UNIVERSITY OF WARWICK

Summer Examinations 2017/18

Applied Econometrics

Time Allowed: 3 hours, plus 15 minutes reading time during which notes may be made (on the question paper) **BUT NO ANSWERS MAY BE BEGUN.**

Answer **ALL** questions in **SECTION A** (32 marks total) and **ANY FOUR** questions from **SECTION B** (17 marks each). Answer Section A questions in one booklet and Section B questions in a separate booklet.

Statistical Tables and a Formula Sheet are provided. 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 Questions

---

1. Y_i is the income of individual i. Y_{1i} and Y_{0i} represent the potential incomes of individual i if the individual had ($T_i = 1$) and hadn't ($T_i = 0$) entered an in-work training program, respectively. State whether the following statements are true or false:

(a) Y_{1i} is only observable if $T_i = 1$. **(1 mark)**

(b) $Y_{1i} - Y_{0i}$ represents the individual causal effect of the program and can be observed if individuals are randomised into and out of the program. **(1 mark)**

(c) $E[Y_1|T = 1] - E[Y_0|T = 0]$ represents the average causal effect. **(1 mark)**

(d) The average selection effect is $E[Y_0|T = 1] - E[Y_0|T = 0]$ and is zero in expectation if individuals were randomised into the program. **(1 mark)**

(Continued overleaf)

2. The following regression is estimated using OLS:

$$\hat{Y}_i = \underset{(0.02)}{0.03} + \underset{(0.01)}{0.04} X_i \tag{1}$$

We are interested in estimating the causal effect of X on Y.

(a) An important variable, T, has been omitted from the model. T is thought to be correlated with both X and Y. Explain how this is likely to be biasing the OLS estimator of the causal effect. **(2 marks)**

(b) Unfortunately, there is no data on T. A further regression is run:

$$\hat{X}_i = \underset{(0.01)}{0.02} + \underset{(0.01)}{0.05} Z_i \tag{2}$$

Explain under what conditions Z will be useful as an instrument for X. **(2 marks)**

---

3. The mean and variance of weekly wages are 112.01 and 150.21, respectively, for a random sample of 19 male nurses. The mean and variance of weekly wages are 103.99 and 136.98, respectively, for a random sample of 18 female nurses. Stating any assumptions you make, answer the following:

(a) Test for equality of variances in male and female wages. **(1 mark)**

(b) Test if mean wages of male and female nurses are equal against the two-sided alternative. **(2 marks)**

(c) Calculate and intepret the p-value of the test in part (b). **(1 mark)**

---

4. Let $X \sim N(20, 100)$, $Y \sim N(30, 25)$ and $W = 10 + 2X + Y$. Further, suppose $cov(X, Y) = 9.5$. Answer the following:

(a) What is the probability that $X \leq 10$? **(1 mark)**

(b) What is the probability that $Y \geq 30$? **(1 mark)**

(c) What is the expected value and variance of W. **(1 mark)**

(d) What is the probability that $45 \leq W \leq 81$? **(1 mark)**

---

5. 25 households randomly entered into a microcredit program. Their mean income was 100.21 with a standard deviation of 10.32. For an independent random sample of 22 households not entered into the microcredit program mean income was 90.21 with a standard deviation of 11.21.

(Continued overleaf)

(a) Suppose the true effect of the program on mean income is 5.3. Choosing a 10% significance level, what is the probability of rejecting a null of no effect against a two-sided alternative. **(2 marks)**

(b) On a diagram represent whether your calculated probability in (a) would increase/decrease or remain the same if the true impact of the program was a reduction of 2.5 and not 5.3. **(2 marks)**

---

6. In February 2015, State A in the US made it illegal to own a gun. In the nearby state, State B, gun ownership was legal. Tim suggests using a crime survey to estimate the impact of the law change on number of gun crimes. Let G_{pst} represent the number of gun crime arrests in police station p, state s at time t where $t = 0$ if the arrests took place between March 2014 and January 2015 and $t = 1$ if the crimes took place between March 2015 and Jan 2016.

