EX0012

BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF COMMERCE

DEPARTMENT OF ECONOMICS

NOV 2024

MSc ECONOMICS

ADVANCED ECONOMETRICS (MEC 502)

EXAMINATION **DURATION: 3 HOURS**

INSTRUCTIONS TO CANDIDATES

1. Answer question 1 in Section A and any other three questions from Section B.
2. Question 1 carries 40 marks.
3. All the questions in Section B carry equal marks of 20 each.
4. Cell-phones are not allowed into the examination room.

SECTION A (COMPULSORY)

Question 1

a. i. Under what assumptions is the Regression Discontinuity Design (RDD) valid? **(6 marks)**

ii. With the aid of a graph illustrate what you understand by the common trends ussumption in Difference in Difference (DiD) estimation. **(6 marks)**

iii. Using an example, explain what you understand by contamination bias in experimental research design. **(8 marks)**

A researcher estimated an earnings equation and obtained the following results:

$$\widehat{logearn} = 1340.4 + 0.0193wkhrs + 0.4579edu + 0.530age - 0.134age^2 - 0.0015male$$
$$(206.31)\ (0.009)\qquad (0.061)\qquad (0.021)\qquad (0.032)\qquad (0.0002)$$

$n = 1200$, $R^2 = 0.62$, $\overline{R}^2 = 0.567$ $DW = 1.832$ F- stat (probability) = 61.456 (0.001). The variable *logearn* is the natural logarithm of total earnings in US dollars per month, *wkhrs* is total monthly hours spent working, *edu* and *age* is education and age of the worker measured in years respectively, and *male* is a gender dummy. Values in parentheses are p-values.

(i) Can you estimate the above model using logit? Why or why not? If not suggest a more suitable estimator. **(4 marks)**

(ii) All other factors being equal, is there evidence of gender discrimination against men? How strong is the evidence? **(5 marks)**

(iii) Test the null hypothesis that, holding other factors fixed, education has no effect on earnings? **(5 marks)**

(iv) Give an economic interpretation of the variables *male* and age^2. **(6 marks)**

[40 marks]

SECTION B (ANSWER ANY THREE QUESTIONS)

Question 2

After receipt of your MSc degree you are employed by Picnic, an online supermarket that offers free home delivery. You are hired to optimize the presentation of the Picnic-app, a key element of the Picnic enterprise. The app allows for easy online shopping and tracking of the delivery. The company is accustomed to doing "A/B" testing where clients randomly receive a different version of the app, version A or version B. In week 42 of 2019, you extend this methodology and add a third version, C. The aim is to find which version of the app generates the largest revenue. After randomizing over all clients that have installed the app and collecting data, you run the following code:

```
. desc *
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
memberdur       double  %10.0g                Membership duration in weeks
male            double  %10.0g                Male client
rural           double  %10.0g                Client lives in rural area
android         double  %10.0g                Client uses android phone
A               double  %10.0g                Treatment A: order by alphabet
B               double  %10.0g                Treatment B: order by product type
C               double  %10.0g                Treatment C: order by type and purchase prob.
spent           double  %10.0g                Client total expenditure in week 42 (euros)
bought          double  %10.0g                Client bought one or more items in week 42

. sum *, format
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   memberdur |    147,874       10.10        8.97      1.00      48.00
        male |    147,874        0.60        0.49      0.00       1.00
       rural |    147,874        0.20        0.40      0.00       1.00
     android |    147,874        0.72        0.45      0.00       1.00
           A |    147,874        0.52        0.50      0.00       1.00
-------------+---------------------------------------------------------
           B |    147,874        0.24        0.43      0.00       1.00
           C |    147,874        0.24        0.43      0.00       1.00
       spent |    147,874      121.78       79.40      0.00     485.88
      bought |    147,874        0.92        0.28      0.00       1.00
```

EX0012

```
. qui eststo a: reg spent B C, r
. qui eststo b: reg spent B C memberdur male rural android, r
. qui eststo c: reg spent B C memberdur male rural android if android==0, r
. qui eststo d: reg spent B C memberdur male rural android if android==1, r
. esttab a b c d,  se b(a2) nogaps compress noomit
```

	(1) spent	(2) spent	(3) spent	(4) spent
B	16.9***	13.8***		13.8***
	(0.51)	(0.56)		(0.56)
C	22.0***	19.2***		19.2***
	(0.51)	(0.56)		(0.56)
memberdur		1.87***	1.86***	1.87***
		(0.022)	(0.041)	(0.026)
male		27.9***	27.0***	28.3***
		(0.40)	(0.73)	(0.48)
rural		-35.4***	-34.5***	-35.7***
		(0.46)	(0.84)	(0.55)
android		-0.79		
		(0.54)		
_cons	112.5***	86.0***	86.3***	85.0***
	(0.28)	(0.46)	(0.64)	(0.57)
N	147874	147874	41327	106547

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Given the Stata output, answer the following:

a. Interpret the size and significance of the estimated coefficient of *B* in col. (1). **(5 marks)**

b. Judging from column (1), is there a significant difference between version C and version B of the app? Calculate. (use $t_{5\%}^{critical} = 1.96$) **(7 marks)**

c. Does the Stata output provide supporting evidence for random assignment or does it contain counter evidence? Provide 3 pieces of evidence shown in the results table. **(8 marks)**

[20 marks]

