**Homework – Microeconometrics (Part I), IMAE and IMQF studies**

**Professors: Aleksandra Nojkovic, Aleksandra Anic and Marko Vladisavljevic**

**Due date: Sunday, 9th February 2025**

**In-person defense date: 12th February 2025, room TBA**

**Total marks: 25**

**General note: Clearly state the hypothesis, test statistics, and rejection rule in every related homework requirement below**

**First part**

**Data:** [**http://ekonometrija.ekof.bg.ac.rs/imae\_micec.html**](http://ekonometrija.ekof.bg.ac.rs/imae_micec.html)

1. Suppose that an event occurs with probability p. In a sample of size *n,* the event occurs *m* times. Define the joint probability function and loglikelihood function of the event occurring.

(1 mark)

1. The following *Probit* estimates were calculated from a sample of 2380 mortgage applications filed in 1990 (HMDA data from FED of Boston, *Boston\_hmda\_data*):



The dependent variable *deny* takes the value 1 if the mortgage application is denied, *pi\_rat* is a ratio of total monthly debt payments to total monthly income, *ccred* is consumer credit score (goes from 1 to 6) and *married* if the applicant reported being married.

1. Test the null hypothesis that the coefficient of *pi\_rat* is zero. Explain briefly the common way of calculating and reporting marginal effect for continuous variable *pi\_rat*.

(2 marks)

1. Using Probit estimates, calculate the marginal effect of being *married* by using sample statistics:



(1 mark)

1. Test the null hypothesis that all the slope coefficients are jointly equal to zero, given following estimates:



(1 mark)

**Second part**

1. Write and explain the probability equation(s) for the multinomial logit model if there are only individual specific variables in the model. (1 mark)
2. Explain how to test independence of irrelevant alternatives assumption by Hausman test. (don’t need to derive) (1 mark)
3. Use the restaurant.dta dataset. Estimate alternative-specific conditional logit model by using all the explanatory variables available in the dataset. Hint: check which variables are alternative and individual specific and which variables are only individual specific. Show that average predicted probabilities by restaurant type are the same as frequences for chosen restaurants. (2 marks)
4. Choose between alternative-specific conditional logit model and multinomial logit model. (1 mark)
5. Estimate multinomial logit model. Hint: Only individual specific variables may be included. It is not necessary to reshape the database, there is simpler solution how to apply mlogit on the dataset. (2 marks)
6. Use tobacco.dta dataset. Estimate ordered probit model among smokers. Use parent, female and income as explanatory variables. Interpret estimates of parameters. (2 marks)
7. Add age variable into the model and test between restricted and unrestricted model by using LR test. (1 mark)

**Third part**

Use the following data set <http://www.stata-press.com/data/r15/mroz87> (type use and then copy paste the link into the do file). Keep your analyses in the do file and send the do file together with the homework.

1. Use descriptive statistics functions to analyse female working hours (variable *whrs75*)
   1. How many censored observations are there in the variable *whrs75*
   2. What is the mean and what is the median of the conditional working hours distribution (excluding zero working hours)?
   3. If we exclude women with zero working hours, is there an asymmetry in the distribution of the variable *whrs75*?

(1 mark)

1. Estimate a tobit model, where the dependent variable is female working hours (*whrs75*) and independent variables are
   1. wedyrs (Wife's educational attainment (in years))
   2. wexper (Actual years of wife's previous labor market experience)
   3. wifeage (Wife's age)
   4. kl6 (Number of children less than 6 years old in household)
   5. k618 (Number of children between ages 6 and 18 in household)

(1 mark)

1. Calculate the appropriate type of marginal effects interpret the coefficients for the variables *wedyrs*, *kl6* and *k618*.

(2 marks)

1. Estimate a two-part model with the same variables as in the previous tobit model. Use *lfp* – labor force participation variable as a dependent variable in the first part of the model

(1 mark)

1. Interpret the effects of the varialbes *wedyrs*, *kl6* and *k618* for both parts of the model. Are there variables that affect labour force participation but do not affect conditional working hours (working hours higher than zero)?

(2 marks)

1. Estimate a heckman model (ml estimation) with the same variables as in the previous twopart model. Use variable *wedyrs* in the selection equation, but do not use it in the main equation (working hours equation). Additionally, use *hfedyrs* variable (Husband's father's education level) in the selection equation. Use *lfp* – labor force participation variable as a dependent variable in the selection part of the model.

(1 mark)

1. Are the two equations independent? Compare the coefficient for working experience in the working hours equation with and without selection effect. Are there significant changes?

(2 marks)