CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TA: Jelena Plazonja, Gega Todua
Date: 03/07/2014

ECONOMETRICS II

Exercise session #9
(Censored and Truncated Models)

Problem 1. Consider a censored data model with censoring from above

$$y_i^* = x_i\beta + \epsilon_i$$

where ϵ_i is $N(0,\, \sigma^2)$ and $x_i\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}$. The observed data is x_i and y_i where
$y_i = y_i^*$ if $y_i^* < \mu$
$y_i = \mu$ if $y_i^* \geq \mu$

(a) Write down the log likelihood function of the model.

(b) Which parameters are identified?

(c) Derive the formula for $\mathbb{E}(y_i^*|x_i)$ and $\mathbb{E}(y_i|x_i)$.

(d) Derive the effect of x_{ki} on $\mathbb{E}(y_i^*|x_i)$ and on $\mathbb{E}(y_i|x_i)$ respectively.

(e) Assume that you estimate an OLS regression of y_i on x_i. Are the estimates consistent? Justify your answer.

Problem 2. You are asked to model weekly ice-cream consumption as a function of average outside temperature, local advertisement exposure, and individual's age. A non-negligible part of the individuals in your sample have zero ice-cream consumption in a given week. You have a random sample of individuals drawn from the entire population. You observe the actual amount of ice-cream consumed in a given week (positive amount or zero) and all the RHS variables perfectly.

(a) Without making any distributional assumptions, write down the expected value of ice-cream consumption in the population represented by your sample.

(b) Propose a maximum likelihood estimation of the model. State any assumptions you make and write down the log-likelihood function.

(c) Derive the average marginal effect of increasing the advertisement exposure on the expected value of the observed ice-cream consumption in the population. Interpret the different components of this effect.

(d) You are asked to estimate the same econometric model as proposed above but with different (more limited data). For each case described below, propose an estimation method, based on maximum likelihood, which yields consistent estimates of the parameters of the model. For each case, discuss what is the main advantage of the proposed method compared to the simple OLS regression of the observed LHS variable that measures ice-cream consumption on a constant and the three RHS variables.

   i. You have a random sample drawn from individuals, who had a positive ice-cream consumption in a given week. You observe the actual amount of ice-cream consumed and all the RHS variables perfectly.

   ii. You have a random sample drawn from individuals who are between 13 and 25 year old. again, you observe the actual amount of ice-cream consumed and all the RHS variables perfectly.

   iii. You have a random sample of individuals drawn from the entire population. You have all information on all the RHS variables you need in your data but instead of observing the actual amount of ice-cream consumed, you only observe a binary indicator of whether an individual consumed any ice-cream in a given week or not.

Problem 3. 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 4. You are asked to model annual expenditure on skiing as a function of income and proximity to a ski resort using a random cross-sectional sample from the adult sample in Czech Republic. Proximity is measured by a dummy variable that equals one if an individual lives in a region that contains a ski resort, and zero otherwise. A non-negligible part of the individuals in your sample do not ski and have zero expenditure on skiing.

(a) With the data described above, propose a method of estimating the effect of income on ski-related expenditure. State any assumptions you make.

(b) Write down the log.likelihood function and state which parameters of the model are identified.

(c) What is the expected value of skiing expenditure among the skiers (those with positive skiing expenditure)? What is the expected value of skiing expenditure in the adult population?

(d) Derive the effect of a unit increase in income on the expected value of the ski-related expenditure in the adult population. Describe and interpret the two components of this effect.

(e) Derive the effect of living in a region with a ski resort on the expected value of the ski-related expenditure in the adult population.