THE AUSTRALIAN NATIONAL UNIVERSITY

Second Semester Practice Final Examination – October, 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 9 pages (including cover page) with 5 exam questions plus an Appendix. 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 models:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u},$$

   where $\bar{Y} := \sum Y_i / n$ and similarly for $\bar{X}$.

   (a) [5 marks] Subtract both equations from each other and mathematically define an OLS estimator for β_1.

   (b) [10 marks] Derive the OLS estimator for β_1.

   (c) [5 marks] Using your result from part (b), propose an estimator for β_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) Consistency of an estimator is a property that does not depend on the size of the random sample drawn from the population.

   (b) The coefficient estimates of a linear probability model can be interpreted as marginal probabilities.

   (c) In order to deal with time trends in panel data estimation, each subject/entity needs to be observed for at least three time periods.

3. [20 marks]

   Evans and Schwab (1995) study the causal effect of attending Catholic high school on the probability of attending university. They consider the following equation:

$$\text{Uni}_i = \beta_0 + \beta_1 \text{CathHS}_i + \text{other factors} + u_i,$$

   where Uni_i is a dummy variable equal to one if person i attends university and zero otherwise, and CathHS_i is a dummy variable equal to one if person i attended Catholic high school and zero otherwise. (Note: 'other factors' is shortcut for explanatory variables on gender, race, family income and parental education.)

   (a) Why might CathHS_i be correlated with u_i?

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

   (c) Let CathRel_i be a dummy variable equal to 1 if a person is Catholic and zero otherwise. Discuss the two requirements needed for CathRel_i to be a valid instrumental variable. Which of these two requirements can be tested?

   (d) Is CathRel_i a convincing instrumental variable?

4. [20 marks] Consider the research question

Do workplace smoking bans induce smokers to quit?

To answer this question, you have available a data set with 10,000 observations on the following variables:

Variable name	Variable description
smoker	dummy equal 1 if person smokes
smkban	dummy equal 1 if smoke ban at workplace
age	age of person
hsdrop	dummy equal 1 if person is high school dropout
hsgrad	dummy equal 1 if person finished with high school degree
colsome	dummy equal 1 if person attained some college
colgrad	dummy equal 1 if person finished with college degree
black	dummy equal 1 if person is black
hispanic	dummy equal 1 if person is hispanic
female	dummy equal 1 if person is female

Make use of the attached Stata log-file to answer the following questions. Provide short yet comprehensive answers.

(a) Using a linear probability model which only includes *smoker* and *smkban*, how would you answer the above question?

(b) Now, adding all remaining explanatory variables to the linear probability model, how would you answer the above question?

(c) What could explain the change in the estimated effect of a smoke ban between parts (a) and (b)?

(d) Using a probit model, how would you answer the above question?

(e) What are the predicted smoking probabilities for a 20 year old white male who dropped out of high school when a smoke ban is absent? What if a smoke ban is in place? Study both the linear probability model and the probit model.

(f) How do the predicted probabilities from part (e) change when the person has a completed college degree instead? Study both the linear probability model and the probit model. Explain.

5. [25 marks]

Consider the following linear model for panel data:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it},$$

where $E[u_{it} X_{it}] = 0$ and $E[\alpha_i X_{it}] \neq 0$.

(a) Assuming $T = 2$ (that is, there are only two time periods in your panel data), mathematically define and derive the first difference estimator for β_1.

(b) Still assuming that $T = 2$, how would you obtain an estimate for β_0?
(Note: You can use your result from part (a).)

(c) Letting $T > 2$, if $E[\alpha_i X_{it}] = 0$, suggest a consistent estimator for β_1. Is it efficient?

