Module Code
ECO00005M

BA, BSc, MSc Degree Examinations 2020-1

Department:
Economics

Title of Exam:
Applied Microeconometrics

Time Allowed: 24 Hours
(PLEASE NOTE: Submissions up to 30 minutes late will receive a 5 mark penalty deduction. Submissions late by more than 30 minutes, will not be marked.)

Time Recommended:
TWO hours

Word limit:
There is a word limit of 1,200 words per question.

You have the flexibility on how to split the word limit between the different parts of each question.

The word count does not include: figures, graphs and equations.
A bibliography is **NOT** required. In-text references (Author, date) are sufficient.

Allocation of Marks:
All questions have the same weight. The weights of the parts of each question vary and are stated in percentage points.

Instructions for Candidates:
Answer TWO out of the four questions.

If you answer more than two questions, only the first two answers in the order they appear in the script, will be marked.

Any answers you do not wish to be included in the marking, **must** be clearly crossed out.

Materials Required:
Calculator

A note on Academic Integrity

We are treating this online examination as a time-limited open assessment. You are therefore permitted to refer to written, and online materials, to aid you in your answers.

However, you must ensure that the work you submit is entirely your own, and for the whole time the assessment is live you must not:

- communicate with departmental staff on the topic of the assessment
- communicate with other students on the topic of this assessment
- seek assistance with the assignment from the academic and/or disability support services, such as the Writing and Language Skills Centre, Maths Skills Centre and/or Disability Services. (The only exception to this will be for those students who have been recommended an exam support worker in a Student Support Plan. If this applies to you, you are advised to contact Disability Services as soon as possible to discuss the necessary arrangements)
- seek advice or contribution from any third party, including proofreaders, friends, or family members.

We expect, and trust, that all our students will seek to maintain the integrity of the assessment, and of their award, by ensuring that these instructions are strictly followed.

Failure to adhere to these requirements will be considered a breach of the Academic Misconduct regulations, where the offences of plagiarism, breach/cheating, collusion and commissioning are relevant - see AM.1.2.1" *(Note this supersedes section 7.3 of the Guide to Assessment).*

Question 1
1,200 words limit

(a) [25%]
Discuss an empirical example (different from lectures and seminars' examples) of a linear regression model with endogeneity. Write the regression equation and provide details on the dependent and explanatory variables and on the interpretation of the coefficients. Explain the cause of endogeneity and why the ordinary least squares estimation would be inconsistent. More points will be given to realistic empirical examples with more than one explanatory variable and with an appropriate choice of explanatory variables.

(b) [25%]
Explain how you would estimate your model in (1.a) using a two-stage least squares estimation and define the instrument(s) you would use. Provide details on how the two-stage least squares estimation is computed. Write the formula for the two-stage least squares estimation considering the model defined in point (1.a).

(c) [15%]
Explain what assumptions your instrumental variable(s) must satisfy to produce a consistent estimation of the model defined in point (1.a). Show that the instrumental variable estimation defined in (1.b) is consistent under these assumptions.

(d) [15%]
Explain how you would test for the validity of your instrumental variable(s). Explain also how you would test for whether there is an endogeneity issue in your model. Provide details on how you would perform these tests using the model defined in point (1.a).

(e) [20%]
Discuss potential drawbacks of the instrumental variable(s) you proposed in point (1.b). Discuss also what is the consequence of using an instrument whose effect on the endogenous variable conditional on the remaining control variables is not statistically very significant.

Question 2
1,200 words limit

(a) [30%]
Discuss an empirical example (different from lectures and seminars' examples) of a panel data model where you would use a fixed effect estimation rather than a random effect estimation. Write the regression equation and provide details on the dependent and explanatory variables, on the error term and on the interpretation of the coefficients. Explain how you would compute the fixed effect estimation using your defined model.

(b) [35%]
Explain what the unobserved individual effects in the model defined in (2.a) capture.
Explain the differences in the assumptions needed for the consistency of the fixed effect estimation and of the random effect estimation for the model you discussed in (2.a). Why is the fixed effect estimation more appropriate than the random effect

Question continued overleaf

estimation in the empirical example you discussed in point (2.a)? Explain how you would perform a test to decide whether to adopt a random effect or a fixed effect estimation.

(c) [35%]
Discuss an empirical example of a panel data model where one of the explanatory variables is endogenous because it is correlated with unobserved variables that are relevant to explain both the dependent variable and the endogenous variable. Explain under which conditions the fixed effect estimation can solve such an issue of endogeneity. Explain which type of estimation you would adopt to solve the endogeneity issue if the conditions for the consistency of the fixed effect estimation were not satisfied.

