BEE2031

UNIVERSITY OF EXETER

BUSINESS SCHOOL

January 2018

Econometrics

Module Convenor: Dr Xiaohui Zhang

Duration: ONE and HALF HOURS

Answer ALL questions

Answer ALL questions in the answer booklet.

Materials to be supplied:
Answer booklet

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 materials distributed in any form are allowed.

Section I Cross-Sectional Data Analysis [60 marks]

A. [40 marks] Crime of Individual Men

Suppose you have data on arrests during the year 1986 and other information on 2,725 men born in either 1960 or 1961 in Devon, UK. Each man in the sample was arrested at least once prior to 1986. Variables available in this dataset are as follows:

Variable name	Variable label
narr86	number of times arrested in 1986
pcnv	proportion of arrests prior to 1986 that led to conviction
avgsen	average sentence length (in months) served for prior convictions
tottime	total months spent in prison since 18 years old
ptime86	months spent in prison during 1986
qemp86	number of quarters employed in 1986
inc86	legal income in 1986, £100s
durat	recent unemployment duration
pcnvsq	square of pcnv
pt86sq	square of ptime86
inc86sq	square of inc86

The variable *narr86* is zero for most men in the sample (72.29%), and it varies from 0 to 12. The percentage of men arrested once during 1986 was 20.51.

A.1 [15 marks] Two regression models are estimated and the Stata outputs are as follows:

Model 1: $narr86 = \beta_0 + \beta_1 pcnv + ... + \beta_7 durat + u$

Source	SS	df	MS	Number of obs	=	2,725
				F(7, 2717)	=	20.99
Model	103.118857	7	14.7312653	Prob > F	=	0.0000
Residual	1907.2283	2,717	.701961096	R-squared	=	0.0513
				Adj R-squared	=	0.0488
Total	2010.34716	2,724	.738012906	Root MSE	=	.83783

narr86	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pcnv	-.1540919	.0406969	-3.79	0.000	-.233892	-.0742919
avgsen	-.0085966	.0123659	-0.70	0.487	-.0328442	.0156509
tottime	.0127	.0095384	1.33	0.183	-.0060032	.0314032
ptime86	-.0386847	.0090322	-4.28	0.000	-.0563953	-.0209741
qemp86	-.0593741	.0158381	-3.75	0.000	-.0904302	-.0283181
inc86	-.001664	.0003437	-4.84	0.000	-.0023379	-.0009901
durat	-.0038305	.0040536	-0.94	0.345	-.0117789	.0041179
_cons	.7064818	.0411838	17.15	0.000	.625727	.7872366

Model 2: $narr86 = \beta_0 + \beta_1 pcnv + \beta_2 pcnvsq + ... + \beta_{10} durat + u$

Source	SS	df	MS	Number of obs	=	2,725
				F(10, 2714)	=	26.30
Model	177.573199	10	17.7573199	Prob > F	=	0.0000
Residual	1832.77396	2,714	.675303595	R-squared	=	0.0883
				Adj R-squared	=	0.0850
Total	2010.34716	2,724	.738012906	Root MSE	=	.82177

narr86	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pcnv	.6493028	.1542444	4.21	0.000	.3468544	.9517512
pcnvsq	-.8527109	.1559456	-5.47	0.000	-1.158495	-.5469267
avgsen	-.0149904	.0121442	-1.23	0.217	-.0388032	.0088224
tottime	.0124167	.0093577	1.33	0.185	-.0059322	.0307656
ptime86	.3053407	.0446319	6.84	0.000	.2178248	.3928567
pt86sq	-.0312617	.0039071	-8.00	0.000	-.0389228	-.0236006
qemp86	-.025244	.0184954	-1.36	0.172	-.0615104	.0110225
inc86	-.0037379	.0008089	-4.62	0.000	-.0053241	-.0021517
inc86sq	7.87e-06	2.58e-06	3.06	0.002	2.82e-06	.0000129
durat	-.0077057	.0039956	-1.93	0.054	-.0155404	.0001289
_cons	.6411596	.041764	15.35	0.000	.5592672	.723052

1a) [5 marks] Interpret the estimated impacts of income on the number of times arrested from both *Model 1* and *Model 2*.

