Microeconometrics
Spring Semester 2019/20
Problem set 2
Due: March 26^{th}

Note: Send a copy of your solutions to Sónia Félix at sfelix@novasbe.pt and to me (teresa.molina@novasbe.pt). The solutions should include a .pdf or .doc document with the answers to each one of the questions ***and*** *the log files or do files of the exercises done using stata. Late problem sets will not be accepted.*

Exercise 1 (Discrete choice models)

For this exercise please create a do-file and corresponding log-file to hand in with your solution.

Use the dataset **laborparticipation.dta** that provides information on labor force participation of married women during 1975. To model labor market participation we consider the dependent variable $inlf$, which is equal to one if the woman reports working for a wage outside the home at some point during the year, and zero otherwise. The independent variables are the husband's income ($nwifeinc$, measured in thousands of dollars), years of education ($educ$), years of labor market experience ($exper$), age, number of children aged less than six years old ($kidslt6$), and number of kids between six and 18 years of age ($kidsge6$).

(a) Estimate a linear probability model (LPM) with heteroskedasticity-robust standard errors. Why might a linear probability model not be suitable for modelling the probability of labor force participation?

(b) Estimate the same model specification using the probit and the logit estimators. Interpret the results.

(c) Compare the LPM coefficients with the average marginal effects for the probit and the logit models.

(d) Estimate the probability of participation in the labor market for women with 0, 1, and 2 kids aged under 6 years using the probit model at the average value of the regressors. How do these probabilities compare with the probabilities estimated by the LPM?

(e) Given the results from part 1.b), set the regressors at the sample mean and draw a picture of the probability of participating in the labor market as a function of the number of kids.

Exercise 2 (Quantile regression models)

For this exercise please create a do-file and corresponding log-file to hand in with your solution.

Consider the **netfinancialwealth.dta** dataset, which is a subset of the data used in Abadie (2003).

(a) Provide detailed descriptive statistics of the variables net total financial wealth $nettfa$, income inc, age, and a binary variable indicating whether an individual is eligible for a (401 (k)) pension fund through her employer. Financial wealth and income are measured in thousands of dollars. Draw a histogram of the variable $nettfa$. What do you find?

(b) Run an OLS regression model with $netfa$ as dependent variable and inc, age, $agesq$, and $e401k$ as explanatory variables. Interpret the results. At what age does net financial assets increase with age?

(c) Test the presence of heteroskedasticity in the model. State clearly the null and alternative hypotheses, the test statistic, and the decision rule. What do you conclude?

(d) Obtain quantile regression estimates for the same model at the 0.1, 0.25, median, 0.75, and 0.9 quantiles. Interpret carefully the estimated coefficients on the inc variable and compare them with the OLS estimate.

(e) Run the same quantile regressions using only $e401k$ as explanatory variable. Why did the coefficients change from those estimated in (d)?

Exercise 3 (Multinomial discrete choice models)

For this exercise please create a do-file and corresponding log-file to hand in with your solution.

Consider the **travel_choice_MNL.dta** dataset on individual choice to travel between Sydney and Melbourne. The dataset is comprised of 210 observations on choice among four travel modes: air, train, bus, and car. The attributes used are choice-specific: gc, which is a measure of the generalized cost of the travel, $ttme$, which is the terminal time (zero for cars), and for the choice between air and the other modes $hinc$, the household income. The model specified is:

$$U_{ij} = \alpha_{air} d_{i,air} + \alpha_{train} d_{i,train} + \alpha_{bus} d_{i,bus} + \beta_g gc_{ij} + \beta_t ttme_{ij} + \gamma_h d_{i,air} hinc_i + \varepsilon_{ij}$$

(a) State the probabilities for a four-outcome conditional logit model.

(b) Provide sample means of the attributes for the 210 observations and for the observations that made that choice.

(c) Estimate a Conditional Logit Model (CLM). Interpret the coefficient on household income.

(d) Obtain the predicted probabilities of choice of each mode. Are they consistent with the actual frequencies?

(e) Calculate the marginal effects of a change in household income at mean values. Interpret the results.

(f) Reshape the data in wide format in order to end with a single line by individual. Estimate a multinomial logit model of *choicetravel* on the alternative-invariant variable. Comment on the estimated coefficients.

(g) Is the IIA likely to hold in this model? Perform a formal test.