CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TAs: Jelena Plazonja, Gega Todua
Due Date: 15/07/2014

ECONOMETRICS II

HOMEWORK #5

Instructions. You can (but don't have to) work in groups of two. Make sure to report the commands used to generate the empirical output via email (do file needed). Please, add to your written solutions the STATA output obtained (sent or printed). The solutions are to be submitted by Tuesday July 15^{st}, at the beginning of Final Exam. In data exercise answer the questions using **50% subsample** of the original data. **NOTE:** *set the seed to be your birth date!* (e.g. 01121988).

Problem 1. You estimate a Tobit model: $\hat{y}^* = 20 + 40x_1 - 60x_2$, with $\sigma = 100$. At $x_1 = x_2 = 1$:

(a) Compute $\hat{y}^*$, $\mathbb{P}(y > 0)$, $\mathbb{E}(y|y > 0)$ and $\mathbb{E}(y)$.

(b) Compute $\frac{\partial \mathbb{E}(y|x)}{\partial x_2}$ and $\frac{\partial \mathbb{P}(y>0|x)}{\partial x_2}$.

(c) Recompute (a) and (b) at $x_1 = 1$ and $x_2 = 2$.

(d) Compare the effect of increasing x_2 from one to two on the $\mathbb{P}(y > 0)$, first treating x_2 as a continuous and then as a discrete variable.

Problem 2. Tobit and Selection Model.
Using RAND data on medical expenditure over a 12 month period used in chapter 16 of Cameron and Trivedi's *Microeconometrics*, and using a similar model specification, we wish to consider the following broad question: Which model is appropriate for modeling the expenditure data?

(a) Using the data summary of the expenditure variable, analyze the implications of the high proportion of zero expenditures observed. Is this a violation of the normality assumption? Is there a transformation of expenditure that would make the assumption of normality more appropriate?

(b) Two candidate models are considered, each with the same set of covariates. These covariates are $LC, IDP, LINC, FEMALE, EDUDEC, XAGE, BLACK$, $HLTHG$, $HLTHF$, $HLTHP$. The models are Tobit model, and the selection model. Explain how each one of these will be set up, the relationship

and connections among them, and how one might compare and choose among them. if you are likely to encounter any specific specification or estimation problems, state them and suggest how you might handle them. Pay attention to the choice of exclusion restrictions.

(c) Estimate in turn the Tobit model and the selection models. in the case of the selection model, use both the MLE and the two-step Heckman estimators. Discuss your reasons underlying the exclusion restriction required in the estimation of the selection problem. Is there evidence that the selection problem is a serious issue?

Note: For details on the selection models, it is quite useful to actually read chapter 16 from Cameron and Trivedi's *Microeconometrics*.

Problem 3. Consider a family saving function for the population of all families in the United States:

$$sav = \beta_0 + \beta_1 inc + \beta_2 hhsize + \beta_3 age + u$$

where *hhsize* is household size, *educ* is years of education of the household head, and *age* is age of the household head. Assume that $\mathbb{E}(u|inc, hhsize, educ, age) = 0$.

(i) Suppose that the sample includes only families whose head is over 25 years old. If we use OLS on such a sample, do we get unbiased estimators of the β_j? Explain.

(ii) Now suppose our sample includes only married couples without children. Can we estimate all of the parameters in the saving equation? Which ones can we estimate?

(iii) Suppose we exclude from our sample families that save more than \$25,000 per year. Does OLS produce consistent estimators of the β_j?

Problem 4. You have cross sectional data of N households with information on household debt, household annual income and an indicator that equals one if the household lives in a city and that equals zero otherwise. You are interested in the effect of income and city on the level of household debt. There is a significant proportion of households in your sample, which do not borrow at all (they have zero debt).

(a) How would you estimate the model by maximum likelihood? Write down the log-likelihood function and state which parameters of the model are identified.

(b) What is the marginal effect of income on the probability that household has a zero debt? How would you construct a 95% confidence interval for this effect?

(c) What is the effect of a unit increase in income on the expected value of the debt (conditional on explanatory variables) among indebted households?

(d) What is the effect of a unit increase in income on the expected value of the debt, conditional on explanatory variables, in the population? Describe and interpret the two components of this effect.

(e) What is the partial effect of moving from the countryside to the city on the expected value of the debt, conditional on explanatory variables, among indebted households?

(f) Is the OLS estimate of income on household debt *using only households with positive debt* consistent? What is the direction of the bias (if there is any)? Justify your answer.

(g) Is the OLS estimate of the effect of income on household debt *using all households in the sample* consistent? What is the direction of the bias (if there is any)?

Problem 5. Consider the following model

$$y_i = \alpha + \beta x_i + \epsilon_i$$
$$I_i^* = \delta + \gamma z_i + \eta_i$$

and y_i is observed only when $I^* > 0$. Assume that ϵ and η have bivariate normal distribution, with zero means, $VAR(\epsilon) = \sigma_\epsilon^2$, $VAR(\eta) = \sigma_\eta^2 = 1$, and with correlation ρ. Assume $\gamma < 0$, $\rho < 0$, and $COV(x,\ z) < 0$. What is the expected direction of bias in the OLS estimate of β when selection is ignored?