UNIVERSITY OF WARWICK

Summer Examinations 2015/16

Econometrics 2: Microeconometrics

Time Allowed: 1.5 Hours

Answer ALL SIX questions in Section A (72 marks) and ONE question in Section B (28 marks). Answer Section A questions in one booklet and Section B questions in a separate booklet.

Approved pocket calculators are allowed.

Read carefully the instructions on the answer book provided and make sure that the particulars required are entered on each answer book. If you answer more questions than are required and do not indicate which answers should be ignored, we will mark the requisite number of answers in the order in which they appear in the answer book(s): answers beyond that number will not be considered.

Section A: Answer ALL SIX questions

---

1. Imagine we want to estimate the impact of receiving some treatment (e.g. participating in a job training programme) on a continuous outcome (e.g. wages).

(a) Explain what is meant by "selection on observables" or the "conditional independence assumption" in this context. **(3 marks)**

(b) Explain briefly how matching works and discuss the situations in which matching might be more appropriate than OLS (without matching) for estimating this treatment effect. **(3 marks)**

(c) Which of the estimators covered on the course would be inconsistent if there was "selection on unobservables" or "unobserved heterogeneity"? **(2 marks)**

(d) Which of the estimators covered on the course would be consistent if the only source of unobserved heterogeneity was constant within units (e.g. did not vary within individuals over time)? **(2 marks)**

2. Consider a panel data model of the following form:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i + u_{it}, \qquad\qquad u_{it} \sim IID(0, \sigma_u^2),$$

where i indexes individuals ($i = 1, \cdots, N$), t indexes time periods ($t = 1, \cdots, T$) and the u_{it} are assumed to be independently and identically distributed over i and t, with mean zero and variance σ_u^2. $\boldsymbol{x}_{it}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of parameters.

(a) State the additional assumptions you would need to make in order to consistently and efficiently estimate the above model using the random effects (RE) estimator. **(2 marks)**

(b) Write down the RE estimator in a way that makes clear it is a weighted average of the within and between estimators. **(2 marks)**

(c) Define the weighting parameter and discuss the circumstances under which the random effects estimator would approach the within estimator or the pooled OLS estimator. **(8 marks)**

---

3. During the late 1980s and early 1990s, the government gave all secondary schools in England the option of becoming "grant maintained" (GM). GM schools were granted independence from the local authority, offering greater control over their land and buildings, and power to set their own admissions policies. To become grant maintained, schools had to secure a majority vote amongst parents of children at the school (i.e. more than 50% had to vote in favour). Assume that the vote is binding, i.e. that the school had to become grant maintained if parents voted in favour, and had to stay under local authority control if they voted against.

(a) How might you estimate the causal effect of a school becoming grant maintained on the exam results of its pupils in this scenario? Describe the intuition behind the method you would use. State and explain the assumptions required for this method to consistently estimate the treatment effect, and whether you believe these assumptions are likely to hold in this case. **(9 marks)**

The results of this evaluation suggest that becoming grant maintained significantly increases the average test scores of children in the school. On the basis of these results, the government decides to force all schools to become grant maintained.

(b) Discuss any concerns you might have about the results being used in this way. **(5 marks)**

(Continued overleaf)

4. Let $Y_1, Y_2, \cdots, Y_n$ be n independent and identically $N(\mu, 1)$ distributed random variables, and let $y_1, \ldots, y_n$ denote their realisations.