Appendix follows below.

```
------------------------------------------------------------------------------
      name:  <unnamed>
       log:  ,Ä¶
  log type:  text
 opened on:  17 Oct 2014, 11:31:15

.
. // end preamble
.
. use "../Stata/Stock_data/Smoking.dta"

.
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      smoker |     10000       .2423    .4284963          0          1
      smkban |     10000       .6098    .4878194          0          1
         age |     10000     38.6932    12.11378         18         88
      hsdrop |     10000      .0912    .2879077          0          1
      hsgrad |     10000      .3266     .468993          0          1
-------------+--------------------------------------------------------
     colsome |     10000      .2802    .4491193          0          1
     colgrad |     10000      .1972    .3979045          0          1
       black |     10000      .0769     .266446          0          1
    hispanic |     10000      .1134     .317097          0          1
      female |     10000      .5637    .4959505          0          1

.
. generate agesq = age^2
```

```
. regress smoker smkban, robust

Linear regression                                      Number of obs =   10000
                                                       F(  1,  9998) =   75.06
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0078
                                                       Root MSE      =  .42684

------------------------------------------------------------------------------
             |               Robust
      smoker |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smkban |  -.0775583    .008952    -8.66   0.000    -.0951061   -.0600106
       _cons |   .2895951   .0072619    39.88   0.000     .2753604    .3038298
------------------------------------------------------------------------------


. regress smoker smkban female age agesq hsdrop hsgrad colsome colgrad black hi
> spanic, robust

Linear regression                                      Number of obs =   10000
                                                       F( 10,  9989) =   68.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0570
                                                       Root MSE      =  .41631

------------------------------------------------------------------------------
             |               Robust
      smoker |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smkban |  -.0472399   .0089661    -5.27   0.000    -.0648153   -.0296645
      female |  -.0332569   .0085683    -3.88   0.000    -.0500525   -.0164612
         age |   .0096744   .0018954     5.10   0.000      .005959    .0133898
       agesq |  -.0001318   .0000219    -6.02   0.000    -.0001747   -.0000889
      hsdrop |   .3227142   .0194885    16.56   0.000     .2845128    .3609156
      hsgrad |   .2327012   .0125903    18.48   0.000     .2080217    .2573807
     colsome |   .1642968   .0126248    13.01   0.000     .1395495     .189044
     colgrad |   .0447983   .0120438     3.72   0.000       .02119    .0684066
       black |  -.0275658   .0160785    -1.71   0.086    -.0590828    .0039513
    hispanic |  -.1048159   .0139748    -7.50   0.000    -.1322093   -.0774226
       _cons |  -.0141099   .0414228    -0.34   0.733    -.0953069    .0670872
------------------------------------------------------------------------------
```

```
. margins, at(smkban=(0 1) female=0 age=20 agesq=400 hsdrop=1 hsgrad=0 colsome=
> 0 colgrad=0 black=0 hispanic=0)

Adjusted predictions                              Number of obs   =      10000
Model VCE    : Robust

Expression   : Linear prediction, predict()

1._at        : smkban          =           0
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           1
               hsgrad          =           0
               colsome         =           0
               colgrad         =           0
               black           =           0
               hispanic        =           0

2._at        : smkban          =           1
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           1
               hsgrad          =           0
               colsome         =           0
               colgrad         =           0
               black           =           0
               hispanic        =           0
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .4493721   .0204352    21.99   0.000     .4093151    .4894292
          2  |   .4021323    .021035    19.12   0.000     .3608994    .4433651
------------------------------------------------------------------------------
```

```
. margins, at(smkban=(0 1) female=0 age=20 agesq=400 hsdrop=0 hsgrad=0 colsome=
> 0 colgrad=1 black=0 hispanic=0)

Adjusted predictions                              Number of obs   =      10000
Model VCE    : Robust

Expression   : Linear prediction, predict()

1._at        : smkban          =           0
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           0
               hsgrad          =           0
               colsome         =           0
               colgrad         =           1
               black           =           0
               hispanic        =           0

2._at        : smkban          =           1
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           0
               hsgrad          =           0
               colsome         =           0
               colgrad         =           1
               black           =           0
               hispanic        =           0
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1714562   .0146762    11.68   0.000     .1426879    .2002245
          2  |   .1242163   .0139756     8.89   0.000     .0968214    .1516112
------------------------------------------------------------------------------
```

```
. probit smoker smkban female age agesq hsdrop hsgrad colsome colgrad black his
> panic, robust

Iteration 0:   log pseudolikelihood = -5537.1662
Iteration 1:   log pseudolikelihood = -5238.7464
Iteration 2:   log pseudolikelihood =  -5235.868
Iteration 3:   log pseudolikelihood = -5235.8679

Probit regression                                 Number of obs   =      10000
                                                  Wald chi2(10)   =     542.93
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -5235.8679                 Pseudo R2       =     0.0544

------------------------------------------------------------------------------
             |               Robust
      smoker |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smkban |    -.15863   .0291099    -5.45   0.000    -.2156843   -.1015757
      female |  -.1117313    .028841    -3.87   0.000    -.1682585    -.055204
         age |   .0345114   .0068839     5.01   0.000     .0210192    .0480035
       agesq |  -.0004675   .0000826    -5.66   0.000    -.0006295   -.0003056
      hsdrop |    1.14161   .0729708    15.64   0.000     .9985902    1.284631
      hsgrad |   .8826708   .0603706    14.62   0.000     .7643467    1.000995
     colsome |   .6771192   .0614448    11.02   0.000     .5566896    .7975488
     colgrad |   .2346839   .0654163     3.59   0.000     .1064703    .3628976
       black |  -.0842789   .0534536    -1.58   0.115    -.1890461    .0204883
    hispanic |  -.3382743   .0493523    -6.85   0.000     -.435003   -.2415457
       _cons |  -1.734926   .1519802   -11.42   0.000    -2.032802   -1.437051
------------------------------------------------------------------------------
```

```
. margins, at(smkban=(0 1) female=0 age=20 agesq=400 hsdrop=1 hsgrad=0 colsome=
> 0 colgrad=0 black=0 hispanic=0)

Adjusted predictions                              Number of obs   =      10000
Model VCE    : Robust

Expression   : Pr(smoker), predict()

1._at        : smkban          =           0
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           1
               hsgrad          =           0
               colsome         =           0
               colgrad         =           0
               black           =           0
               hispanic        =           0

2._at        : smkban          =           1
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           1
               hsgrad          =           0
               colsome         =           0
               colgrad         =           0
               black           =           0
               hispanic        =           0
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    .464102   .0236973    19.58   0.000     .4176562    .5105479
          2  |    .401783    .024021    16.73   0.000     .3547027    .4488634
------------------------------------------------------------------------------
```

```
. margins, at(smkban=(0 1) female=0 age=20 agesq=400 hsdrop=0 hsgrad=0 colsome=
> 0 colgrad=1 black=0 hispanic=0)

Adjusted predictions                              Number of obs   =      10000
Model VCE    : Robust

Expression   : Pr(smoker), predict()

1._at        : smkban          =           0
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           0
               hsgrad          =           0
               colsome         =           0
               colgrad         =           1
               black           =           0
               hispanic        =           0

2._at        : smkban          =           1
               female          =           0
               age             =          20
               agesq           =         400
               hsdrop          =           0
               hsgrad          =           0
               colsome         =           0
               colgrad         =           1
               black           =           0
               hispanic        =           0
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1593747   .0130461    12.22   0.000     .1338049    .1849446
          2  |   .1239099   .0108087    11.46   0.000     .1027252    .1450946
------------------------------------------------------------------------------
. log close       // close log-file
      name:  <unnamed>
log type:  text
 closed on:  17 Oct 2014, 11:31:15
-------------------------------------------------------------------------------
```

End of Appendix.
End of Exam.