

# EFIM20011 Econometrics 1

## Reassessment Assignment

### Instructions Related to the Coursework

- You should answer ALL questions.
- Use no more than 15 sides of A4. Assignments longer than this are unlikely to be sufficiently concise.
- Hand write and scan your answers.
- Justify your answers and show the basis of your calculations.
- Assignments must be submitted through the online submission point on Blackboard.
- Your assignment will be checked for plagiarism through TurnItIn.
- You should attach the coversheet (available on Blackboard) to your assignment before submission.
- The DEADLINE for submission is 13:00 21 August.

### Late Penalties

Assignments handed in after the deadline, without a pre-arranged extension will be subject to the following penalty:

- A fixed absolute penalty of 10 marks is applied for each working day work is submitted after the agreed submission deadline. A mark of zero is applied to work submitted five or more working days after the agreed deadline if this threshold is not already reached.

### Overall Mark

Assignments will be marked out of 100 marks with reference to the stated marking criteria which are located at the end of this document. Section A is worth 30 marks. Section B is worth 35 marks. Section C is worth 35 marks.

## Section A. [30 marks]

Answer all of the following 6 questions. Each question is worth **5 marks**.

1. Given the random sample  $\{X_i\}_{i=1}^N$ , with each  $X_i \sim \text{Bernoulli}(1/4)$ , derive the expectation and variance of  $X_i$ . Now derive the expectation and variance of the sample mean  $\bar{X} = (1/N) \sum_{i=1}^N X_i$ . Appealing to the Central Limit Theorem, explain how  $\bar{X}$  is approximately/asymptotically distributed in large samples. **[5 marks]**
2. Let  $w_i$  denote the wage, and let  $f_i$  denote a female dummy variable, i.e.,  $f_i = 1$  if individual  $i$  is female and zero otherwise. In a linear regression of  $w_i$  on  $f_i$  including an intercept parameter, assuming that  $f_i$  is exogenous, what is the interpretation of the coefficient on the female dummy variable  $f_i$ ? If  $f_i = 1$  for all  $i = 1, 2, \dots, N$ , i.e., all individuals in the sample are female, what are the implications for the ordinary least squares (OLS) estimators for the slope and intercept parameters? **[5 marks]**
3. Given the random sample  $\{y_i, x_{1i}\}_{i=1}^N$ , each observation satisfies

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i,$$

where  $y_i$  is the dependent variable,  $x_{1i}$  is the independent variable,  $u_i$  is the error term, and  $\beta_0$  and  $\beta_1$  are parameters. Suppose that you inadvertently omit the independent variable in your specification and regress the dependent variable on just a constant, i.e., each observation instead satisfies  $y_i = \delta_0 + v_i$ , where  $y_i$  is the dependent variable,  $v_i$  is the error term, and  $\delta_0$  is a parameter. If you estimate  $\delta_0$  using OLS, when is the OLS estimator for  $\delta_0$  an unbiased estimator for  $\beta_0$ ? **[5 marks]**

4. Under which conditions is measurement error problematic in terms of bias and inconsistency of the OLS estimator? Suggest an alternative estimation method that addresses the issue of measurement error. **[5 marks]**
5. You want to analyse the effect of a tax credit, i.e., the amount of money that taxpayers can subtract from taxes owed, on labour supply. You know that the tax amount that can be subtracted depends on the number of children that one has, and that in 2000 there have been large increases in this amount for those with children. If the following STATA output aims to use a differences-in-differences method for your analysis, explain the function of *each* of the five STATA command lines below. Discuss any errors in the code and propose solutions for said errors. (Assume that the common trends assumption holds; the variable *year* identifies the year, *kids* measures the number of children that one has, and *work* is a dummy variable equal to 1 if in the labour force and 0 otherwise.)

```
1      gen after=(year >= 2000)
2      gen treat=(kids >= 1)
3      gen treat_after=treat*after
4      /* work is the labour force participation rate */
5      reg work after treat_after
```

**[5 marks]**

6. Explain when one can use a sharp regression discontinuity design and give a concrete example (different from the ones covered in class). **[5 marks]**

## Section B. [35 marks]

Given the random sample  $\{y_i, (x_{1i}, x_{2i}, \dots, x_{Ki})\}_{i=1}^N$ , each observation satisfies

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_K x_{Ki} + u_i,$$

where  $y_i$  is the dependent variable,  $(x_{1i}, x_{2i}, \dots, x_{Ki})$  are the independent variables,  $u_i$  is the error term, and  $(\beta_1, \beta_2, \dots, \beta_K)$  are parameters.

