CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TAs: Jelena Plazonja, Gega Todua
Due Date: 01/07/2014 , 1:30 p.m.

ECONOMETRICS II

HOMEWORK #4

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 1^{st}, 1:30 p.m. (at the beginning of exercise session or to mailbox of either of TAs). In all data exercises answer the questions using **90% subsample** of the original data. **NOTE:** *set the seed to be your birth date!* (e.g. 01121988).

Problem 1. For this exercise use the data in PENSION.DTA. Note that the ordered response model should be estimated for the following three types of a pension plan portfolio: a bond-portfolio, a mixed-portfolio, and a stock-portfolio, as given by the variable *pcttstck*, which indicates the prevalence of stocks in the portfolio.

(a) Consider the following explanatory variables *choice*, *age*, *educ*, $female$, *black*, *married*, $finc25$ - $finc101$, $wealth89$, and $prftshr$ and estimate a linear model for *pctstck*. Why there might arise a need for heteroskedasticity-robust standard errors?

(b) The sample contains separate observations for some husband-wife pairs. Compute standard errors of the estimates from the model in part (a) that account for the cluster correlation within family (these should also be heteroskedasticity-robust). Compare the standard errors to usual OLS errors and to heteroskedasticity-robust errors (and comment!).

(c) Estimate the model from part (a) by ordered probit. Estimate $\mathbb{E}(pctstck|\mathbf{x})$ for a single, nonblack female with 12 years of education who is 60 year old. Assume she has net worth in 1989 equal to $\$150,000$ and earns $\$45,000$ a year, and her plan is not profit sharing. Compare this with the estimate of $\mathbb{E}(pctstck|\mathbf{x})$ from the linear model.

(d) If you want to choose between the linear model and ordered probit based on how well each estimates $\mathbb{E}(y|\mathbf{x})$, how would you proceed?

(e) Estimate the model by ordered logit , compare with your previous findings, and answer question (c) based on your new estimates.

(f) Use the ordered probit and ordered logit model to predict for each individual one of the three possible outcomes of the dependent variable. Comapre the (unconditional) distribution of individuals across the predicted outcomes based on the two models with the actual distribution.

(g) Consider the ordered probit model from above. Based on the estimated coefficient of $prftshr$, what sign do you expect the effect of being a woman to have on the probability of having a bond-portfolio, a mixed-portfolio and a stock portfolio respectively? Compute the average partial effect of being a woman on these three probabilities.

Problem 2. For this exercise again use the data in PENSION.DTA.
Define a variable $invest = 0$ if $pctstck = 0$, $invest = 1$ if $pctstck = 50$, and $invest = 2$ if $pctstck = 100$.

(a) Estimate the ordered probit model from the previous problem but with *invest* as the dependent variable. What do you conclude?

(b) Are there any interesting quantities that would differ between using *pctstck* and *invest* as the dependent variables? Explain.

(c) Estimate the same model with ordered logit and compare your estimates to ones previously obtaine. Comment.

Problem 3. We are interested in the ordered probit model. Our data consists of 250 observations, of which the response are

y	0	1	2	3	4
n	50	40	45	80	35

Using the preceding data, obtain maximum likelihood estimates of the unknown parameters of the model.
Hint: Consider the probabilities as the unknown parameters.

Problem 4. For this exercise follow the fishing example (Cameron & Trivedi, chpt 15) and use the data in FISHING.DTA from previous homework.

(a) Estimate the conditional logit model.

(b) Comment on the statistical significance of parameter estimates.

(c) What is the effect of an increase in income on the various models of fishing?

Problem 5. Introductory Wooldridge (International Edition) 17 C8

Problem 6. Suppose that a random variable Y has the following probability distribution: $Pr(Y = 1) = p$, $Pr(Y = 2) = q$, $Pr(Y = 3) = 1 - p - q$. A random sample of size n is drawn from this distribution and the random variables are $Y_1, Y_2, \ldots, Y_n$. Derive the likelihood function for the parameters p and q and derive the ML estimates of p and q.

Problem 7. You have an iid sample with N observations on y_i, a variable which describes the duration of unemployment ($y_i > 0$ for all i). Assume that y_i is distributed by the following exponential distribution: $f(y|\gamma) = \gamma exp(-\gamma y)$. Derive ML estimator of γ.

Problem 8. Let y be a binary outcome variable, which takes on values 0 and 1, x is a continuous variable and d is a dummy variable.

(a) Find the effect of x and d on the probability $Prob(y_i = 1|x_i, d_i)$ in the following models:

- $Prob(y_i|x_i, d_i) = \Phi(\alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \delta d_i)$
- $Prob(y_i|x_i, d_i) = \Phi(\alpha + \beta_1 x_i + \beta_2 x_i^2 + \delta d_i + \gamma_1 d_i x_i + \delta_2 d_i x_i^2)$

(b) How would you estimate these effects?

(c) How would you obtain the standard errors of these effects?