Question 3
1,200 words limit

(a) [35%]
Suppose that a researcher has information on the type of health insurance for a random sample of individuals. The researcher observes a categorical variable, *insure*, taking value 1 for individuals who choose an *indemnity* plan (a fee-for-service insurance), 2 for individuals who choose a *prepaid* plan (a fixed up-front payment with unlimited use) and 3 for individuals who are uninsured (*uninsure)*. The researcher wants to analyse the demographic factors linked with the three choices of insurance and observes for each individual the following variables: *age* in years, *male* which is a dummy variable taking value 1 for men and 0 otherwise, and *nonwhite* which is a dummy variable taking value 1 for people who are not of white ethnicity and 0 otherwise. Provide an interpretation of the results that are reported in Table 3.1 below. Write down the model that the researcher has estimated. Explain how the estimation has been computed.

Question continued overleaf

Table 3.1

```
. mlogit insure age male nonwhite

Iteration 0:   log likelihood = -555.85446
Iteration 1:   log likelihood = -545.60089
Iteration 2:   log likelihood = -545.58328
Iteration 3:   log likelihood = -545.58328

Multinomial logistic regression                 Number of obs     =        615
                                                LR chi2(6)        =      20.54
                                                Prob > chi2       =     0.0022
Log likelihood = -545.58328                     Pseudo R2         =     0.0185
```

insure	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Indemnity	(base outcome)					
Prepaid						
age	-.0111915	.0060915	-1.84	0.066	-.0231305	.0007475
male	.5739825	.2005221	2.86	0.004	.1809665	.9669985
nonwhite	.7312659	.218978	3.34	0.001	.302077	1.160455
_cons	.1567003	.2828509	0.55	0.580	-.3976773	.7110778
Uninsure						
age	-.0058414	.0114114	-0.51	0.609	-.0282073	.0165245
male	.5102237	.3639793	1.40	0.161	-.2031626	1.22361
nonwhite	.4333141	.4106255	1.06	0.291	-.371497	1.238125
_cons	-1.811165	.5348606	-3.39	0.001	-2.859473	-.7628578

(b) [25%]
Compute the log-odds ratio Log[Pr(insure=2|x)/Pr(insure=1|x)], where Pr(insure=2|x) is the probability that the categorial variable for the health insurance choice (insure) takes value 2 (prepaid) conditional on x, which is the vector of explanatory variables used by the researcher in Table 3.1, and Pr(insure=1|x) is the probability that the categorial variable for the health insurance choice (insure) takes value 1 (indemnity) conditional on x. Provide an interpretation of the results in Table 3.1. using the log odds ratio.

(c) [15%]
The researcher also computed the results reported in Table 3.2 below. What has the researcher computed? Provide an interpretation of the results in Table 3.2.

Question continued overleaf

Table 3.2

```
. predict p1 p2 p3 if e(sample), p

.
. gen Pinsure= .
(615 missing values generated)

.
. egen atest = rmax(p1 p2 p3)  if e(sample)

.
. foreach n of numlist 1/3 {
  2.         replace Pinsure = `n' if p`n'==atest
  3.                 }
(379 real changes made)
(236 real changes made)
(0 real changes made)

.
. gen pdiff = (Pinsure == insure)

. ta pdiff
```

pdiff	Freq.	Percent	Cum.
0	289	46.99	46.99
1	326	53.01	100.00
Total	615	100.00	

(d) [25%]

Discuss how you would estimate a model to predict the effect of gender, non-white ethnicity and age on the probability of being uninsured if you were able to observe whether an individual is insured or not but were unable to observe the type of their insurance if they are insured. Define the model and explain how you would estimate it.

Question 4
1,200 words limit

(a) [20%]

A researcher used a random sample of 4,642 child births to study how the number of prenatal medical visits (*nprenatal*) is related to a set of explanatory variables which are: *mmarried*, a dummy variable taking value 1 if the mother is married and 0 otherwise; *foreign*, a dummy variable taking value 1 if the mother is foreign and 0 otherwise; *mage*, the mother's age in years; *medu*, the mother's education in years; *shortgap*, a dummy variable taking value 1 if the gap between the last child pregnancy and the current pregnancy is lower or equal to 18 months and 0 if it is longer or if the current child birth is the first; *fbaby*, a dummy variable taking value 1 for first child births and 0 otherwise; *order*, the birth order of the new-born. The researcher produced the results reported in Table 4.1. What type of estimation method has the

Question continued overleaf

researcher adopted to study the relationship between *nprenatal* and *the* explanatory variables? Explain how the estimation has been computed. Provide a detailed interpretation of the results reported in Table 4.1.

