ECON 861
Empirical MicroEconomics II
Winter Term 2012/13

Assignment 3
Due: In Class on ***March 5th*** *2013*

Rules of the 'game':

1) you can work in groups (actually, I encourage you to do so), but you MUST submit your own set of answers

2) you MUST list the people you worked with in the front page

3) late submissions will NOT be accepted. If you cannot make it to the class, you can pop into my office (DH306) and submit the assignment whenever you are ready, before the 5th

4) I will keep the set of answers you are going to hand in, so please make copies if you think you are going to need them

5) you have to dowload the datasets for the exercises from the webpage of the course:
http://qed.econ.queensu.ca/pub/faculty/cozzi/Webpage/Cozzi_ECON861.htm

Devote some time to give the graphs, plots and tables a format easy to understand. Also the way you present your answers matter for the final grade. Even if a question is mainly analytical, briefly explain what you are doing, stressing the economic meaning of the various steps.

Ex.1: Estimation of Linear Probability and Probit/Logit Models.

Download the first dataset: PS3_1.dta. It includes a set of variables on a sample of American males, some of whom were arrested in 1986.

a) Define a binary variable, say *arr86*, equal to unity if a man was arrested at least once during 1986, and zero otherwise. Estimate the following LPM:

$$arr86_i = \beta_0 + \beta_1 pcnv_i + \beta_2 avgsen_i + \beta_3 tottime_i + \beta_4 ptime86_i + \beta_5 inc86_i + \beta_6 black_i + \beta_7 hispan_i + \beta_8 born60_i + \varepsilon_i$$

Report the usual and heteroskedasticity-robust standard errors. What is the estimated effect on the probability of arrest if *pcnv* goes from .25 to .75? (*pcnv* is the proportion of times the individual was sent to prison when he was arrested in the past).

b) Test the joint significance of *avgsen* and *tottime* using a robust and a non-robust test. Compute the predicted values and comment.

c) Now estimate the model by Probit. At the average values of the other explanatory variables (setting $black = 1$, $hispan = 0$, and $born60 = 1$), what is the estimated effect on the probability of arrest if *pcnv* goes from .25 to .75? Compare this result with the answer from part a.

d) For the Probit model estimated in part c, obtain the percent correctly predicted. What is the percent correctly predicted when $narr86 = 0$? When $narr86 = 1$? What do you make of these findings?

e) In the Probit model, add the terms $pcnv_i^2$, $ptime86_i^2$, and $inc86_i^2$ to the model. Are these individually or jointly significant? Describe the estimated relationship between the probability of arrest and *pcnv*. In particular, at what point does the probability of conviction have a negative effect on the probability of arrest?

f) Estimate the same model of part c, this time with a Logit model. At the average values of the other explanatory variables (setting $black = 1$, $hispan = 0$, and $born60 = 1$), what is the estimated effect on the probability of arrest if *pcnv* goes from .25 to .75? Compare this result with the answer from part c.

g) Plot the predicted probabilities of the LPM and the Probit against each other (if they are close, they should lie on the 45-degree line). Do the same with the predicted probabilities of the Logit and the Probit.

Ex.2: Specifying a Simple Discrete Choice Model.

Download the second dataset: PS3_2.txt. It's a dataset on extramarital affairs used by Fair (JPE 1978). The sample contains a survey of 6,366 married women, conducted by *Redbook* magazine. The variables in the data set are as follows (if you need more details, please refer to the paper):

1: *id*= an identification number;
2: *c*= Constant, value = 1;
3: *v1*= self-reported rating of the marriage, coded 1 to 5;
4: *v2*= age, in years, aggregated;
5: *v3*= number of years married;
6: *v4*= number of children, top coded at 5;
7: *v5*= religiosity, coded 1 to 4, 1= not, 4= very;
8: *v6*= education, coded 9, 12, 14, 16, 17, 20;
9: This variable is not used
10: *v7*= occupation;
11: *v8*= husband's occupation;
12: *yrb*= a constructed measure of time spent in extramarital affairs;
13: This variable is not used;
14: This variable is not used.

This time the file is not in STATA format. You are going to learn the 'beauty' of dealing with 'quasi-raw' data.

1) You need to import the data in STATA. You can either use the software StatTransfer, or the infile/insheet commands in STATA.

2) Generate the binary variable *aff=1* if *yrb*>0, and 0 otherwise. The regressors of interest are *v1* to *v8*. Choose some of those to specify a binary choice model for *aff*. Start with a Probit, try different specifications, and provide some tests to support your choice.

3) Explain in 'economic terms' why the chosen specification is 'better'; try to think about the possible endogeneity problems that your model might suffer from (but do not try to address them at this stage). In particular, explain how the self-reported rating of the marriage (*v1*) might bias your estimates.

4) Compute the marginal effects for the variables you included.

5) Compare your results to those of a Logit, with the same regressors as before: are there any substantial differences between the two models?

Ex.3: Ordered Data.

Keep the same dataset you used for Ex.2. The dependent variable is now *v1* that is the rating of the marriage. It goes without saying that this is an ordered discrete variable.

1) Analyze *v1* using an ordered probit model of your choice.

2) Explain in 'economic terms' why you chose that specification, arguing with simple 'stories' what is the 'theoretical' reason that persuaded you to include that regressor.

3) Briefly comment on the marginal effects for this model.