BEE2031

UNIVERSITY OF EXETER

BUSINESS SCHOOL

January 2017

Econometrics

Module Convenor: Dr Xiaohui Zhang

Duration: ONE and a HALF HOURS

Answer ALL Questions

Answer all questions in the answer booklet.

Materials to be supplied:
None

Materials to be supplied on request:
None

Approved calculators are permitted.

This is a restricted note paper.
An original, hand-written formula sheet, consisting of at most two sides of A4 paper is permitted. No other material distributed in any form is allowed.

Section I Analysis on Individual Wage with Cross-Sectional Data (60 marks)

You have a cross-sectional dataset for 935 working males. Table 1 provides information on variables, and their definitions. Please answer the following questions.

Table 1

Variable	Definition
wage	monthly earnings
hours	average weekly hours
IQ	IQ score
KWW	knowledge of world work score
educ	years of education
exper	years of work experience
tenure	years with current employer
age	age in years
married	=1 if married
black	=1 if black
south	=1 if live in south
urban	=1 if live in urban areas
sibs	number of siblings
brthord	birth order
meduc	mother's education
feduc	father's education
lwage	natural log of wage

(A) Consider a wage equation that recognizes that ability (*abil*) affects *lwage*, as

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + Z\gamma + u \qquad \textbf{(a)}$$

where Z represents all other observed variables, including *tenure*, *married*, *black*, *south* and *urban*. As you can see, ability is something we cannot observe. Suppose your primary interest is to get consistent estimation for β_1, β_2 and γ . One possibility is to use a proxy variable to take the place of *abil*, i.e. the plug-in solution.

A.1 Suppose we can write $abil = \delta_0 + \delta_1 IQ + v$, what do we require of *IQ*, so that it can be a proper proxy variable for *abil*? What assumptions are needed for the plug-in solution to provide consistent estimation for β_1 , β_2 and γ ? (5 marks)

A.2 Using the above data, two models are estimated by OLS and results are provided in the following table (Table 2). Why is return to education estimated differently from the two models? Which one is more reliable? (10 marks)

Table 2

	Model 1		**Model 2**	
	Coef.	Std. Err.	Coef.	Std. Err.
educ	0.065	0.006	0.054	0.007
exper	0.014	0.003	0.014	0.003
tenure	0.012	0.002	0.011	0.002
married	0.199	0.039	0.200	0.039
black	-0.188	0.038	-0.143	0.039
south	-0.091	0.026	-0.080	0.026
urban	0.184	0.027	0.182	0.027
IQ			0.004	0.001
_cons	5.395	0.113	5.176	0.128
Observations	935		935	
R-squared	0.253		0.263	

(B) Suppose the data described in Table 1 were collected in year 2000, and you have another dataset collected in year 2005, which includes the same information (variables) for another 1,000 working males. You then pool these two cross-sectional datasets together.

B.1 State a model you want to estimate, so that you can tell if there is any change in wage return to education and/or wage return to experience over the two time periods. (5 marks)

B.2 Explain how you can test there is change in return to education. You need to state your null hypothesis and alternative hypothesis; provide calculation method for test statistics; and the rejection rule. (10 marks)

(C) For equation (a) stated in (A) above, if we put *abil* into the error term, and we are left with the following model specification,

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + Z\gamma + e \qquad \textbf{(b)}$$

It is very likely that *educ* is an endogenous variable in Model (b), and you assume all other variables are exogenous. Suppose you use the variable of *sibs* as the instrumental variable for *educ*, and Model (b) can be estimated by IV estimator (using data described by Table 1). The estimation results from both OLS and IV estimator are presented in Table 3 below.

Table 3

	OLS estimator		IV estimator	
	Coef.	Std. Err.	Coef.	Std. Err.
educ	0.065	0.006	0.084	0.033
exper	0.014	0.003	0.018	0.008
tenure	0.012	0.002	0.011	0.003
married	0.199	0.039	0.202	0.039
black	-0.188	0.038	-0.170	0.048
south	-0.091	0.026	-0.087	0.027
urban	0.184	0.027	0.179	0.028
_cons	5.395	0.113	5.092	0.524
Observations	935		935	
R-squared	0.253		0.245	

C.1 What are the assumptions required to make *sibs* a valid instrumental variable for *educ*? (5 marks)

C.2 Explain in detail how to test if *educ* is endogenous. (10 marks)

(D) For data described in Table 1, suppose for each male you also know the number of cigarettes he smoked per day (*cigs*), and you want to figure out the effects of smoking on wage. The model you are going to estimate is

$$lwage = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 exper + u \qquad \textbf{(c)}$$

To reflect the fact that cigarette consumption might be jointly determined with income, a demand for cigarettes equation is

$$cigs = \gamma_0 + \gamma_1 lwage + \gamma_2 educ + \gamma_3 exper + \gamma_4 log(cigpric) + \gamma_5 restaurn + v \qquad \textbf{(d)}$$

where *cigpric* is the price of a pack of cigarettes (in pennies), and *restaurn* is a dummy variable equal to 1 if the person lives in a county with restaurant smoking restrictions, 0 otherwise. Assume these are exogenous to the individual.

