The International College of Economics and Finance
Econometrics-2021-2022.
Home assignment 6. MLE, Binary Choice Models
To be submitted by December, 22, 23:55 (before Winter exam)
No changes to the deadline are possible, please do and upload your work beforehand!
Answer questions shortly and accurately, avoiding redundant information

1. [30 marks] Consider a linear probability model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \,;\; i = 1, 2, ..., n \,,$$

where $Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise and $E(u_i) = 0$.

1.1. [10 marks] □ What is the interpretation of this model and its coefficients?

□ What problems are connected with linear probability model? (*Answer exactly to the question, no mathematical details are expected here*)

1.2. [10 marks] The alternative approach to estimation of linear regression is based on the using of so called logit model

□ Outline the idea and general structure of the logit model (*no explanation of ML estimation is expected here and no mark will be given for this*).

□ What is the difference in the interpretation of its coefficients and evaluation of the marginal effects of the factors?

1.3. [10 marks] How to evaluate statistical quality of the binary choice (logit and probit) models?

□ How to evaluate whether the coefficients of the logit model are significant?

□ What are McFadden R^2 and *LR* statistic? How they can be used?

2. [20 marks] Consider simple linear regression

$$Y_i = \beta_1 + \beta_2 X_i + u_i \,;\; i = 1, 2, ..., n \,, \quad (1)$$

where $E(u_i) = 0$. (*Answer exactly to the question, no mathemetical details connected with normal distribution of disturbance term are expected here*)

2.1 [10 marks] □ Explain the concept of the likelihood function and the maximum likelihood estimators of the regression coefficients

□ Explain why the likelihood function is usually used in the form of loglikelihood function/

2.2. [10 marks] □ Describe briefly the properties of MLE estimators.

□ Under what additional conditions does estimating regression (1) by the maximum likelihood method become preferable to estimating by the OLS method? (*Answer exactly to the question, no details are expected here*)

In answering the practical questions, no general theoretical/mathematical explanations are acceptable, nor any information on the topic that is not directly related to the question.

3. [50 marks] Your data file **ha06_data_** contains data on students who studied econometrics course during one of the last years. **R –** rating in % for the second year, **A –** number of seminars attended before October exam (from 0 to 8, but some students attended 2-3 seminars per week), **H –** average grade for homework, **S –** dummy variable, equal to 0 for students majoring in Banking and 1 for other specialties. Binary variable **P** (passed) equal to 1, for students whose assessment exceeds a minimum of 25%, 0 for others.
A student tries to understand what factors contribute to success in an econometrics exam so that he can use them in his studies. His friend helps (and sometimes hinders) him. **Follow the logic of their dialogue, carry out all the calculations mentioned in the text, and comment on the results and opinions of the dialogue participants.**

3.1. [10 marks] The student starts by selecting the explanatory variables for his model. To do this, he decided to use a linear probability model of the relationship between **P** and the variables **H, A, S, M, R (*equation 1*) (M**=1 for male students). His friend, after looking at the results, advised him to exclude the variables **M** and **R**, and the student agreed with him.
□ Try to reconstruct the friend's arguments doing appropriate calculations.
□ You can also disagree with him by giving counter-arguments.

3.2. [10 marks] Next, the student, following his friend's advice, calculates a logit model with the dependent variable **P** and the independent variables **H, A, S (*equation 2*)**, and is surprised to find that variable A has a negative coefficient.
□ A friend advises him to remove variable **H** from the regression in the hope that this will make variable **A** significant **(*equation 3*).** Comment.
□ Suggest a meaningful explanation for the reasons why attendance **A** is insignificant in the student's regression (*equation 2*)**.**

3.3. [10 marks] The student himself (majoring in banking) has an average grade of 10 on his homework so far at the end of September and has only attended one seminar. His girlfriend (ME major) already has an average grade of H=80 and has attended all 4 seminars in September. The student wants to compare the probability of passing the exam for himself and for his girlfriend. Help him to do this.

3.4. [10 marks] The student wants to understand whether he should pay more attention to doing his homework (student characteristics in relation to the logit model under consideration are given in paragraph 3.2). Help him calculate the marginal effect of one extra point gained from homework based on (*Equation 2*)
□ by means of a direct comparison of the probabilities of passing the exam;
□ via partial derivatives;
□ compare the marginal effect obtained with the maximum effect.

3.5. [10 marks] Are all of the variables in Equation 2 significant in the whole? Is the group of variables directly related to student activity **H** and **A** significant in this equation?

If you have any questions please ask at Vladimir.tcherniak@gmail.com