(a) Write down the log-likelihood function for observation i.**(1 mark)**
Hint: $Y \sim N(\mu, 1)$ if:

$$f_Y(y; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-\mu)^2\right).$$

(b) Briefly motivating each step, derive the Maximum Likelihood (ML) estimator of μ. **(4 marks)**

---

5. Suppose we are interested in studying married women's labour force participation. For this purpose, we observe a random sample of married women in the UK $\{(\boldsymbol{x}_i, y_i) \colon i = 1, \ldots, N\}$, where y_i takes value 1 if woman i is working and 0 otherwise, and $\boldsymbol{x}_i$ is a $1 \times K$ vector of explanatory variables including, among other things, an intercept, husband's income, age, experience, years of education, number of children of 4 years of age or younger, and number of children above age 4. Suppose

$$P(y_i = 1 \mid \boldsymbol{x}_i) = \Lambda(\boldsymbol{x}_i\boldsymbol{\beta}) \equiv \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i\boldsymbol{\beta}}}.$$

(a) Derive the log-likelihood function for individual i.**(1 mark)**

(b) Let x_j be one of the continuous variables in $\boldsymbol{x}$. Write both the "average partial effect" (APE) and the "partial effect at the average" (PEA) of x_j on y. If you were asked to report only one of them, which one would you choose in this empirical application? Briefly motivate your answer. **(4 marks)**

(c) Let x_K be a discrete variable in $\boldsymbol{x}$. Write both APE and PEA of x_K on y. If you were asked to report only one of them, which one would you choose in this empirical application? Briefly motivate your answer. **(4 marks)**

(d) Let $\boldsymbol{x}_i\boldsymbol{\beta} = \boldsymbol{w}_i\boldsymbol{\delta} + \boldsymbol{z}_i\boldsymbol{\gamma}$, where $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are two vectors of length H and Q, respectively; $H + Q = K$. Describe how would you test the null hypothesis $H_0 \colon \boldsymbol{\gamma} = \boldsymbol{0}$ against the alternative $H_1 \colon \boldsymbol{\gamma} \neq \boldsymbol{0}$ **(4 marks)**

(Question 5 continued overleaf)

Let $\hat{\boldsymbol{\beta}}_N$ be the ML estimate of $\boldsymbol{\beta}$ obtained by using the observations $\{(\boldsymbol{x}_i, y_i)\colon i = 1, \ldots, N\}$. For each i, you could compute the predictions as

$$\tilde{y}_i = \begin{cases} 1 & \text{if } \Lambda(\boldsymbol{x}_i \hat{\boldsymbol{\beta}}_N) \geq 1/2 \\ 0 & \text{otherwise} \end{cases}.$$

The *percent correctly predicted* is the percentage of times that $\tilde{y}_i = y_i$, $i = 1, \cdots, N$. Some have criticised this prediction rule for two reasons: firstly, because the threshold value 1/2 is arbitrary. Secondly, because it is possible to get high correctly predicted percentages even when the least likely outcome is poorly predicted.

(e) Show formally why the second concern may arise. **(7 marks)**

6. Some response variables in economics can come in the form of a *duration*, which is the time elapsed before or since a certain event occurs. A few examples include weeks unemployed, months spent on welfare, and days until next arrest after incarceration.

Let T^* denote the duration of some event, such as unemployment, measured in continuous time. Consider the following model:

$$T^* = \exp(\boldsymbol{x}\boldsymbol{\beta} + u)$$
$$u \mid \boldsymbol{x} \sim N(0, \sigma^2),$$

where $\boldsymbol{x}$ is a vector of observable explanatory variables, $\boldsymbol{\beta}$ is a vector of unknown parameters of interest, and u is the unobservable error term.

Suppose T^* is not fully observable; instead of observing T^*, we are only able to observe $T = \min(T^*, c)$, where $c > 0$ is a censoring constant.

(a) Briefly motivating each step, derive $P(T = c \mid \boldsymbol{x})$ in terms of the standard normal CDF $\Phi(\cdot)$, c, $\boldsymbol{x}$, $\boldsymbol{\beta}$ and σ. **(8 marks)**

(b) What happens to $P(T = c \mid \boldsymbol{x})$ when $c \to \infty$? Explain briefly the intuition behind this result. **(3 marks)**

Section B: Answer ONE question

7. (Inspired by Ashenfelter and Krueger 1994, and Bonjour et al., 2003). In this exercise, we are interested in estimating returns to education using a data set of female, genetically identical U.K. twins. Each twin was asked questions about, among other things, their earnings, their own schooling, and the schooling of the other twin.

