UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: ECON3150/4150 – Introductory Econometrics

Date of exam: Wednesday, May 18, 2016 **Grades are given: June 9, 2016**

Time for exam: 09:00 a.m. – 12.00 noon

The problem set covers 7 pages (incl. cover sheet)

Resources allowed:

- All written and printed resources – as well as calculator - is allowed

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Exam ECON3150/4150: Introductory Econometrics.
18 May 2016; 09:00h-12.00h.

This is an open book examination where all printed and written resources, in addition to a calculator, are allowed. If you are asked to derive something, give all intermediate steps. Do not answer questions with a "yes" or "no" only, but carefully motivate your answer.

Question 1

The head of department wants to investigate whether a mandatory term paper increases the probability that a student passes the exam. He has asked the econometrics teacher to set up an experiment where students are randomly assigned to a treatment group that has to write a mandatory term paper and to a control group that does not have to write a mandatory term paper. At the end of the course the teacher has a data set with information on 150 students about whether they passed the exam ($Passed_i$) and about whether they had written a mandatory term paper ($Termpaper_i = 1$) or not ($Termpaper_i = 0$). The teacher decides to estimate the following regression model by OLS

$$Passed_i = \beta_0 + \beta_1 \cdot Termpaper_i + u_i \tag{1}$$

and obtains the following estimation results

```
. regress Passed Termpaper, robust

Linear regression                               Number of obs     =        150
                                                  F(1, 148)        =       4.23
                                                  Prob > F         =     0.0415
                                                  R-squared        =     0.0278
                                                  Root MSE         =     .39706
```

Passed	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Termpaper	.1333333	.0648398	2.06	0.042	[redacted]	[redacted]
_cons	.7333333	.0514066	14.27	0.000	.6317475	.8349192

a) Give an interpretation, in words, of the two estimated coefficients.

b) Compute a 95 percent confidence interval for $\widehat{\beta}_1$.

c) The teacher wants to analyze whether the effect of the mandatory term paper depends on the age of the student. Describe in detail how you would extend model (1), such that you can test the null hypothesis that the effect of the mandatory term paper does not depend on the age of the student.

d) Because the dependent variable is binary the teacher decides to estimate a logit model and obtains the following estimation results.

```
. logit Passed Termpaper, robust

Iteration 0:   log pseudolikelihood =   -75.060364
Iteration 1:   log pseudolikelihood =   -72.978111
Iteration 2:   log pseudolikelihood =   -72.944236
Iteration 3:   log pseudolikelihood =   -72.944223
Iteration 4:   log pseudolikelihood =   -72.944223

Logistic regression                               Number of obs     =        150
                                                  Wald chi2(    1)      =       4.00
                                                  Prob > chi2       =     0.0454
Log pseudolikelihood =  -72.944223                  Pseudo R2         =     0.0282

------------------------------------------------------------------------------------
             |               Robust
      Passed |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------------
   Termpaper |   .8602013   .4298819     2.00   0.045      .0176483    1.702754
       _cons |   1.011601   .2619912     3.86   0.000      .4981075    1.525094
------------------------------------------------------------------------------------
```

What is the effect of writing a term paper on the predicted probability of passing the exam?

e) The teacher finds out that 180 students enrolled in the course and 90 were randomly assigned to the treatment group and the other 90 to the control group. The teacher doesn't observe exam results for all 180 students but only for 150 students. She suspects that some of the students in the treatment group didn't want to write a term paper and dropped out of the course. Explain whether in this case the OLS estimator of β_1 (from part a)) is a consistent estimator of the causal effect of writing a mandatory term paper on the probability of passing the exam.

f) The administration informs the teacher that 30 students got the flu and therefore did not take the exam. Explain whether in this case the OLS estimator of β_1 (from part a)) is a consistent estimator of the causal effect of writing a mandatory term paper on the probability of passing the exam.

Question 2

The federal government of the U.S. wants to know whether increasing seat belt usage reduces traffic fatalities. A government employee has data on the number of traffic fatalities per million of traffic miles ($fatalityrate_{it}$) and on the seat belt useage rate ($sb\,useage_{it}$) for 51 U.S. States, for the years 1990-1997. He estimates the following regression model by OLS

$$fatalityrate_{it} = \beta_0 + \beta_1 \cdot sb\,usage_{it} + u_{it} \tag{2}$$

and obtains the following estimation results

```
. regress fatalityrate sb_useage, robust

Linear regression                              Number of obs     =        408
                                                 F(1, 406)         =   [redacted]
                                                 Prob > F          =   [redacted]
                                                 R-squared         =     0.0309
                                                 Root MSE          =     .00448
```

fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
sb_useage	-.0060436	.0018963	[redacted]	[redacted]	[redacted]	[redacted]
_cons	.0220271	.0012039	18.30	0.000	.0196605	.0243937

a) Is the relation between seat belt usage and the traffic fatality rate significantly different from zero at a 5 percent significance level?