Table 4.1

```
. sum  nprenatal

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   nprenatal |      4,642    10.75808    3.681084          0         40

. regress nprenatal mmarried foreign  mage medu shortgap fbaby order , robust

Linear regression                               Number of obs     =      4,642
                                                F(7, 4634)        =      68.41
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1174
                                                Root MSE          =     3.4608

------------------------------------------------------------------------------
             |               Robust
   nprenatal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    mmarried |   1.481734    .141469    10.47   0.000     1.204388    1.759081
     foreign |  -.3721951   .2550216    -1.46   0.145    -.8721589    .1277686
        mage |   .0634913   .0121004     5.25   0.000     .0397686    .0872139
        medu |   .1193324   .0259211     4.60   0.000     .0685148      .17015
    shortgap |  -1.169561   .2308484    -5.07   0.000    -1.622133   -.7169879
       fbaby |  -.0959504   .1521806    -0.63   0.528    -.3942968    .2023961
       order |  -.5328748   .0795283    -6.70   0.000    -.6887881   -.3769615
       _cons |   7.685075   .4028786    19.08   0.000     6.895241    8.474909
------------------------------------------------------------------------------

. test foreign fbaby

 ( 1)  foreign = 0
 ( 2)  fbaby = 0

       F(  2,  4634) =    1.28
            Prob > F =    0.2768
```

(b) [30%]

Using the data described in point (4.a), the researcher produced the results reported in Table 4.2. What type of estimation method has the researcher adopted? Explain how the estimation results were computed? Provide a detailed interpretation of the results reported in Table 4.2. Discuss the advantages and disadvantages of the two estimation methods used in Tables 4.1 and 4.2.

Question continued overleaf

Table 4.2

```
. poisson nprenatal mmarried foreign  mage  medu shortgap fbaby order

Iteration 0:   log likelihood = -12692.054
Iteration 1:   log likelihood = -12692.054

Poisson regression                              Number of obs     =      4,642
                                                LR chi2(7)        =     711.54
                                                Prob > chi2       =     0.0000
Log likelihood = -12692.054                     Pseudo R2         =     0.0273
```

nprenatal	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mmarried	.1456463	.0116131	12.54	0.000	.122885	.1684076
foreign	-.0374512	.0204958	-1.83	0.068	-.0776222	.0027198
mage	.0057634	.0010393	5.55	0.000	.0037264	.0078004
medu	.0113288	.0021391	5.30	0.000	.0071363	.0155213
shortgap	-.1250814	.0190538	-6.56	0.000	-.1624262	-.0877365
fbaby	-.019074	.0133913	-1.42	0.154	-.0453204	.0071724
order	-.0561452	.0065546	-8.57	0.000	-.068992	-.0432985
_cons	2.096118	.033365	62.82	0.000	2.030723	2.161512

(c) [15%] Compute the average partial effect of the mother's age on the number of prenatal visits using results in Tables 4.1 and 4.2, respectively. How do they compare?

(d) [15%]
The researcher also produced the results in Table 4.3 below. Explain how you would test for the null hypothesis that the coefficients for *fbaby* and *foreign* are jointly equal to zero using results in Tables 4.2 and 4.3. How does this test compare with the corresponding F-test presented in Table 4.1? Based on the test statistics you computed which estimation results would you employ? Provide details for your answer.

Question continued overleaf

Table 4.3

```
. poisson nprenatal mmarried mage  medu shortgap  order

Iteration 0:   log likelihood = -12694.794
Iteration 1:   log likelihood = -12694.794

Poisson regression                              Number of obs     =      4,642
                                                LR chi2(5)        =     706.06
                                                Prob > chi2       =     0.0000
Log likelihood = -12694.794                     Pseudo R2         =     0.0271
```

nprenatal	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mmarried	.1459826	.0115862	12.60	0.000	.1232741	.1686911
mage	.0058007	.0010358	5.60	0.000	.0037706	.0078307
medu	.0117107	.0021295	5.50	0.000	.007537	.0158844
shortgap	-.1223061	.0188008	-6.51	0.000	-.159155	-.0854571
order	-.0499453	.0048372	-10.33	0.000	-.0594261	-.0404646
_cons	2.067832	.028318	73.02	0.000	2.01233	2.123334

(e) [20%]

Discuss an empirical example (different from lectures and seminars' examples) of a Poisson model. Define the empirical model and provide details on the dependent and explanatory variables and on why these explanatory variables should be included.

END OF EXAMINATION PAPER