Suppose the (log of) wage of twin i in family f is determined by

$$y_{if} = \alpha + \beta S_{if} + A_{if} + \varepsilon_{if}, \qquad (i = 1, 2;\ f = 1, \cdots, n),$$

where S_{if} is years of schooling, A_{if} is "ability", broadly defined as all other variables affecting wages outside those of schooling (intelligence, motivation, family background, access to educational funds, etc.), and ε_{if} is an independently and identically distributed (i.i.d.) error. $u_{if} = A_{if} + \varepsilon_{if}$ is unobserved. n is the number of twin-pairs, and $N = 2n$ is the total number of observations in the dataset. From now onwards, we will assume that, for all $i, j = 1, 2$,

$$cov(\varepsilon_{if}, S_{jf}) = 0 \tag{A1}$$
$$cov(\varepsilon_{if}, A_{jf}) = 0. \tag{A2}$$

It is possible to show (but you are not asked to do it) that, as $n \to \infty$,

$$\hat{\beta}_N^{(\text{OLS},1)} = \frac{\hat{cov}(S_{if}, y_{if})}{\hat{V}(S_{if})} \xrightarrow{p} \frac{cov(S_{if}, y_{if})}{V(S_{if})}.$$

(a) Show that, unless $cov(S_{if}, A_{if}) = 0$, $\hat{\beta}_N^{(\text{OLS},1)}$ is an inconsistent estimator of β. Would you expect $\hat{\beta}_N^{(\text{OLS},1)}$ to be larger or smaller than β? Explain. **(4 marks)**
Hint 1: Recall that an estimator $\hat{\beta}$ is said to be inconsistent for β when $\hat{\beta} \xrightarrow{p} c \neq \beta$.
Hint 2: Recall that, for any two random variables X and Y, and any two scalars a and b,

$$cov(Y, a + bX) = bcov(Y, X), \quad V(aY + bX) = a^2 V(Y) + b^2 V(X) + 2abcov(Y, X).$$

Given that the data set consists of identical twins, one could assume that $A_{1,f} = A_{2,f}$ for all $f = 1, \ldots, n$, and estimate β by using variables in differences instead of in levels:

$$\Delta y_f = \beta \Delta S_f + \Delta \varepsilon_f, \qquad (f = 1, \ldots, n),$$

(Question 7 continued overleaf)

where $\Delta S_f = S_{1,f} - S_{2,f}$ (Δy_f and $\Delta\varepsilon_f$ are defined similarly). It is possible to show (but you are not asked to do it) that, as $n \to \infty$,

$$\hat{\beta}_n^{(\text{OLS},2)} = \frac{\hat{cov}(\Delta S_f, \Delta y_f)}{\hat{V}(\Delta S_f)} \xrightarrow{p} \frac{cov(\Delta S_f, \Delta y_f)}{V(\Delta S_f)}.$$

(b) Explain briefly what $A_{1,f} = A_{2,f}$ means and why it is useful in this context. **(3 marks)**

(c) Do you think that $A_{1,f} = A_{2,f}$, $f = 1, \ldots, n$, is a reasonable assumption? Would you be able to provide at least one argument in favour, and one against it? **(3 marks)**

(d) Show that, if $A_{1,f} = A_{2,f}$, $f = 1, \ldots, n$, $\hat{\beta}_n^{(\text{OLS},2)}$ is a consistent estimator of β. **(2 marks)**

