UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON3150/ECON4150 – Introductory Econometrics**

Date of exam: Thursday, May 18, 2017 **Grades are given:** June 12, 2017

Time for exam: 09.00 a.m. – 12.00 noon

The problem set covers 6 pages (incl. cover sheet)

Resources allowed:

- Open book exam, where 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 2017; 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

A researcher wants to investigate whether class size affects long term labour market outcomes. He has a data set with labour market outcomes of 100 000 30-year-old individuals that live in Norway. The variable $employed_i$ equals one if an individual is employed and zero otherwise and the variable $class\,size_i$ equals the number of students in the class of the individual when he/she was in school.

a) The researcher decides to estimate the following regression model by OLS

$$employed_i = \beta_0 + \beta_1 \cdot class\ size_i + u_i \tag{1}$$

and obtains the following estimation results

```
. regress employed class_size, robust

Linear regression                               Number of obs     =        100,000
                                                F(1, 99998)       =        1201.81
                                                Prob > F          =         0.0000
                                                R-squared         =         0.0119
                                                Root MSE          =         .29821
```

employed	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-.0158525	.0004573	-34.67	0.000	-.0167488	-.0149562
_cons	1.264691	.0102837	122.98	0.000	1.244535	1.284847

Give an interpretation, in words, of the estimated coefficient on class size.

b) Compute a 99 percent confidence interval for $\widehat{\beta}_0$.

c) The researcher decides to estimate a logit model and obtains the following estimation results

```
. logit employed class_size, robust

Iteration 0:   log pseudolikelihood =   -32508.297
Iteration 1:   log pseudolikelihood =   -31919.121
Iteration 2:   log pseudolikelihood =   -31906.557
Iteration 3:   log pseudolikelihood =   -31906.551
Iteration 4:   log pseudolikelihood =   -31906.551  (backed up)

Logistic regression                               Number of obs     =        100,000
                                                  Wald chi2(    1)      =    1220.95
                                                  Prob > chi2       =        0.0000
Log pseudolikelihood =  -31906.551                  Pseudo R2         =        0.0185
```

employed	Coef.	Robust Std. Err.	[redacted]	[redacted]	[redacted]	[redacted]
class_size	-.1794293	.0051351	[redacted]	[redacted]	[redacted]	[redacted]
_cons	6.379385	.121634	52.45	0.000	6.140987	6.617783

Is the coefficient on $class\,size_i$ significantly different from zero at a 10 percent significance level?

d) Using the results from the logit model, what is the predicted change in the probability of being employed at age 30 that is associated with a reduction in class size from 25 to 20 students?

e) The data set also contains information about yearly income and the researcher decides to estimate the following regression model by OLS

$$ln(income_i) = \beta_0 + \beta_1 \cdot class\,size_i + u_i \tag{2}$$

and obtains the following estimation results

```
. regress ln_income class_size, robust

Linear regression                               Number of obs     =        100,000
                                                   F(1, 99998)       =    15530.38
                                                   Prob > F          =      0.0000
                                                   R-squared         =      0.1236
                                                   Root MSE          =      .05719
```

ln_income	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-.0103901	.0000834	-124.62	0.000	-.0105535	-.0102267
_cons	3.181756	.0019192	1657.88	0.000	3.177995	3.185518

Give an interpretation, in words, of the estimated coefficient on class size.

f) Name and explain one threat to internal validity that might apply when estimating equation (2) by OLS.

g) The researcher decides to include the years of completed education of both the father ($edu\ father_i$) and the mother ($edu\ mother_i$) as explanatory variables. He estimates the following equation by OLS

$$ln(income_i) = \beta_0 + \beta_1 \cdot class\ size_i + \beta_2 \cdot edu\ father_i + \beta_3 \cdot edu\ mother_i + \varepsilon_i \qquad (3)$$

and obtains the following estimation results

```
. regress ln_income class_size edu_father edu_mother, robust

Linear regression                               Number of obs     =        100,000
                                                F(3, 99996)       =      10633.90
                                                [redacted]        =      [redacted]
                                                R-squared         =        0.2417
                                                Root MSE          =        .05319
```

ln_income	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-.0007064	.0001123	-6.29	0.000	-.0009265	-.0004863
edu_father	.0098685	.000473	20.87	0.000	.0089415	.0107955
edu_mother	.0153749	.0003337	46.07	0.000	.0147208	.0160289
_cons	3.008313	.0023325	1289.75	0.000	3.003741	3.012884

Test the null hypothesis that the coefficient on class size and the coefficients on the years of completed education of both the father and the mother are equal to zero at a 5 percent significance level.

h) The researcher wants to analyze whether the effect of class size depends on the education of the parents. Describe in detail how you can test the null hypothesis that the effect of class size does not depend on the education of the mother and the father.

i) The researcher decides to use an instrumental variable approach. He thinks that class size is on average higher in Oslo than in the rest of Norway and therefore uses $Oslo_i$ as instrument for $class\ size_i$. $Oslo_i$ equals 1 when an individual lived in Oslo when he/she was in school and 0 otherwise. The researcher obtains the following estimation results.

```
. regress class_size Oslo, robust

Linear regression                               Number of obs     =        100,000
                                                F(1, 99998)       =      97275.24
                                                Prob > F          =        0.0000
                                                R-squared         =        0.4943
                                                Root MSE          =        1.4695
```

class_size	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Oslo	1.006355	.0032266	311.89	0.000	1.000031	1.012679
_cons	20.48096	.0093368	2193.57	0.000	20.46266	20.49926

Is $Oslo_i$ a weak instrument?

j) Do you think that, when using $Oslo_i$ as an instrument to estimate the effect of $class\ size_i$ on $ln(income_i)$, the instrument exogeneity condition holds ? Explain why or why not.

k) The researcher estimates the following two equations by OLS

$$ln(income_i) = \delta_0 + \delta_1 Oslo_i + \epsilon_i$$

$$class\ size_i = \pi_0 + \pi_1 Oslo_i + \upsilon_i$$

and obtains the following estimation results.

```
1 . regress ln_income Oslo, robust noheader
```

ln_income	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Oslo	-.0002051	.0001339	-1.53	0.126	-.0004675	.0000573
_cons	2.943244	.0003881	7583.53	0.000	2.942483	2.944004

```
2 . regress class_size Oslo, robust noheader
```

class_size	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Oslo	1.006355	.0032266	311.89	0.000	1.000031	1.012679
_cons	20.48096	.0093368	2193.57	0.000	20.46266	20.49926

What is the instrumental variable estimate of the effect of $class\ size_i$ on $ln(income_i)$?

Question 2

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

a) If the error terms are homoskedastic, hypothesis tests that are based on heteroskedasticity-robust standard errors are invalid.

b) An estimator is consistent if the expected value of the estimator equals the true value of the population parameter.

c) The p-value is the smallest significance level at which the null hypothesis can be rejected.

d) In a regression model with a only a constant term and no explanatory variables the R^2 equals zero.

Question 3

A researcher wants to estimate the effect of a job training program (T_i) on wages (W_i).

$$W_i = \beta_0 + \beta_1 T_i + u_i$$

A colleague of the researcher thinks that the wage affects participation the job training program:

$$T_i = \delta_0 + \delta_1 W_i + v_i$$

If the colleague is right, will the researcher obtain a consistent estimate of the causal effect of the job training program (T_i) on wages (W_i) when he estimates the following regression model $W_i = \beta_0 + \beta_1 T_i + u_i$ by OLS? Show why or why not (Hint: first derive $Cov(T_i, u_i)$ and assume that $Cov(u_i, v_i) = 0$).