1b) [10 marks] What kind of misspecification may *Model 1* suffer from? Explain how to test it by RESET.

A.2 [17 marks] From the data provided above, you can generate a new variable, arr86, which is 1 if narr86>0, otherwise is 0. A probit model and a logit model are estimated in Stata and the outputs are as follows:

Probit regression — Number of obs = 2,725; LR chi2(7) = 202.82; Prob > chi2 = 0.0000

Log likelihood = -1506.7733 — Pseudo R2 = 0.0631

arr86	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pcnv	-.5639463	.0711265	-7.93	0.000	-.7033516	-.4245409
avgsen	.0138712	.0208739	0.66	0.506	-.0270408	.0547833
tottime	-.0045815	.0164409	-0.28	0.781	-.036805	.027642
ptime86	-.0687106	.0181213	-3.79	0.000	-.1042277	-.0331934
qemp86	.0261471	.0263773	0.99	0.322	-.0255514	.0778456
inc86	-.005138	.000684	-7.51	0.000	-.0064787	-.0037974
durat	.0119642	.0062027	1.93	0.054	-.000193	.0241213
_cons	-.2268796	.0643515	-3.53	0.000	-.3530062	-.100753

Logistic regression — Number of obs = 2,725; LR chi2(7) = 202.67; Prob > chi2 = 0.0000; Log likelihood = -1506.8464; Pseudo R2 = 0.0630

arr86	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pcnv	-.9415939	.1220454	-7.72	0.000	-1.180799	-.7023893
avgsen	.0225457	.0345375	0.65	0.514	-.0451467	.090238
tottime	-.0074597	.0275246	-0.27	0.786	-.0614069	.0464875
ptime86	-.1097529	.0311666	-3.52	0.000	-.1708384	-.0486675
qemp86	.061598	.0445023	1.38	0.166	-.0256249	.1488209
inc86	-.009486	.0012802	-7.41	0.000	-.0119952	-.0069768
durat	.0194932	.0100699	1.94	0.053	-.0002435	.0392299
_cons	-.364998	.1053356	-3.47	0.001	-.5714519	-.1585442

2a) [7 marks] Compare the results from the probit model and the logit model. For example, the estimated coefficient of inc86 by the probit model is -0.005, while it is -0.009 estimated by logit model. Are they showing that income has different impacts on the probability of being arrested, estimated by the two models? Explain your answer.

2b) [10 marks] In order to better understand the impact of the regressors in the above probit and logit models, you can calculate their partial effect at the average (PEA) and average partial effect (APE). Explain the difference between PEA and APE. Do you expect that the APE's estimated from the probit model are similar to those estimated from the logit model or not? Explain your answer.

A.3 [5 marks] The output from the command of "*tab narr86*" in Stata is as follows:

.

# times arrested, 1986	Freq.	Percent	Cum.
0	1,970	72.29	72.29
1	559	20.51	92.81
2	121	4.44	97.25
3	42	1.54	98.79
4	12	0.44	99.23
5	13	0.48	99.71
6	4	0.15	99.85
7	1	0.04	99.89
9	1	0.04	99.93
10	1	0.04	99.96
12	1	0.04	100.00
Total	2,725	100.00	

Explain why a Poisson regression is more suitable than a linear regression model.

A.4 [3 marks] A Poisson regression model is estimated and the Stata output is as follows:

```
Poisson regression                              Number of obs     =      2,725
                                                LR chi2(7)        =     293.64
                                                Prob > chi2       =     0.0000
Log likelihood = -2295.1017                     Pseudo R2         =     0.0601
```

narr86	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pcnv	-.432861	.0833474	-5.19	0.000	-.5962189	-.2695031
avgsen	-.0182917	.0195457	-0.94	0.349	-.0566005	.0200171
tottime	.0257075	.014473	1.78	0.076	-.0026591	.0540741
ptime86	-.0923174	.0207045	-4.46	0.000	-.1328976	-.0517373
qemp86	-.0485177	.0307015	-1.58	0.114	-.1086916	.0116562
inc86	-.0087713	.0010463	-8.38	0.000	-.0108221	-.0067206
durat	-.0072274	.0066736	-1.08	0.279	-.0203074	.0058526
_cons	-.282345	.0653299	-4.32	0.000	-.4103891	-.1543008

Interpret the estimated coefficient of "*inc86*".

