SCHOOL OF ECONOMICS

AUCKLAND PARK KINGSWAY CAMPUS

SECOND SEMESTER SUPPLEMENTARY EXAMINATION (2018)

Module: **Econometrics 4B (EKN4815 / ECM8X02)**

Examiner: Dr Kwame Osei-Assibey

Exam Duration: 3 Hours

Marks: 100

Instructions

- **All sections and questions are compulsory.**

- **No examination paper is to be removed from the room**

Section A (Theory and Concepts)

Question One   **[11]**

You have a data set that has two years of individual data on wages, experience, education level, and score on a job psychometric tests.

i. Write out the model assuming you organized the data set as pooled cross-section. **(1 Mark)**

ii. Write out a new model that allows the intercept to change between the two years. **(2 Marks)**

iii. Write out another model that allows both the slope and the intercept to change. **(2 Marks)**

iv. Describe in detail how you will test if the intercept is different between the time periods. **(3 Marks)**

v. Describe in detail how you will test if the slope is different between the time periods.**(3 Marks)**

Question Two [6]

Consider the following regression model;

$$y_{it} = \beta_0 + \beta_1 x_{it} + \varepsilon_{it};$$

Where $\varepsilon_{it} = \alpha_i + \mu_{it}$.

You are asked to estimate β_1. In a short essay, compare and contrast the following strategies in estimating β_1 (i) First differencing (ii) Fixed Effects. Be sure to clearly state the assumptions about α_i as well as illustrating how data is transformed for both strategies.

Question Three [11]

a. Explain why the IV variance of β_1 from the IV estimation will always be greater than the OLS variance. Given you answer, state in what case the two variance will be exactly the same. **(3 Marks)**

b. Consider the following simple regression model
$y_i = \alpha + \beta x_i + u_i$: Where $Cov(x_i, u_i) = \rho \; and \; \rho > 0$. Suppose there exist a variable z such that $Cov(z_i, u_i) = 0$

   i.   State the probability limit ($plim$) for both the OLS and IV estimator of β_1 **(2 Marks)**

   ii.   Is the IV estimator consistent? **(2 Marks)**

   iii. What if $Cov(x_i, u_i) > Cov(z_i, u_i) > 0$? Is the IV estimator still consistent? Explain. **(2 Marks)**

   iv.   If not, can you tell whether the IV or the OLS estimator is more biased (asymptotically)? **(2 Marks)**

Question Four [12]

a. What is the difference between a probit and a logit model? On what basis can you choose which one to use, and why does it matter? **(5 Marks)**

b. This hypothetical question looks at how education, age and husband's income are related to a married woman labour supply :

   i.   Construct a latent variable for a probit model on the probability of employment, with education (in years), age, and age squared as explanatory variables. Give an interpretation of the latent variable. **(4 Marks)**

ii. What is the partial effect of education on the employment probability? How is this different from partial effect of the average **(3 Marks)**

Section B (Practice and Applications Using STATA)

Question Five [35]

This exercise uses the WAGE2 dataset to assess wage-education relationship. Consider a log-wage (lwage) regression with explanatory variables : Years of education (educ) ; experience (exper) ; years with current employer (tenure) ; dummy for marital statuts (married=1 if married, 0 otherwise) ; dummy for living in the South (South =1 if living in the South, 0 otherwise) ; dummy for living in the urban neighbourhood (SMSA=1 if living in urban neighbourhood, 0 otherwise) ; race dummy (black= 1 if respondent is black, 0 otherwise)

a. Write down the population regression model to be estimated. **(1 Mark)**

b. Estimate the model using OLS and present your estimated model with t-statistics in brackets right below the estimated coefficient. Perform tests to assess the overall fit of the estimated model and the statistical significance of the estimated coefficients. **(4 Marks)**

c. Why might you expect an endogeneity problem with this regression ? How would this affect our OLS estimation of the coefficient of educ ? **(3 Marks)**

d. Assume ability influences wages but is not observed and contributes to the endogeneity problem (omitted variable problem). Explain why you think a provided proxy like IQ in the dataset cannot be used as an instrument variable? **(2 Marks)**

e. From the texts and practise, three variables namely number of siblings, father and mother levels of education have been used as instruments for education in many experiments. Discuss the characteristics that qualifies each of them as a potential instrument. **(5 Marks)**

f. In the presence of three potential intruments, what is the best estimator to use ? Explain. **(3 Marks)**

g. Use the instruments to perform an IV regression. In your presentation ensure that you:

   i. Include analyses on the overall fit of the estimated model and the statistical significance of the estimated coefficients and compare results to those from the OLS estimation. **(5 Marks)**

   ii. Provide evidence on the relevance and strength of the instruments. (**5 Marks**)

   iii. Perform endogeneity tests to support your choice of estimation technique. **(4 Marks)**

    iv. Test to ascertain if there is an over-identification problem. **(3 Marks)**

Question Six [25]

This exercise uses the FERTIL2 dataset. The information comes from Botswana's 1988 Demographic and Health Survey. We are going to use some information contained in the dataset to analyse factors that influences womens' usage of birth control.

For our exercise, the dependent variable is the binary variable usemeth (usemeth=1 if the respondent has ever used a birth control method, 0 otherwise). The explanatory variables for this exercise are years of education (educ); age (both age and agesq); number of living children (children); having electricity (electric=1 if respondent has electricity, 0 otherwise) and husband's education (heduc).

    a. Write down the population regression model to be estimated? **(1 Mark)**
    b. Use STATA to perform an appropriate linear probability model (LPM) regression and present your estimated models with t-statistics in brackets right below the estimated coefficients. Perform tests to assess the overall fit of the estimated model and the statistical significance of the estimated coefficients. **(3 Mark)**
    c. From your estimated coefficients, explain the impact of the explanatory variables on the probability of a respondent's using birth control. Are these consistent with your expectations ? **(3 Marks)**
    d. Why would you advise the use of logit regression over the LPM? **(1 Mark)**
    e. Use the random utility model (RUM) to construct a latent variable for a logit model on the probability of using a birth control with reference to the explanatory variables. Give an interpretation of the latent variable. **(4 Marks)**
    f. Use STATA to perform a logit regression and present your estimated models with t-statistics in brackets below the estimated coefficients. Perform tests to assess the overall fit of the estimated model and the statistical significance of the estimated coefficients and compare results to those from the LPM estimation. **(4 Marks)**
    g. Compare the expected probability under both the LPM and the logistic model of using birth control for an 18-year old woman, with both husband and wife having 10 years of education, have electricity, and have 7 children who are alive. What do you observe? **(4 Marks)**
    h. Estimate the odds ratios for all explanatory variables. Present your estimated models with t-statistics in brackets below the estimated odds

ratios. Using the result for education as an example, explain what they represent. **(3 Marks)**

i. What is the value of percent correctly predicted? What does this value mean? **(2 Marks)**

END