1. Derive (but do not solve) the first order conditions for the ordinary least squares (OLS) estimator  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K)$ . **[4 marks]**

Suppose that  $K = 2$ , i.e., that each observation satisfies  $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ ; we will refer to this multiple linear regression model for the rest of this section.

2. Consider estimating the linear models given by  $y_i = \delta x_{2i} + v_i$  and  $x_{1i} = \eta x_{2i} + w_i$ , where  $v_i$  and  $w_i$  are error terms, and where  $\delta$  and  $\eta$  are parameters. Derive the OLS estimators for  $\delta$  and  $\eta$ . **[6 marks]**
3. Using the OLS estimators for  $\delta$  and  $\eta$  from Part 2, consider the predictions  $\hat{y}_i$  and  $\hat{x}_{1i}$ , and also the residuals  $\hat{v}_i$  and  $\hat{w}_i$ . Now suppose that  $\hat{v}_i = \theta \hat{w}_i + \varepsilon_i$ , where  $\varepsilon_i$  is an error term, and  $\theta$  is a parameter. What does the OLS estimator for  $\theta$  capture? **[3 marks]**

When  $K = 2$ , the OLS estimators for  $\beta_1$  and  $\beta_2$  are denoted by  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , with  $\hat{\beta}_1$  given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N x_{2i}^2 \sum_{i=1}^N x_{1i} y_i - \sum_{i=1}^N x_{1i} x_{2i} \sum_{i=1}^N x_{2i} y_i}{\sum_{i=1}^N x_{1i}^2 \sum_{i=1}^N x_{2i}^2 - \left( \sum_{i=1}^N x_{1i} x_{2i} \right)^2},$$

and similarly for  $\hat{\beta}_2$ . Suppose that  $E(u_i | x_{1i}, x_{2i}) = 0$  and  $\text{Var}(u_i | x_{1i}, x_{2i}) = \sigma^2$ .

4. Prove that  $\hat{\beta}_1$  is unbiased. **[7 marks]**
5. Suppose that  $x_{2i} = \lambda x_{1i}$  for some fixed constant  $\lambda > 0$ . What is the implication for the existence of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ ? **[4 marks]**
6. Suppose that  $x_{1i}$  and  $x_{2i}$  are dummy variables, i.e., they can only take on values of 0 and 1, and also that  $x_{2i} = 1 - x_{1i}$ . Show that  $\hat{\beta}_1 = \sum_{i=1}^N x_{1i} y_i / \sum_{i=1}^N x_{1i}^2$  and  $\hat{\beta}_2 = \sum_{i=1}^N x_{2i} y_i / \sum_{i=1}^N x_{2i}^2$ . Why does this occur? **[4 marks]**
7. When  $x_{1i}$  and  $x_{2i}$  are dummy variables with  $x_{2i} = 1 - x_{1i}$ , re-parametrize the model so that  $y_i = \gamma_0 + \gamma_1 x_{1i} + u_i$ , where  $\gamma_0 = \beta_2$  and  $\gamma_1 = \beta_1 - \beta_2$ , and derive the standard error of  $\hat{\gamma}_1$ , the OLS estimator for  $\gamma_1$ . **[7 marks]**

### Section C. [35 marks]

You are interested in estimating the effect of smoking during pregnancy on child birth weight. You have individual data on a sample of 1 million mothers from various regions within a country for a given year and you want to estimate the following model:

$$\ln(bwght) = \beta_0 + \beta_1 packs + \beta_2 age + u, \quad (1)$$

where  $\ln(bwght)$  is the natural logarithm of the child birth weight,  $packs$  is the weekly average number of packs of cigarettes smoked by the mother during pregnancy,  $age$  is the mother's age at the time of birth,  $u$  is an error term, and  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  are parameters. We will refer to this equation as model (1) for the rest of the section.