The variable "years of schooling" (S_{if}) is often observed with error. In other words, we would like to observe S_{if}, whereas we actually observe S'_{if}, where

$$S'_{if} = S_{if} + v'_{if}, \qquad (j = 1, 2;\ f = 1, \cdots, n),$$

and v'_{if} is an unobserved *measurement error*. From here onwards, we will assume that, for all $i, j = 1, 2$,

$$cov(v'_{if}, S_{jf}) = 0 \tag{B1}$$
$$cov(v'_{if}, A_{jf}) = 0 \tag{B2}$$
$$cov(v'_{if}, \varepsilon_{jf}) = 0 \tag{B3}$$

and

$$cov(v'_{1,f}, v'_{2,f}) = 0. \tag{B4}$$

The OLS estimator in a regression of y_{if} on a constant and S'_{if} is equal to

$$\hat{\beta}_N^{(\text{OLS},3)} = \frac{\hat{cov}(S'_{if}, y_{if})}{\hat{V}(S'_{if})} \xrightarrow{p} \frac{cov(S'_{if}, y_{if})}{V(S'_{if})}.$$

(e) Show that, unless $cov(S_{if}, A_{if}) = 0$ and $V(v'_{if}) = 0$, $\hat{\beta}_N^{(\text{OLS},3)}$ is an inconsistent estimator of β. What happens to $\hat{\beta}_N^{(\text{OLS},3)}$ when $V(v'_{if})$ becomes larger and larger? Provide some intuition for why this is the case. **(10 marks)**

Hint 3: You may want to substitute $S_{if} = S'_{if} - v'_{if}$ in the equation of y_{if}.

(f) Assume $V(v'_{if}) \neq 0$, i.e. that there is unobserved measurement error. How could you obtain a consistent estimator of β using the data at your disposal in this exercise? Explain the method you would use and how this helps to overcome the measurement error problem. **(6 marks)**

8. Some corner solution responses can take on two or more values with positive probability. For instance, when the response variable is a fraction or a percent, the corners are usually at zero and one, or zero and 100, respectively. Another example is when institutional constraints impose corners at other values. For instance, if workers are allowed to contribute at most 15% of their earnings to a tax-deferred pension plan, and y_i is the fraction of income contributed for worker i, the corners are at zero and 0.15.

Generally, let $a_1 < a_2$, and consider the following model:

$$y = \begin{cases} a_1 & \text{if } y^* \leq a_1 \\ y^* & \text{if } a_1 < y^* < a_2 \\ a_2 & \text{if } y^* \geq a_2 \end{cases}$$

$$y^* = \boldsymbol{x}\boldsymbol{\beta} + u, \qquad u \mid \boldsymbol{x} \sim N(0, \sigma^2).$$

This specification ensures that $P(y = a_1) > 0$ and $P(y = a_2) > 0$ but $P(y = a) = 0$ for $a_1 < a < a_2$. Therefore, this model is applicable only when we actually see pileups at the two endpoints and then a (roughly) continuous distribution in between.

(a) Briefly motivating each step, derive $P(y = a_1 \mid \boldsymbol{x})$ and $P(y = a_2 \mid \boldsymbol{x})$ in terms of the standard normal CDF $\Phi(\cdot)$, $\boldsymbol{x}$, $\boldsymbol{\beta}$ and σ. Also, derive the PDF of y given $\boldsymbol{x}$ for realisations of y in the interval (a_1, a_2). **(7 marks)**

(b) Briefly motivating each step, derive $E(y \mid \boldsymbol{x}, a_1 < y < a_2)$ and $E(y \mid \boldsymbol{x})$. **(14 marks)**
Hint: If $Z \sim N(0, 1)$, then for $c_1 < c_2$ we have that

$$E(Z \mid c_1 < Z < c_2) = \frac{\phi(c_1) - \phi(c_2)}{\Phi(c_2) - \Phi(c_1)}.$$

(c) Consider the following method for estimating $\boldsymbol{\beta}$. Using only the nonlimit observations, that is, observations for which $a_1 < y_i < a_2$, run the OLS regression of y_i on $\boldsymbol{x}_i$. Explain why this does not generally produce a consistent estimator of $\boldsymbol{\beta}$.
(3 marks)

(d) How would you estimate $E(y \mid \boldsymbol{x}, a_1 < y < a_2)$ and $E(y \mid \boldsymbol{x})$? **(4 marks)**

(End)