**UNIVERSITY OF DUBLIN**

**TRINITY COLLEGE**

**FACULTY OF ARTS, HUMANITIES AND SOCIAL SCIENCES**

**DEPARTMENT OF ECONOMICS**

**Senior Sophister Trinity Term 2013**

**BESS, TSM,**

**Visiting Students**

**Econometrics**

**DATE VENUE TIME**

**Professors O. Henry, C. Newman, A. Bénétrix and Mr M. Curran**

EXAM INSTRUCTIONS

Section A: Please answer Q1 **and** one of **either** Q2 **or** Q3

Section B: Please answer TWO from Q4, Q5, Q6, Q7

**Materials Permitted for this Examination**

Standard non-programmable calculator

**You may not start this examination until you are instructed to do so by the Invigilator.**

**Section A**

**Please answer Q1 and EITHER Q2 OR Q3**

**Question 1 (100 marks)**

1. Let  be a random sample from a population with a normal distribution with mean  and variance . Show that the sample mean is an unbiased estimator of the population mean. Explain what is meant by an efficient estimator and comment on the efficiency of the sample mean as an estimator of the population mean. Briefly describe two different approaches for estimating population parameters using a sample of data.

[25 marks]

1. Consider the following linear regression model:



Explain why the assumption that , where  is a constant, is required when estimating this model using Ordinary Least Squares estimation and provide an example of a real life application where such an assumption is unlikely to hold.

*[25 marks]*

1. Consider the following estimated regression equation (standard errors in parenthesis) for a sample of 1,000 manufacturing firms:



Where,  is the output of firm , and  is firm ’s capital inputs and  is firm ’s labour inputs. Interpret the coefficients on  and  and comment on their statistical significance. Why do you think a log transformation of the data was used when estimating this model?

*[25 marks]*

1. Write down a regression model of a dependent variable  on two independent variables,  and , that is linear in parameters. Explain what is meant by the ‘omitted variable problem’ in the context of this regression model. What are the consequences for Ordinary Least Squares estimation of the model?

*[25 marks]*

**Question 2 (100 marks)**

Consider the following population model:



1. Derive Ordinary Least Squares (OLS) estimators for  and  stating any assumptions required.

*[10 marks]*

1. Give an example of an econometric application that you might estimate using this model and explain how you would interpret the estimated parameters.

*[10 marks]*

1. Show that the OLS estimator, , is an unbiased estimator of  stating any assumptions required. Comment on what these assumptions mean in the context of the example you provided for part b).

*[20 marks]*

1. Derive an estimator for the variance of  and comment on each of its components. What additional assumptions are required in order to derive this estimator? Comment on what these assumptions mean in the context of the example you provided for part b).

*[20 marks]*

1. Derive a test of the null hypothesis, , against the two sided alternative clearly stating any distributional assumptions required in the construction of the test statistic.

*[20 marks]*

1. What are the consequences for OLS estimation if  and . Explain what this means in the context of the example you provided for part b).

*[20 marks]*

**Question 3 (100 marks)**

Consider the following equation to explain a household’s level of health expenditure in a given year () on its income level () and the average age of household members ():



Using data on 1,200 households the following equation was estimated using OLS (estimated standard errors are given in parenthesis):



  



a) Interpret the estimated parameters of this model in terms of economic and statistical significance. Explain how  is constructed and comment on the  value from this regression.

*[25 marks]*

b) The model is extended to include the square of the average age of household members () and the number of household members ().



The  obtained from estimating this model for the same 1,200 households using OLS is .

1. What signs might you expect on the parameters  and ? Explain your answers. Comment, in particular, on how you would expect the inclusion  to affect your interpretation of the impact of age on health expenditures. *[15 marks]*
2. Perform a test of the overall significance of this model. *[10 marks]*
3. Perform a test between this and the original regression model given in part (a). *[10 marks]*

c) How will the following affect the estimation and statistical testing of the parameters of the model given in part b):

(i) The number of household members is highly correlated with household income? *[10 marks]*

(ii) The average age of household members is highly correlated with the variance of the error term? *[10 marks]*

d) What assumptions are required to show that OLS estimation of the model given in part b) will yield unbiased estimates of the population parameters? Are these assumptions likely to hold? Explain your answer.

*[20 marks]*

**Section B**

**Please answer TWO from Q4, Q5, Q6, Q7**

Question 4 (100 Marks) – Identification & Simultaneous Equations Models.

Part (a): (50 Marks)

Imagine results for a fraction of students from EC3090 disappear. Suppose from a class of 120, 6 students have missing results. Let  denote result in percentage terms where students need at least 40 to pass and  denote whether we observe the result or not. You can assume that the only scores possible are . The distribution of marks for students we have data on is given in table 1.