(a) Without controlling for any other factors, specify a regression model that would allow you to test whether the law impacted the number of gun crimes. Interpret each of the coefficients in your regression model. **(2 marks)**

(b) In order for your regression to provide strong evidence of a causal effect of the law what assumption needs to hold? Suggest one way you could provide supporting evidence for the assumption. **(2 marks)**

---

7. The following OLS regression is estimated on a random sample of 1201 individuals from the UK population:

$$ln(wage_i) = \underset{(0.122)}{5.198} + \underset{(0.0073)}{0.0571}school_i + \underset{(0.0032)}{0.0195}exper_i + \underset{(0.0147)}{0.0870}ZIQ_i \qquad (3)$$

Where $ln(wage_i)$ is the natural log of wages; $school_i$ is years of school; $exper_i$ is years of experience and ZIQ_i is standardised IQ score (that is, mean 0 and variance 1).

(a) At the 10% significance level test the hypothesis that the wage return to an additional year of school is equal to half the wage return of a one standard deviation increase in IQ score. Note $cov(b_{school}, b_{ZIQ}) = -0.00005$. **(3 marks)**

(b) Calculate the p-value of the test in part (a) and illustrate on a diagram. **(1 mark)**

---

8. The following regression is estimated via OLS on a random sample of 254 individuals:

$$\hat{H}_i = \underset{(4.53)}{150.12} - \underset{(1.23)}{3.51} age_i + \underset{(0.1)}{0.03} age_i^2$$

Where H_i is happiness (between 0 and 100) and age_i and age_i^2 are the age and age squared of individual i. Standard errors are given in parentheses.

(a) Illustrate the regression with a diagram. **(2 marks)**

(b) Let $M_i = 1$ if an individual is male and $M_i = 0$ otherwise. Outline how would you test for a difference in the relationship between happiness and age between males and females. **(2 marks)**

(Continued overleaf)

Section B: Answer FOUR Questions
Please use a separate booklet

9. S_i are the observed end of second year test scores for individual i. Further, let S_{1i} and S_{0i} represent the potential test score of individual i if the individual had $(T = 1)$ and hadn't $(T = 0)$ attended support classes during the year, respectively. Answer the following questions:

(a) What does $S_{1i} - S_{0i}$ represent? **(2 marks)**

(b) What does $E[S_1|T = 1] - E[S_0|T = 0]$ represent? **(2 marks)**

(c) What does $E[S_1|T = 1] - E[S_0|T = 1]$ represent? **(2 marks)**

(d) When could you reasonably expect $E[S_1|T = 1] - E[S_0|T = 0] = E[S_1|T = 1] - E[S_0|T = 1]$? **(3 marks)**

(e) Suppose a random sample of 26 students was taken. 12 students had taken support classes and scored 72.11 on average with a standard deviation of 4.21. 14 students had not taken support classes and scored 68.11 on average with a standard deviation of 4.12. Test for a difference in mean test scores between those who took and did not go to support classes. State any assumptions you are making. **(4 marks)**

(f) Did the support classes increase the average test score marks? Briefly explain your answer. **(4 marks)**

---

10. The following regression is estimated using OLS on a random sample of 524 individuals:

$$\widehat{SPORT}_i = \underset{(4.32)}{10.51} + \underset{(2.31)}{7.6}\,AGE_i - \underset{(0.01)}{0.06}AGE_i^2 + \underset{(4.56)}{50.21}M_i + \underset{(10.32)}{60.32}UG_i \quad (4)$$

$SPORT_i$ is number of minutes of sport activities per week individual i does; AGE_i and AGE_i^2 are the individuals age and age squared; M_i is $= 1$ if male and $= 0$ otherwise; UG_i is $= 1$ if they have an undergraduate degree and $= 0$ otherwise. The $R^2 = 0.32$ and $RSS = 13100$.

(a) Excluding the intercept, interpret the coefficients in the model. **(3 marks)**

(b) Test the hypothesis that age impacts the amount of sport if the R^2 of the model estimated without the age variables is 0.25. **(3 marks)**

(Question 10 continued overleaf)

(c) Test the hypothesis that the impact of UG and male on sport is the same. Assume all the OLS estimators are pairwise uncorrelated. **(3 marks)**

(d) Estimating the model for 225 individuals who had an UG degree gave an $RSS = 11202$, while for those without an UG degree the $RSS = 9000$. Test for a structural break between individuals with and without a degree. **(4 marks)**

(e) Suppose the true effect of male on minutes of sport is 10. Calculate the power of rejecting a null hypothesis of no effect, at the 5% level, against the two sided alternative. **(4 marks)**