1. Assume that  $packs$  is exogenous. An ordinary least squares (OLS) regression of model (1) gives an estimate of  $\beta_1$  equal to  $-0.299$  with a standard error of  $0.450$ . How would you interpret this parameter estimate? **[5 marks]**
2. Discuss whether it is reasonable to assume that  $packs$  is exogenous. Show whether the OLS estimator of  $\beta_1$  is expected to be biased upwards or downwards. **[6 marks]**

Assume that the exogeneity assumption for  $packs$  is not holding. You therefore consider instrumenting  $packs$  with excise taxes on cigarettes (denoted by  $cigtax$ ), which vary across regions within a country based on local laws.

3. Present the two conditions that are required for  $cigtax$  to be a relevant and exogenous instrument for  $packs$  and discuss whether they are likely to hold. **[7 marks]**
4. Explain how you can test whether  $cigtax$  is a strong instrument. **[5 marks]**
5. Show how you can implement the two stage least squares estimation of model (1) using  $cigtax$  as an instrument for  $packs$ . **[7 marks]**
6. Discuss how your answer to part [5.] changes if the excise tax is set at country level instead. **[5 marks]**

## **Level 5 Marking and Assessment Criteria (Second Year)**

### 1st (70+)

- For essay-based subjects:
  - Excellent knowledge and understanding of the subject and understanding of theoretical methodological issues.
  - A coherent argument that is logically structured and supported by evidence.
  - Demonstrates a capacity for intellectual initiative/independent thought and an ability to engage with the material critically.
  - Use of appropriate material from a range of sources extending beyond the reading list.
  - High quality organisation and style of presentation (including referencing); minimal grammatical or spelling errors; written in a fluent and engaging style.
  - A very high level of skill in problem solving, which demonstrates powers of critical analysis (NB: where problem solving is an important key learning outcome).
- For mathematical subjects:
  - Perfect, or near-perfect answers to a high proportion of the parts of the questions attempted, and a firm grasp of the central issues covered.
  - Answers are presented fluently and logically.
  - Explanations, where required, show evidence of an excellent comprehension of the material.
  - Interpretations, where required, often display a strong critical appreciation of the material.
  - Excellent use of common standard mathematical notation and conventions.

### 2:1 (60–69)

- For essay-based subjects:
  - Very good knowledge and understanding of the subject and displays awareness of underlying theoretical and methodological issues.
  - A generally critical, analytical argument that is reasonably well structured and well-supported.
  - Some critical capacity to see the implications of the question, though not able to 'see beyond the question' enough to develop an independent approach.
  - Some critical knowledge of relevant literature; use of works beyond the prescribed reading list; demonstrating some ability to be selective in the range of material used and to synthesise rather than describe.
  - Well presented: no significant grammatical or spelling errors; written clearly and concisely; fairly consistent referencing and bibliographic formatting.
  - A very good ability to apply principles effectively in the solution of factual problems and to deal with problems in an orderly manner, with realism and discrimination (NB: where problem solving is an important key learning outcome).

- For mathematical subjects:
  - A very good knowledge of much of the important material, possibly excellent in places, but with a limited account of some significant topics, or with some omissions/misunderstandings.
  - There is a good fluency and logical structure to most of the answers.
  - Explanations, where required, show evidence of good comprehension of the material though there may be some limited understanding of some areas.
  - Interpretations, where required, show some evidence of a critical appreciation of the material.
  - Some good use of common standard mathematical notation and conventions.

