UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: **ECON3150/4150 – Introductory Econometrics**

Date of exam: Thursday, June 2, 2016

Time for exam: 09:00 a.m. – 12:00 noon

The problem set covers 5 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.

Postponed Exam ECON3150/4150: Introductory Econometrics.
2 June 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 government of a developing country wants to know whether completing primary school has a positive effect on future income. A government employee has a data set with income ($income_i$) in US dollars of 1000 individuals living in this developing country in 2012. The data set also contains a variable $primary_i$ that equals 1 if individual i completed primary school and zero if individual i did not complete primary school. The government employee estimates the following equation by OLS

$$ln(income_i) = \beta_0 + \beta_1 primary_i + u_i$$

and obtains the following estimation results:

```
. regress ln_income primary, robust

Linear regression                               Number of obs     =          1,000
                                                F(1, 998)         =         952.76
                                                Prob > F          =         0.0000
                                                R-squared         =         0.4838
                                                Root MSE          =         .22116
```

ln_income	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
primary	.439359	.0142341	30.87	0.000	.4114268	.4672911
_cons	5.255183	.0090622	579.90	0.000	5.2374	5.272966

a) Interpret the estimated coefficient on $primary_i$.

b) Construct a 90 percent confidence interval for the coefficient on $primary_i$.

c) Explain whether we can interpret the estimated OLS coefficient on $primary_i$ as the causal effect of completing primary school on future income.

d) The researcher decides to estimate the effect of completing primary school using an instrumental variable approach. He has information on where each individual lived at the age of 10 and computes the distance to the nearest primary school. He uses the variable Z_i as instrumental variable which equals 1 if the primary school was less than 5 kilometers away and zero if the nearest primary school was more than 5 kilometers away. He obtains the following first stage estimation results.

```
. regress primary Z, robust

Linear regression                               Number of obs     =          1,000
                                                [redacted]
```

primary	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
Z	.228	.0299638	[redacted]	0.000	.1692008	.2867992
_cons	.272	.0199205	13.65	0.000	.2329092	.3110908

Do you think that the instrument relevance condition holds? Is Z a weak instrument?

e) Do you think that the instrument exogeneity condition holds?

f) The following table shows the averages of $ln(income_i)$ and $primary_i$ for those who lived less than 5 kilometers from the nearest primary school ($Z_i = 1$) and for those who lived more than 5 kilometers from the nearest primary school ($Z_i = 0$). Use the results in the table below to obtain the instrumental variable estimate of the effect of completing primary school on future income and interpret the magnitude of this instrumental variable estimate.

	$Z_i = 1$	$Z_i = 0$
$\widehat{E}\left[ln(income_i)\middle\|Z_i = x\right]$	5.446	5.403
$\widehat{E}\left[primary_i\middle\|Z_i = x\right]$	0.50	0.27

Question 2

A teacher wants to know whether study time affects the probability of passing an exam. She has a data set with 500 students that contains a variable $passed_{it}$ that equals 1 if a student passed the exam that took place at time t and a variable $studytime_{it}$ that contains the number of hours that student i spent on preparing for the exam that was taken at time t. The data set contains in total 5000 observations, with 500 students ($n = 500$) that each took 10 different exams ($T = 10$). The teacher estimates

$$passed_{it} = \beta_0 + \beta_1 studytime_{it} + u_{it}$$

by OLS and obtains the following estimation results.

```
. regress passed studytime, robust

Linear regression                               Number of obs     =          5,000
                                                F(1, 4998)        =       2121.16
                                                Prob > F          =       0.0000
                                                R-squared         =       0.2854
                                                Root MSE          =       .36611
```

passed	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
studytime	.0804765	.0017474	[redacted]	[redacted]	[redacted]	[redacted]
_cons	.3448519	.0127398	27.07	0.000	.3198762	.3698276

a) Test the null hypothesis that $\beta_1 = 0$ at a 1% significance level.

b) Interpret the two estimated coefficients.

c) Explain whether we can interpret the estimated OLS coefficient on $studytime_{it}$ as the causal effect of study time on the probability of passing an exam.

d) The teacher decides to augment the model with student fixed effects (α_i)

$$passed_{it} = \beta_0 + \beta_1 studytime_{it} + \alpha_i + \varepsilon_{it}$$

Explain how you could estimate this model.

e) The teacher is confident that by including student fixed effects the estimated coefficient on $studytime_{it}$ cannot suffer from omitted variable bias problems. Explain whether you agree with the teacher.

Question 3

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

a) If the sample size is large and we perform a t-test with a critical value equal to 1.96, the probability of rejecting the null hypothesis when it is true is 5%

b) The R^2 can never be equal to 0.

c) If we have measurement error in the explanatory variable(s) we can solve this by computing heteroskedasticity robust standard errors.

Question 4

Consider the following population regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$ with $E[u_i|X_i] = E[u_i|W_i] = 0$ and $E[W_i|X_i] = \alpha$. A researcher does not observe W_i and estimates the following regression model by OLS

$$Y_i = \beta_0 + \beta_1 X_i + v_i$$

Show whether the OLS estimator is a biased or unbiased estimator of β_1.