---

11. A university is trailing online courses and is interested in understanding teacher evaluations. $E_i \in [0, 100]$ represents the evaluation of teacher i where 0 means completely unsatisfied and 100 is completely satisfied. All course material is online: there are no lectures or classes and the only correspondence is via forums. The following regression was estimated on an independent random sample of 232 tutors.

$$E_i = \underset{(4.32)}{65.51} + \underset{(2.32)}{10.11}M_i + \underset{(3.32)}{6.21}EXP_i + \underset{(1.32)}{5.40}YR2_i - \underset{(3.32)}{10.51}YR3_i \tag{5}$$

$M_i = 1$ if the teacher is male and zero otherwise. $EXP_i = 1$ if the teacher has 2 or more years experience at university level teaching and zero otherwise; $YR2_i = 1$ for year 2 modules and zero otherwise; $YR3_i = 1$ for year 3 modules and zero otherwise.

(a) Interpret the intercept in the above model. **(2 marks)**

(b) The courses are three years in length, why is there no dummy variable for year 1 modules? **(2 marks)**

(c) Draw a graphic to represent the relationship between evaluation and year of study for males with less than two years of experience. Now draw the same graphic for females with less than two years experience. **(3 marks)**

(d) Outline how you would use an F-test to test $H_0 : \beta_{YR3} = -3\beta_{YR2}$. **(3 marks)**

(e) The university was surprised to see the positive and significant sign on the male coefficient. The regression was re-estimated with a new variable: $Q_i = 1$ if the module was quantitative and zero otherwise. The coefficients on Q and M in the new regression are 5.12 and 5.06, respectively. What does this imply about the relationship between male and quantitative models? **(3 marks)**

The coefficient on male was still statistically significant. To get experimental variation in gender in the next year of the course it was decided not to reveal the real name of the teacher. Instead the name of the teacher was randomised between male and female. Two models were specified:

$$E_i = \alpha + \beta_1 M_i + \epsilon_i \tag{6}$$

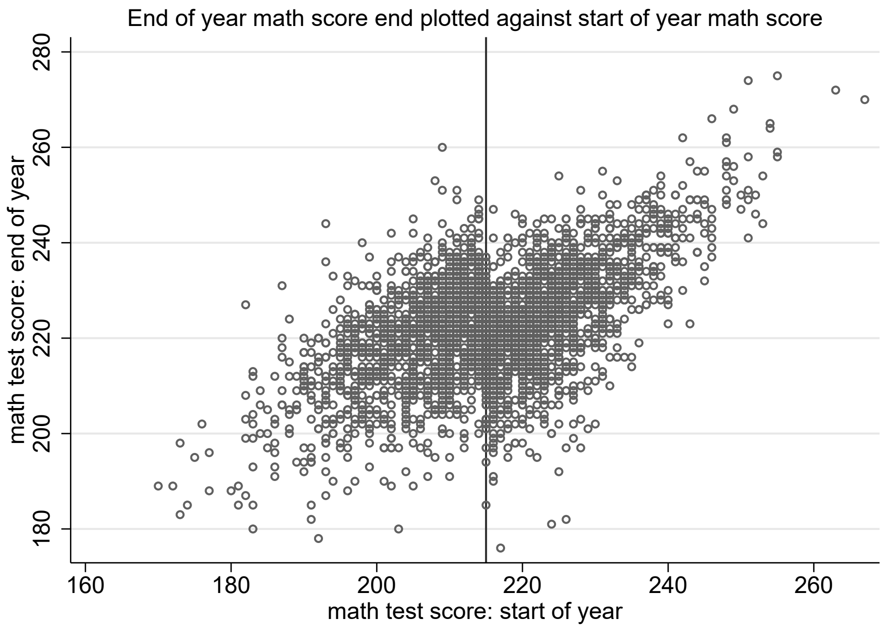
$$E_i = \alpha + \beta_1 M_i + \beta_2 EXP_i + \beta_3 YR2_i + \beta_4 YR3_i + \epsilon_i \tag{7}$$

(f) If the randomisation worked would you expect the estimator of β_1 in (7) to be greater than, less than, or equal to the estimator in model (6)? Briefly explain. **(2 marks)**

(g) If the randomisation worked would you expect the standard error of the estimator of β_1 in (7) to be greater than, less than, or equal to the standard error of the estimator in model (6)? Briefly explain. **(2 marks)**

(Continued overleaf)

12. The scatter plot below plots the end of year math score against the start of year test score for a random sample of 15 year olds. If students scored 215 and below in the start of year test (the grey line) they had to attend extra tuition classes. Those who score above 215 didn't attend extra tuition classes.



