



SCHOOL OF ECONOMICS
AUCKLAND PARK KINGSWAY CAMPUS
SECOND SEMESTER EXAMINATION (2017)

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

Examiner: Dr Kwame Osei-Assibey

Examination Duration: 3 Hours

Marks: 100

Instructions

- **All sections and questions are compulsory.**
- **No examination paper is to be removed from the exam room**

Section A (Theory and Concepts)

Question One

[11]

- a. Explain why panel data techniques such as first-differencing and time demeaned fixed effects cannot be applied to an independent cross-section analysis. **(2 Marks)**
- b. 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. Describe how you would estimate this model through the time-demeaned fixed effects estimation. **(3 Marks)**
 - iii. Will you get different results if you use either first differencing or fixed effects estimations? **(2 Marks)**
 - iv. List three reasons why you would prefer a panel data analysis over a cross-sectional or time series analysis? **(3 Marks)**

Question Two

[6]

Consider the following population regression model;

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + \alpha_i + u_{it}$$

Where u_{it} is the idiosyncratic error term; α_i is the unobserved heterogeneity component.

- a. The within group version of the fixed effects regression model involves subtracting the group mean relationship: $\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \beta_2 \bar{x}_{i2} + \cdots + \beta_k \bar{x}_{ik} + \alpha_i + \bar{u}_i$ from the population regression model above. Explain why the between effects model is in general inappropriate for estimating the parameters of a model using panel data. **(2 Marks)**
- b. You are presented with data and has been asked to construct the random effect estimator for the given model and the data. The software you are using can only perform OLS estimation but not the random effect estimation. Luckily for you, a friend whispered to you that $\sigma_\alpha^2 = E(\alpha_i^2)$ and $\sigma_u^2 = E(u_{it}^2)$. Use this information to construct a random-effects estimator for your data and model. **(2 Marks)**
- c. If $\sigma_\alpha^2 \rightarrow 0$ what value does the random effect transformation parameter tend to? Which estimator does it become? **(1 Marks)**
- d. If $\sigma_u^2 \rightarrow 0$ what value does the random effect transformation parameter tend to? Which estimator does it become? **(1 Marks)**

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 your 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 **(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]

Suppose you were hired by the City of Johannesburg to conduct a study of voter willingness to support a tax increase to fund new school spending.

Define

$$V_i = \begin{cases} 1 & \text{if voter } i \text{ supports the tax increase} \\ 0 & \text{if voter } i \text{ does not support the tax increase} \end{cases}$$

Thus V_i is a binary response variable. It is believed that voter support for tax increase depends on income

a.

- i. Present the linear probability model to be estimated. **(1 Mark)**
- ii. What is the interpretation of the estimated slope? **(1 Mark)**
- iii. Would you convincingly argue that your estimate is reliable? Explain. **(3 Marks)**

b.

- i. How would your answer change in part a (ii) if you instead use a probit or logit model ? **(1 Mark)**
- ii. Construct a latent variable for a probit model on the probability of a voter supporting the tax increase with income as the only explanatory variable. Give an interpretation of the latent variable. **(3 Marks)**
- iii. What is the partial effect of income on the probability of a voter supporting a tax increase? 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 ($\ln wage$) regression with explanatory variables being : Years of education ($educ$) ; experience ($exper$) ; years with current employer ($tenure$) ; dummy for marital status ($married=1$ 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 instruments, 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 the estimation technique. **(4 Marks)**
 - iv. Test to ascertain if there is an over-identification problem. **(3 Marks)**

Question Six

[25]

This exercise uses the affairs dataset. The data is collected from surveys filled out by readers for a magazine. One can use a number of characteristics contained in this data to predict the probability of a respondent having an affair.

For our exercise, the dependent variable is the binary variable affairs (=1 if the person had at least one affair, 0 otherwise). The explanatory variables for this exercise are religiosity (relig is a categorical variable from 1 to 5, where 5 is very religious); happiness in marriage (ratemarr is a categorical variable from 1 to

5, where 5 is very happy); gender of respondent (=1 if male, 0 otherwise); age of respondents (age); number of years married (yrsmarr); kids (=1 if respondents have kids, 0 otherwise) and years of education (educ).

- a. Use STATA to perform an appropriate linear probability model (LPM) 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. **(4 Marks)**
- b. From your estimated coefficients, explain the impact of religiosity, happiness in marriage and having kids on respondents' probability of having an affair. Are these consistent with your expectations ? **(4 Marks)**
- c. Use the random utility model (RUM) to construct a latent variable for a logit model on the probability of having an affair with reference to the explanatory variables. Give an interpretation of the latent variable. **(4 Marks)**
- d. Use STATA to perform a logit 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 and compare results to those from the LPM estimation. **(5 Marks)**
- e. Compare the expected probability under both the LPM and the logistic model of having an affair for a very religious (relig=5) 25-year old woman, without kids, who is very happily married (ratemarr=5) for a year and has 9 years of education. What do you observe? **(3 Marks)**
- f. Estimate the odds ratios for all explanatory variables. Present your estimated models with t-statistics in brackets right below the estimated odds ratios. Using the result for education as an example, explain what they represent. **(3 Marks)**
- g. What is the value of percent correctly predicted? What does this value mean? **(2 Marks)**

END