THE AUSTRALIAN NATIONAL UNIVERSITY

Second Semester Final Examination – November, 2014

Econometrics II: Econometric Modelling

(EMET 2008/6008)

Reading Time: 5 Minutes
Writing Time: 90 Minutes
Permitted Materials: None

Instructions:

- This handout of exam questions contains 5 pages (including cover page) with 5 exam questions. Make sure you are not missing any pages!
- Answer **ALL** questions of this handout in the script book provided to you.
- Always provide comprehensive and exhaustive answers. Show your work!
- No partial credit will be given for merely stating results (unless I explicitly ask you to 'state' a result).
- Cheat sheets are not permitted.
- Total marks: 100.
- Good luck!

1. [20 marks]

   Consider the following linear model:

   $$Y_i = \beta_1 X_i + \beta_2 X_i^2 + u_i,$$

   with $E[u_i] = 0$ and $E[X_i^3] = 0$ and $E[X_i u_i] = 0$.

   (a) Suppose an oracle told you that $\beta_1 = \beta_2$. What would be the optimal way of estimating the coefficient(s) of the model?

   (b) For the general case that $\beta_1 \neq \beta_2$, mathematically define and derive a consistent OLS estimator for β_1. Exploit the fact that $E[X_i^3] = 0$.

2. [15 marks]

   Are the following statements true or false? Provide a short explanation.
   (Note: you will not receive any credit without providing a correct explanation.)

   (a) In panel data estimation, when $T = 2$, estimating coefficients by OLS in a model of first differences (change in time) will give the same results as estimating coefficients by OLS in a model with $(n-1)$-binary regressors.

   (b) The probit estimator is computationally simpler than the logit estimator.

   (c) The bias of an estimator will become smaller as the sample size increases (although it may not disappear completely).

3. [20 marks]

   Are girls who attend girls' high school better in math than girls who attend coed schools? Having available a random sample of high school girls from Australia, you are considering the following estimation equation:

   $$\text{Score}_i = \beta_0 + \beta_1 \text{GirlsHS}_i + \text{other factors} + u_i,$$

   where Score_i is the score on a standardized math test, GirlsHS_i is a dummy variable equal to 1 if a student attends a girls' high school. (Note: 'other factors' is shortcut for regressors on race, family income and parental education.)

   (a) Why might GirlsHS_i be correlated with u_i?

   (b) If you estimated the above model by OLS would you expect your estimate for β_1 to be an overestimate or an underestimate of the causal effect of girls' high school?

   (c) Let NumGirlsHS_i be a variable that measures the number of girls' high schools within a 30km radius of a girl's home. Discuss the two requirements needed for NumGirlsHS_i to be a valid instrumental variable. Which of these two requirements can be tested?

   (d) Is NumGirlsHS_i a convincing instrumental variable?

4. [20 marks]

Consider the research question

Do alcohol taxes reduce traffic deaths?

To answer this question, you have available a panel data set from 48 states in the United States from 1982 to 1988 including the following variables:

Variable name	Variable description
fatality rate	number of traffic deaths per 10,000
beer tax	tax on case of beer, in 1988 dollars
drinking age 18	dummy equal 1 if legal drinking age is 18
drinking age 19	dummy equal 1 if legal drinking age is 19
drinking age 20	dummy equal 1 if legal drinking age is 20
drinking age	legal drinking age
mandatory jail/ community service	dummy equal 1 if mandatory jail or community service for first time offenders
avg vehicle miles per driver	average miles per driver
unemployment rate	unemployment rate
real income per capita	real income per capita

Make use of the Table on the next page to answer the following questions. Provide short yet comprehensive answers.

(a) Column (1) of the table presents pooled OLS estimates from the regression of fatality rate on the beer tax. Interpret the coefficient.

(b) Study column (2). What has been changed in the estimation compared to column (1)? How does it affect the beer tax coefficient estimate? Explain.

(c) Study column (3). What has been changed in the estimation compared to column (2)? How does it affect the beer tax coefficient estimate? Explain.

(d) Study column (4). What has been changed in the estimation compared to column (3)? How does it affect the beer tax coefficient estimate? Explain.

(e) Panel data are helpful in addressing certain types of unobserved heterogeneity. For this traffic fatality application, give specific examples for the following types of unobserved heterogeneity:

(i) unobserved heterogeneity that varies across entities but not across time

(ii) unobserved heterogeneity that varies across time but not across entities

(iii) unobserved heterogeneity that varies across entities and across time

Which of these three can be addressed by using panel data?

(f) Briefly summarize what can be learned from the estimations undertaken in columns (5), (6) and (7).

TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: traffic fatality rate (deaths per 10,000).

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**							
Time effects = 0			4.22 (0.002)	10.12 (< 0.001)	3.48 (0.006)	10.28 (< 0.001)	37.49 (< 0.001)
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 (< 0.001)		31.96 (< 0.001)	25.20 (< 0.001)
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and p-values are given in parentheses under the F-statistics. The individual coefficient is statistically significant at the $^{+}$10%, *5%, or **1% significance level.

5. [25 marks]

Consider the simple linear model without constant term

$$Y_i = \beta_1 X_i + u_i, \tag{1}$$

in which X_i and u_i are CORRELATED with each other, that is, there is endogeneity in this model; mathematically $E[u_i X_i] \neq 0$. Note that $\beta_1 \neq 0$.

You have available a variable Z_i which partially explains X_i:

$$X_i = \pi Z_i + v_i, \tag{2}$$

The variable Z_i is uncorrelated with the error term u_i from equation (1).

(a) Suppose you run an OLS regression of Y_i on Z_i. What additional assumption(s) do you need to make for this regression to yield consistent estimates? Which coefficient does this regression estimate?

Now, consider the linear model

$$Y_i = \beta_1 \hat{X}_i + r_i, \tag{3}$$

in which $\hat{X}_i := \hat{\pi} Z_i$ where $\hat{\pi}$ is the OLS estimate of π in equation (2).

(b) What does r_i need to be equal to in order for equations (1) and (3) to be equivalent to each other?

(c) Show that $\hat{X}_i$ in equation (3) is exogenous, that is, prove that $E[\hat{X}_i r_i] = 0$.