CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TA: Jelena Plazonja, Gega Todua
Date: 23/06/2014

ECONOMETRICS II

Exercise session #8
(Multinomial Models)

Problem 1. (From ES 7)
You are interested in the effect of monthly income on monthly food expenditure in the following model:

$$\ln(foodexp_i) = \alpha + \beta \ln(income_i) + \epsilon_i$$

where ϵ_i is distributed $N(0, \sigma^2)$.
However, instead of observing the actual value of food expenditure, you only observe whether the given individual's food expenditure falls into one of the following brackets: $\langle 0, 5000]$, $\langle 5000, 15000]$, and more than 15000.

(a) Given the distributional assumption, how would you estimate the model? Write down the likelihood function and determine which parameters can be estimated.

(b) What is the effect of a 1% increase in income on the probability of having food expenditure above 15000 CZK?

Problem 2. (From ES 7)
You are asked to estimate the effect of preventive health expenditure on the number of days an individual is ill in a given year. However, instead of observing the actual number of days of illness, you only observe whether the number of days falls into one of the following categories: $[0, 7]$ days, $\langle 7, 14]$ days, and more than 14 days.
Assume that there are no or only few individuals who are not ill at all in a given year.

(a) Propose a model of the number of days of illness per year as a function of preventive health expenditure. State any assumptions you make and write down the log-likelihood function. Which parameters of the models are identified?

(b) Derive the effect of preventive health expenditure on the probability of being ill for more than two weeks. How would you summarize this effect for the whole sample?

Problem 3. Discrete choice model

(a) Consider a random utility probit model with three alternatives under random sampling.

$$u_{1i} = z_i\theta_1 + x_{1i}\beta_1 + \epsilon_{1i}$$
$$u_{2i} = z_i\theta_2 + x_{2i}\beta_2 + \epsilon_{2i}$$
$$u_{3i} = z_i\theta_3 + x_{3i}\beta_3 + \epsilon_{3i}$$

where $(\epsilon_{1i},\ \epsilon_{2i},\ \epsilon_{3i})'$ is a zero mean uncorrelated normal random vector with variances σ_1^2, σ_2^2, and σ_3^2. Specify the likelihood and discuss which parameters are identified.

(b) Contrast the interpretation of $\hat{\beta}$ in the Linear Probability Model as opposed to the Logit model. Comment on the imposed properties of the effect of x on y in each model.

(c) How do we obtain standard errors for $\frac{\partial \mathbb{E}[y_i=1|x_i]}{\partial x_i}$ in qualitative choice models?

(d) Discuss how the richness (number of parameters) of a model analyzing choice among ordered intervals of y (i.e. $I_j = 1$ iff $y \in \left\langle y_j^{min}, y_j^{max} \right]$) depends on whether we know the minimum and maximum values y_j^{min}, y_j^{max} for each $j - th$ interval.

(e) Briefly discuss the key differences between the Multinomial Logit and Multinomial Probit.

Problem 4. You have the contraceptive use data from the report of the Demographic and Health Survey conducted in El Salvador in 1985. The table shows 3165 currently married women classified by age, grouped in five-year intervals, and current use of contraception, classified as sterilization, other methods, and no method.

Age	**Steriliz.**	**Other M.**	**No M.**	**All**
15-19	3	61	232	296
20-24	80	137	400	617
25-59	216	131	301	648
30-34	268	76	203	547
35-39	197	50	188	435
40-44	150	24	164	338
45-49	91	10	183	284
All	1005	489	1671	3165

Consider contraceptive use the response variable, and age a predictor. Model and interpret this data with multinomial logit.

Problem 5. Handout!