B. [20 marks] Wage of individual men

You have a cross-sectional dataset for 935 working males and the description of the data is as follows:

Variable Name	Variable Label
wage	monthly earnings in British Pound
lwage	natural log of wage
educ	years of education
exper	years of work experience
tenure	years with current employer
age	age in years
married	=1 if married, =0 otherwise
south	=1 if live in south, =0 otherwise
IQ	IQ score
sibs	number of siblings

Suppose you run a regression model as:

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age + \beta_5 married + \beta_6 south + u$$

B.1 [4 marks] It is very likely that *educ* is an endogenous variable in the above regression model. Name two methods to solve, or at least mitigate, the endogeneity issue.

B.2 [6 marks] Choose one solution you proposed in B.1 and explain in detail what is required in terms of data and assumptions.

B.3 [10 marks] Explain in detail how to test if *educ* is endogenous.

Section II Time Series Data Analysis [20 marks]

Suppose you have data on inflation rates and unemployment rates for the UK from year 1948 to 2003. Description of the data is as follows:

Variable Name	Variable Label
year	1948 through 2003
unem	civilian unemployment rate, %
inf	percentage change in CPI
inf_1	one year lagged *inf*
unem_1	one year lagged *unem*
cinf	*inf - inf_1*
cunem	*unem - unem_1*

Using the above data, you estimate a static Phillip curve, given by $inf_t = \beta_0 + \beta_1 unem_t + u_t$. The Stata output is:

Source	SS	df	MS	Number of obs	=	56
				F(1, 54)	=	3.58
Model	31.599858	1	31.599858	Prob > F	=	0.0639
Residual	476.815691	54	8.8299202	R-squared	=	0.0622
				Adj R-squared	=	0.0448
Total	508.415549	55	9.24391907	Root MSE	=	2.9715

inf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
unem	.5023782	.2655624	1.89	0.064	-.0300424	1.034799
_cons	1.053566	1.547957	0.68	0.499	-2.049901	4.157033

1. [4 marks] Interpret the estimated parameter $\hat{\beta}_1$.

2. [6 marks] Is $\hat{\beta}_1$ an unbiased estimation of β_1? Why?

3. [10 marks] Explain in detail how to informally test if *inf* and *unem* are I(1) or I(0). If one or both are I(1), how can you gain a consistent estimation of the impact of unemployment rates on inflation rates?

Section III Panel Data Analysis [20 marks]

Suppose you have a balanced panel dataset comprising a sample of 545 full-time working males who had **completed their schooling by 1980** and were then followed over the period 1980 to 1987. Description of the data is as follows:

Variable Name	Variable Label
lwage	log(wage)
educ	years of schooling
union	=1 if member of a union, =0 otherwise
exper	labor market experience
married	=1 if married
d81	=1 if year == 1981, =0 otherwise
d82	=1 if year == 1982, =0 otherwise
d83	=1 if year == 1983, =0 otherwise
d84	=1 if year == 1984, =0 otherwise
d85	=1 if year == 1985, =0 otherwise
d86	=1 if year == 1986, =0 otherwise
d87	=1 if year == 1987, =0 otherwise

1. [5 marks] Specify a panel data model, which allows you to estimate the impact of education (years of schooling) on wage, and to test if the return to education has changed over time or not.

2. [3 marks] If you apply first differencing (FD) to the model you specified above, will you obtain an estimate of the impact of education on wage?

3. [12 marks] Suppose you apply two methods, random effects (RE) and fixed effects (FE), to estimate a wage equation using the data provided. The estimated results are summarized as follows:

	Random Effects		Fixed Effects	
	Coef.	Std.	Coef.	Std.
educ	0.092	0.011	-	-
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

a) [6 marks] Discuss why *educ* is dropped in the FE analysis.

b) [6 marks] What might be the reason that the impact of *married* on wage estimated by FE is much lower than that estimated by RE?