## 2:2 (50–59)

- For essay-based subjects:
  - Good comprehension of the subject, though there may be some errors and/or gaps, and some awareness of underlying theoretical/methodological issues with little understanding of how they relate to the question.
  - Capacity for argument is limited with a tendency to assert/state opinion rather than argue on the basis of reason and evidence; structure may not be evident.
  - Tendency to be descriptive rather than critical, but some attempt at analysis.
  - Some attempt to go beyond or criticise the 'essential reading' for the unit; displaying limited capacity to discern between relevant and non-relevant material.
  - Adequately presented: writing style conveys meaning but is sometimes awkward; some significant grammatical and spelling errors; inconsistent referencing but generally accurate bibliography.
  - An efficient attempt at solving problems, but a tendency to overlook a number of points (NB: where problem solving is an important key learning outcome).
- For mathematical subjects:
  - A reasonably good knowledge of several important topics, possibly showing some good understanding in places, but with a limited account of some significant topics, or with some significant omissions/misunderstandings.
  - There is fluency and logical structure to some of the the answers.
  - Explanations, where required, show evidence of good comprehension of the material though with limited understanding in some areas.
  - Interpretations, where required, are generally standard but may in parts show some evidence of a critical appreciation of the material.
  - Limited use of common standard mathematical notation and conventions.

## 3rd (40–49)

- For essay-based subjects:

- Limited knowledge and understanding with significant errors and omissions and generally ignorant or confused awareness of key theoretical/methodological issues.
  - Largely misses the point of the question, asserts rather than argues a case; underdeveloped or chaotic structure; evidence mentioned but used inappropriately or incorrectly.
  - Very little attempt at analysis or synthesis, tending towards excessive description.
  - Limited, uncritical and generally confused account of a narrow range of sources.
  - Satisfactorily presented: but not always easy to follow; frequent grammatical and spelling errors; limited attempt at providing references (e.g. only referencing direct quotations) and containing bibliographic omissions.
  - Attempts to identify relevant areas for focusing problem solving but makes significant mistakes in solutions indicative of either a lack of discrimination or an understanding of a principle (NB: where problem solving is an important key learning outcome).
- For mathematical subjects:
    - A reasonable spread of relevant knowledge but showing a good grasp of only a minority of the material. Some questions may be answered well, others will have major omissions or misunderstandings. Some questions may not be attempted at all.
    - There may be some evidence of a logical structure to the answers in some areas.
    - Explanations, where required, are short and display a limited understanding of the material. Some explanations are not given.
    - Interpretations, where required, are poor and do not show critical appreciation of the material.
    - Very limited use of common standard mathematical notation and conventions.

#### Marginal Fail (35–39)

- For essay-based subjects:
  - Shows very limited understanding and knowledge of the subject and/or misses the point of the question.
  - Incoherent or illogical structure; evidence used inappropriately or incorrectly.
  - Unsatisfactory analytical skills.
  - Limited, uncritical and generally confused account of a very narrow range of sources.
  - Unsatisfactory presentation e.g. not always easy to follow; frequent grammatical and spelling errors and limited or no attempt at providing references and containing bibliographic omissions.

- Limited attempt to identify relevant areas for focusing problem solving but makes significant mistakes in solutions indicative of either a lack of discrimination or an understanding of a principle (NB: where problem solving is an important key learning outcome).
- For mathematical subjects:
  - Considerable deficiencies, or very partial attempts at questions, across large parts of the topics set, but with some relevant material at places.
  - There is little evidence of a logical structure to the answers.
  - Explanations, where required, are poor or missing.
  - Interpretations, where required, are weak or missing and show almost no critical appreciation of the material.
  - Limited or no use of common standard mathematical notation and conventions.

#### Outright Fail (0–34)

- For essay-based subjects
  - Shows little or no knowledge and understanding of the subject, no awareness of key theoretical/methodological issues and/or fails to address the question.
  - Unsuccessful or no attempt to construct an argument and an incoherent or illogical structure; evidence used inappropriately or incorrectly.
  - Very poor analytical skills.
  - Limited, uncritical and generally confused account of a very narrow range of sources.
  - Very poor quality of presentation and limited or no attempt at providing references and containing bibliographic omissions.
  - Overlooks most of the points in a problem (NB: where problem solving is an important key learning outcome).
- For mathematical subjects:
  - Substantial deficiencies, or no attempt, across large parts of the topics set, but with a little relevant material at places.
  - There is little or no logical structure to the answers.
  - Explanations, where required, are poor or missing.
  - Interpretations, where required, are missing or wrong and show no critical appreciation of the material.
  - Very limited or no use of common standard mathematical notation and conventions.