b) The government employee decides to include state fixed effects and obtains the following results.

```
. xtreg fatalityrate sb_useage, fe i(State) robust

Fixed-effects (within) regression              Number of obs      =        408
Group variable:  State                          Number of groups   =         51

R-sq:                                          Obs per group:
     within  =  0.3010                                        min =          8
     between =  0.0033                                        avg =        8.0
     overall =  0.0309                                        max =          8

                                               F(   1,50)          =      65.43
corr(u_i, Xb)  =  -0.2176                       Prob > F           =     0.0000

                               (Std. Err. adjusted for    51 clusters in State)
```

fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
sb_useage	-.0131094	.0016207	-8.09	0.000	-.0163646	-.0098542
_cons	.0261786	.0009522	27.49	0.000	.024266	.0280912
sigma_u	.00433259					
sigma_e	.00167444					
rho	.87004661	(fraction of variance due to u_i)				

Are the results different from the results without fixed effects? Explain why or why not.

c) Explain an alternative way to estimate the effect of seat belt usage on the traffic fatality rate while including state fixed effects.

d) The government employee is worried that there are omitted variables that vary within states over time that cause omitted variable bias. He decides to use an instrumental variable approach and to use the presence of a mandatory seat belt law as instrument. The variable $primary_{it}$ is a binary variable that equals 1 if in State i in year t there is a law which makes it possible for a police officer to stop a car and ticket the driver if the officer observes that someone in the car is not wearing a seat belt. He obtains the following first stage estimation results

```
. xtreg sb_useage primary, fe i(State)

Fixed-effects (within) regression               Number of obs      =       408
Group variable:  State                           Number of groups  =        51

R-sq:                                           Obs per group:
     within  =  0.0306                                        min =          8
     between =  0.3273                                        avg =        8.0
     overall =  0.2212                                        max =          8

                                                F(   1,356)          =  [redacted]
corr(u_i, Xb)  =  -0.4552                        Prob > F          =    0.0009

------------------------------------------------------------------------------
   sb_useage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     primary |   .2957143   .0882233  [redacted]  0.001       .12221    .4692186
       _cons |   .5418801   .0142222    38.10   0.000     .5139101    .5698502
-------------+----------------------------------------------------------------
     sigma_u |   .0992207
     sigma_e |  .08252533
         rho |   .5910923   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Is $primary_{it}$ a weak instrument?

e) The government employee decides to use two instruments. Next to the variable $primary_{it}$ the data set contains the variable $secondary_{it}$. The binary variable $secondary_{it}$ equals 1 if in State i in year t there is a law that makes it possible that a police officer can write a ticket if an occupant is not wearing a seat belt, but only if the police officer has another reason to stop the car. The government official estimates a first stage regression using both instruments and obtains the following first stage estimation results

```
. xtreg sb_useage primary secondary, fe i(State)

Fixed-effects (within) regression               Number of obs      =       408
Group variable:  State                           Number of groups  =        51

R-sq:                                           Obs per group:
     within  =  0.0372                                       min =          8
     between =  0.3319                                       avg =        8.0
     overall =  0.2259                                       max =          8

                                                F(   2,355)          =      6.86
corr(u_i, Xb)  =  -0.4735                        Prob > F          =    0.0012

------------------------------------------------------------------------------
   sb_useage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     primary |   .3055089   .0882704     3.46   0.001     .1319103    .4791074
   secondary |   .0035196   .0022574     1.56   0.120     -.00092     .0079591
       _cons |   .5380134   .0144087    37.34   0.000     .5096761    .5663506
-------------+----------------------------------------------------------------
     sigma_u |  .10001657
     sigma_e |  .08235998
         rho |  .59591494   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Are there weak instrument problems when using these two instrument? Explain whether it is better to use both $primary_{it}$ and $secondary_{it}$ as instruments or whether it is better to use only $primary_{it}$ as instrument for $sb\,usage_{it}$.

Question 3

Discuss whether each of the following statements is correct or not.

a) If we perform a J-test and we don't reject the null hypothesis we know that the instrument relevance condition holds.

b) If the $R^2 = 0.5$ the total sum of squares is twice as large as the sum of squared residuals.

c) When the sample size is small we can estimate the Root Mean Squared Forecast Error by the Standard Error of the Regression.

Question 4

A researcher wants to estimate the effect of a job training program (X_i) on wages in NOK (Y_i). Consider the following population regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ with $E[u_i|X_i] = 0$. The researcher observes X_i but by accident wages are measured in Euro's instead of NOK. Let α be NOK/Euro exchange rate. The researcher performs an OLS regression of Y_i^* (the observed wage rate in Euro's) on X_i. The researcher wants to know the causal effect of the job training program on wages in NOK, show whether the OLS estimator is a consistent or an inconsistent estimator of this causal effect.