D.1 what signs would you expect for γ_4 and γ_5? (5 marks)

D.2 Under what assumption is the wage equation (c) identified? (5 marks)

D.3 Is cigarette equation (d) identifiable? Why? (5 marks)

Section II Analysis on Industrial Wage with Time Series Data (20 marks)

You have a time series dataset on monthly employment and wage information for the industry sector 232, i.e. "Men's and Boys' Furnishings", over 51 years (i.e. T = 612 months). The following table, Table 4, describes the variables included in the dataset.

Table 4

Variable	**Definition**
emp232	employment, 1000s
wage232	hourly wage
lemp232	log(*emp232*)
lwage232	log(*wage232*)
minwage	national minimum wage, £/hour
lminwage	log(*minwage*)
cpi	Consumer Price Index
lcpi	log(*cpi*)
gemp232	*lemp232*(*t*) - *lemp232*(*t*-1)
gwage232	*lwage232*(*t*) - *lwage232*(*t*-1)
gmwage	*lminwage*(*t*) - *lminwage*(*t*-1)
gcpi	*lcpi*(*t*) - *lcpi*(*t*-1)

1. Run the regression *gwage232* on *gmwage*, *gcpi*. The estimated equation is as follows:

$$\underset{}{gwage232} = \underset{(0.0004)}{.0022} + \underset{(0.01)}{0.151}\, gmwage + \underset{(0.082)}{0.244}\, gcpi$$

$$n = 611,\ R^2 = 0.293$$

Interpret $\hat{\beta}_{gmwage}$. (5 marks)

2. To estimate the 12-month long-run effect of minimum wage growth on wage growth in sector 232, what model should you estimate? State your model. Based on your model specification, what is the 12-month long-run effect? (5 marks)

3. After running the regression *gwage232* on *gmwage* and *gcpi*, you want to test the errors for AR(1) serial correlation. Explain how you are going to do the test without the assumption that *gmwage* and *gcpi* are strictly exogenous. (10 marks)

Section III Analysis on Individual Wage with Panel Data (20 marks)

Suppose you have a balanced panel dataset comprising a sample of 545 full-time working males who have completed their schooling by 1980 and then followed over the period 1980 to 1987. Note: their working status had never changed over the time period, i.e. always full time working. Table 5 provides the variable names and their definitions.

Table 5

Variable	**Definition**
lwage	log(wage)
educ	years of schooling
union	=1 if in union
black	=1 if black
hisp	=1 if Hispanic
exper	labor market experience
married	=1 if married
d81	=1 if year == 1981
d82	=1 if year == 1982
d83	=1 if year == 1983
d84	=1 if year == 1984
d85	=1 if year == 1985
d86	=1 if year == 1986
d87	=1 if year == 1987

1. Consider the unobserved effects model

$$\begin{aligned} lwage_{it} = \beta_0 + \delta_1 d81_t + \ldots + \delta_7 d87_t + \beta_1 educ_i + \gamma_1 d81_t educ_i \\ + \ldots + \gamma_7 d87_t educ_i + \beta_2 union_{it} + \beta_3 married_{it} + a_i + u_{it} \end{aligned}, \qquad \textbf{(e)}$$

where a_i is allowed to be corrected with $educ_i$ and $union_{it}$. Which parameters can you estimate using first differencing (FD)? (5 marks)

2. Estimate the equation (e) by fixed effects (FE) estimation, and state how to test that the return to education has not changed over time. You need to state your null hypothesis and alternative hypothesis; provide calculation method for test statistics; and the rejection rule. (10 marks)

3. Suppose you apply two methods, i.e. random effects (RE) and fixed effects (FE), to estimate a wage equation using the data described in Table 5. Results are presented in Table 6 below. Discuss why *exper* is dropped in the FE analysis. (5 marks)

Table 6

	Random Effects		Fixed Effects	
	Coef.	Std. Err.	Coef.	Std. Err.
educ	0.092	0.011	-	-
black	-0.139	0.048	-	-
hisp	0.022	0.043	-	-
exper	0.106	0.015	-	-
expersq	-0.005	0.001	-0.005	0.001
married	0.064	0.017	0.047	0.018
union	0.106	0.018	0.080	0.019
d81	0.040	0.025	0.151	0.022
d82	0.031	0.032	0.253	0.024
d83	0.020	0.042	0.354	0.029
d84	0.043	0.051	0.490	0.036
d85	0.058	0.061	0.617	0.045
d86	0.092	0.071	0.765	0.056
d87	0.135	0.081	0.925	0.069