms of the parameters of the model and describe the estimation steps.

7. (10 points) Consider a regression of saving on income and a constant. Suppose income is measured with an error that satisfies the classical measurement error assumptions. The standard deviation of true level of income is 2.5 and the standard deviation of measurement error is 1.5. Determine the bias in the OLS estimate of the propensity to save from income. (express the OLS estimate of the coefficient of income as a function of the true effect.) Assume that OLS estimation of this regression produces a coefficient of income equal to 0.017. What is the true effect on saving of increasing one's income by one unit?

8. (10 points) Explain the two concepts and state the main difference between them.

   1. Pooled OLS estimator and fixed effects estimator
   2. Classical measurement error and optimal prediction error
   3. Under-specification and over-specification
   4. Exogenous and endogenous sample selection

9. (10 points) Consider the following regression:

$$ln(wage)_{it} = \beta_0 + \beta_1 D_2 + \beta_2 female_i + \beta_3 D_2 \cdot female_i + X_{it}\gamma + \alpha_i + e_{it}$$

where D_2 is a time dummy which indicates second period, and $female$ is a gender dummy which indicates whether the individual is female or not, X is experience in years, $wage$ is in dollars, α represents individual time invariant characteristics. Think of this model as a true one, and work within this simple world, meaning there is no omitted variables whatsoever. Additionally,

$$\mathbb{E}(e_{it}|RHSvariables, \alpha_i) = 0,\ COV(\alpha_i, RHSvariables) \neq 0$$

(a) Interpret all the coefficients of this model. What is the effect of change in *experience* on *wage*? What is the interpretation of coefficient β_3?

(b) If you run a regression without individual time invariant characteristics, will you obtain consistent and efficient estimates? Explain.

(c) Propose a model which allows you to estimate the coefficients consistently. Will all the coefficients be estimated consistently? Explain.