Question 3

After obtaining your MSc degree you are recruited as an economist at the World Health Organization (WHO). There, you are asked to investigate the economic benefits of home-birth compared to giving birth in hospital. Supposedly hospitals are safer, but they are also more costly. You recall that in the Netherlands 25% of babies are born at home, so that should give you a setting where you could compare 7-day newborn *mortalityi* (measured in deaths per thousand) between *hospitali* (either 1 or 0) and home. With data collected by the Dutch statistics bureau, you run the following Stata code:

```
. eststo ols1: reg mortality hospital, r
Linear regression                               Number of obs =     356412
                                                F(  1,356410) =      69.09
                                                Prob > F      =     0.0000
                                                R-squared     =     0.0002
                                                Root MSE      =     41.128
```

mortality	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hospital	1.079941	.1299219	8.31	0.000	.8252877	1.334594
_cons	.9930725	.0891368	11.14	0.000	.8183076	1.167777

Given the Stata output, answer the following:

a. Interpret the size and significance of the estimated coefficient of *hospital*. **(5 marks)**

You are skeptical that the obtained estimate can be given a causal interpretation, so you look in the data for more information. There you find two additional variables, the *distancei* to hospital from the mother's home (in 10km), and a *riski* assessment measure (ranging from 0 to 1) that is made on the basis of medical factors after 6 months of pregnancy. Using these variables you run the following:

```
. eststo ols2: reg mortality hospital risk, r
. qui eststo fs1:  reg hospital  distance , r
. qui eststo fs2:  reg hospital  distance risk , r
. qui eststo iv:   ivregress 2sls mortality risk (hospital = distance), r
. esttab ols2 fs2 iv,  se nogaps b(a2) star(* 0.1 ** 0.05 *** 0.01) sca(N F r2 cmd) sfmt(%8.0f %8.2f)
```

	(1) mortality	(2) mortality	(3) hospital	(4) hospital	(5) mortality
hospital	1.08***	-0.19			-8.68*
	(0.13)	(0.13)			(4.76)
distance			0.0083***	-0.042***	
			(0.0019)	(0.0018)	
risk		7.26***		0.49***	11.3***
		(0.32)		(0.0026)	(2.32)
_cons	0.99***	-1.81***	0.65***	0.43***	1.67
	(0.089)	(0.11)	(0.0012)	(0.0017)	(1.96)
N	356412	356412	356412	356412	356412
F	69.09	264.40	19.10	18172.75	
r2	0.00	0.00	0.00	0.00	.
cmd	regress	regress	regress	regress	ivregress

Standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Now answer the following:

b. Explain the difference between column (1) and (2). **(4 marks)**

c. The IV estimate in column (5) suggests that childbirth in *hospital* indeed reduces new-born *mortality*. Who are the compliers of this instrument? Give your answer by selecting one of the options given below. Explain.

1. People that decide to plan birth in hospital because the hospital is close by.

2. Home births that result in a complication that are quickly moved to hospital because it is close by.

3. High risk pregnancies that are planned for delivery in hospital because of a medical indication.

4. Both **1** and **2.**

5. Both **2** and **3.**

6. Both **1** and **3.**

7. All of the above (**1** and **2** and **3**). **(4 marks)**

d. Do you believe the instrument is exogenous? Does the table in the Stata output provide (suggestive) evidence of non-random assignment of the instrument?

(7 marks)

[20 marks]

Question 4

In Portugal the number of years of schooling increased in the last 20 years, while the participation in the elections decreased in the same period. Therefore, this is a clear evidence that a more educated population caused a reduction in the percentage of people voting in the elections.

Discuss the above assertion using examples citing (i) confounding, (ii) non-randomisation of treatment, and (iii) the counterfactual problem. **(20 marks)**

[20 marks]

Question5

You have the following German panel data, where *docvis* is the number of visits to a doctor over the last three months. *hhkids* is a dummy indicating presence of children in the household. The data includes years 1988, 1991 and 1994. Other variables are self-explanatory.

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
docvis	2,661	3.295378	5.710886	0	95
age	2,661	47.47802	8.8763	29	64
female	2,661	.4227734	.4940329	0	1
hhkids	2,661	.3615182	.4805305	0	1
year1988	2,661	.3333333	.4714931	0	1
year1991	2,661	.3333333	.4714931	0	1
year1994	2,661	.3333333	.4714931	0	1

Next, you estimate:

```
. xtreg docvis female hhkids year1991 year1994,fe
note: female omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =       2,661
Group variable: id                              Number of groups   =         887

R-sq:                                           Obs per group:
     within  = 0.0281                                         min =             3
     between = 0.0324                                         avg =           3.0
     overall = 0.0020                                         max =             3

                                                F(3,1771)          =       17.10
corr(u_i, Xb)  = -0.1205                        Prob > F           =      0.0000

------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |          0  (omitted)
      hhkids |   .8138992   .4523825     1.80   0.072    -.0733605    1.701159
    year1991 |   .3872425    .233322     1.66   0.097    -.0703729     .844858
    year1994 |   1.618385    .237652     6.81   0.000     1.152878    2.083893
       _cons |   2.342386   .2553122     9.14   0.000     1.831651    2.853114
-------------+----------------------------------------------------------------
     sigma_u |  4.1478335
     sigma_e |  4.8410849
         rho |   .4232433   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(886, 1771) = 2.08                     Prob > F = 0.0000
```

a) What kind of a panel data estimation is this? Explain. **(4 marks)**

b) Interpret the estimate for *hhkids*. **(4 marks)**

c) Why does the row for female show zero? **(4 marks)**

d) The output says that R-sq:within is 0.0281. What does this mean? **(4 marks)**

e) Is this a balanced panel data? Explain. **(4 marks)**

[20 marks]

END OF PAPER