Score Frequency

5 3

15 17

25 2

35 10

45 18

55 10

65 15

75 21

85 15

95 3

Table 1: Distribution of marks.

1. What bounds does the probability that a student taken at random from the class passes lie between? Show your calculations.
2. What bounds can we place on the average mark in the entire population? Show your calculations.
3. Finally, assume there is no missing data, i.e. we have the distribution of marks for all 120 students in the class. Let  be a dummy variable equal to one if a student is a member of a water-polo club and zero otherwise. Suppose upon discovering that  and  an examiner states the following:

"Data indicate that membership of a water-polo club increases the probability a student will score highly in EC3090. The effect of such a membership is to increase the probability of scoring at least  from 0.1 to 0.5."

Does this statement accurately describe the empirical findings? Explain.

Part (b): (50 Marks)

Consider the following macroeconomic model in structural form:



1. Which variables are endogenous and which variables are predetermined? What type of relationship does the tax equation describe? Explain your answers.
2. Check the identifiability of the tax equation. Show your workings. Based on your answer, motivate an appropriate estimation technique for the tax equation.

Question 5 (100 Marks) – Limited Dependent Variables & Instrument Variables.

Part (a): (50 Marks)

1. Consider the following regression where  is the hourly wage in Dollars,  is a binary variable equal to  if female and  if male and  is the years of education:

 (1)

Have we avoided the dummy variable trap? Explain. Estimating (1) by OLS, we get  . Interpret this coefficient. Now suppose estimating



yielded estimates , , ,  and . Draw a rough sketch of the return to education.

1. Describe any two limitations of the linear probability model and how to estimate alternatives that overcome some of these issues such as logit & probit models.
2. Describe count as an alternative to . Consider a migration study between European cities via a logit model on how certain circumstances and factors can lead to the migration of families, where  indicates migration. Table 2 shows some results. Calculate count . Show your calculations.

Predicted Predicted

y = 0 y = 1 Total

Actual y = 0 471 16 487

Actual y = 1 183 20 203

Total 654 36 690

Table 2: Migration Study

Part (b): (50 Marks)

Consider the following wage regression

 (2)

where the dependent variable is the log hourly wage,  denotes years of education,  denotes years of labour market experience and  denotes years with the current employer. Let us suppose that you have two excluded exogenous regressors (instruments)  and  where  is the years of mother's education and  is years of father's education. We wish to test if at least one of  and  are correlated with  in a regression of  on , ,  and . We get  and a p-value of approximately .

1. What might be the issue with including  in this model and why might  and  be good instrumental variables?
2. Describe the Hausman test for endogeneity of and interpret the F-value of 51.10 above.

Estimating equation (2) by 2SLS in Stata, we obtain the following results

log(wage) = .047 + .062educ + .043exper + .002tenure

(.399) (.032) (.012) (.006)

n = 428,  = .139

1. Interpret the coefficient on . Is  a cause for concern? Explain.

**Question 6 (100 Marks)**

1. Explain how one can deal with trending variables (explained or explanatory) in a linear regression model. How does the previous approach change if the trend is exponential, as opposed to linear? Discuss whether the inclusion of trending variables violates the Classical Linear Model assumptions. Explain what a spurious regression is and how one can fix this problem.

*[25 marks]*

1. Explain what a stationary process is. List and discuss the conditions that make a stochastic process  with finite second moment  to be covariance stationary.

*[15 Marks]*

1. Identify the following processes and discuss their properties: **(i)**  with  and  being independent and identically distributed (i.i.d), **(ii)**  with 

*[30 marks]*

1. The *efficient markets hypothesis* (EMH) states that information observable in the market prior to period  should not help to predict returns during period . One way of testing the EMH is to estimate the following model:

.

Suppose that after estimating this model we run the following regression

and find that  and , where the numbers in parenthesis are the coefficient’s standard deviation, and . What do these findings imply? Discuss.

*[30 marks]*

**Question 7 (100 Marks)**

1. Explain what serial correlation means and how you would test for its presence. How would you test for serial correlation in the context of an AR(1) process without strictly exogenous general regressors,(i.e. one or more  are correlated with .

*[25 marks]*

1. Discuss the properties of the Durbin-Watson statistic and critically assess its limitations.

*[30 marks]*

1. Discuss the properties of a random walk process. Explain how a random walk without drift differs from a random walk with drift and from a stable AR(1) process. In addition, discuss what the best predictor of  is when  is a random walk process without drift and when  is a random process with drift.

*[25 marks]*

1. It is well known that when a series is I(1) the usual central limit theorem that underlies the asymptotic standard normal distribution for the *t* statistics does not apply. This means that the t-statistic cannot be used to test for the presence of a unit root. Taking this into account, how would you test for the presence of a unit root?

*[20 marks]*

© UNIVERSITY OF DUBLIN 2013