(a) Mean end of year test score for the 1295 students who took extra classes was 218.72 with a standard deviation of 11.38. Mean end of year test score for the 1472 students who didn't take extra classes was 225.91 with a standard deviation of 12.37. The null of no mean difference in test scores is very strongly rejected. Does this very statistically significant difference represent the causal effect of extra tuition on test score? Briefly explain. **(4 marks)**

(b) You think the above set-up is ideal for an RDD. Specify a regression model that would allow you to test the causal impact of extra tuition on end of year math test score. **(4 marks)**

(c) Interpret all the coefficients in your model and outline the hypothesis test you would need to run. **(4 marks)**

(d) Ali suggests checking the balance of genders, age, ethnicity, and free school lunch status of students close to the cut off point. Briefly explain why Ali is making this suggestion. Further explain what you would expect to find. **(5 marks)**

13. How does class attendance impact final year 2 score? To try and answer this question the following regression was estimated:

$$\hat{Score_i} = \underset{(3.41)}{75.32} + \underset{(0.34)}{1.21} Att_i + \underset{(4.34)}{10.12} M_i + \underset{(0.21)}{1.12} yr1_i \tag{8}$$

Where $Score_i$ is final year 2 score; Att_i is number of tutorial attended; $M_i = 1$ if male and zero otherwise; $yr1_i$ is the final score in year 1. The sample size was 412 and the $R^2 = 0.35$.

(a) Test the overall significance of the model. **(3 marks)**

(b) Interpret the coefficients. **(3 marks)**

(c) If you removed year 1 score from the regression how do you think it would change the coefficient on attendance? Briefly explain. **(3 marks)**

(d) Do you think the estimated coefficient on attendance is likely to represent the causal effect of attendance on test score? Briefly explain. **(3 marks)**

(e) A first stage regression was estimated:

$$\hat{Att} = \underset{(3.41)}{75.32} - \underset{(1.34)}{10.13} Early_i - \underset{(0.98)}{10.13} Late_i + \underset{(4.34)}{10.12} M_i + \underset{(4.34)}{10.12} yr1_i \tag{9}$$

where $Early_i = 1$ if the class is early in the day and zero otherwise; $Late_i = 1$ if class is late in the day and zero otherwise. Is time of day the class is run an appropriate IV for attendance? **(5 marks)**

---

14. You are interested in the causal effect of X on Y for a population of $i = 1, ..., N$ individuals in $t = 1, ..., T$ time periods. Consider the following equations:

$$Y_{it} = \gamma_0 + \gamma_1 X_{it} + \gamma_2 Z_{1it} + \mu_{1i} + \epsilon_{1it} \tag{10}$$

$$X_{it} = \pi_0 + \pi_1 Z_{1it} + \pi_2 Z_{2it} + \mu_{2i} + \epsilon_{2it} \tag{11}$$

In all parts of the question below, assume that all variables Y_{it}, X_{it}, Z_{1it} and Z_{2it} are continuous variables which have been standardised (mean 0 and variance 1) and, further, assume Z_{1it} and Z_{2it} are exogenous.

(a) If $E[\mu_{1i} + \epsilon_{1it}|X_{it}] = 0$ how would you estimate the relationship? Briefly explain. **(4 marks)**

(b) If $E[\mu_{1i}|X_{it}] \neq 0$ but $E[\epsilon_{1it}|X_{it}] = 0$ how would you estimate the relationship? Briefly explain. **(4 marks)**

(c) If $E[\mu_{1i}|X_{it}] \neq 0$ and $E[\epsilon_{1it}|X_{it}] \neq 0$ how would you estimate the relationship? Briefly explain. **(5 marks)**

(Question 14 continued overleaf)

(d) If $E[\mu_{1i} + \epsilon_{1it}|X_{it}] = 0$ and an interaction term $(X_{it} * Z_{1it})$ was added to the equation (9). How would your interpretation of γ_1 and γ_2 change? **(4 marks)**

(End)