UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON3150/4150 – Introductory Econometrics, spring 2015**

Date of exam: Monday, May 11, 2015 **Grades are given: June 1, 2015**

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

The problem set covers 6 pages

Resources allowed:

- Open book exam, where all written and printed resources, as well as calculator, are allowed

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

Exam ECON4150: Introductory Econometrics.
11 May 2015; 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. In the grading, each sub-question will count for 1/12th of the total grade.

Question 1

One of the oldest questions in labor economics is the effect of union membership on wages. In order to investigate this question a researcher uses data on 5000 individuals observed in the years 2000-2010 with information on their hourly wage ($Wage_{it}$) in year t, their years of completed education ($Education_i$) and whether or not they are member of a union in year t ($Union_{it}$). The researcher estimates the following equation by OLS

$$ln(Wage_{it}) = \beta_0 + \beta_1 Union_{it} + \beta_2 Education_i + u_{it}$$

and obtains the following regression results

```
. regress ln_Wage Union Education, robust

Linear regression                               Number of obs =      55000
                                                F(  2, 54997) =       4.08
                                                [redacted]
                                                Root MSE      =     .89992
```

ln_Wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Union	.0237997	.0091306	2.61	0.009	.0059036	.0416957
Education	.0152618	.0132334	1.15	0.249	-.0106757	.0411992
_cons	2.346704	.0078609	298.53	0.000	2.331296	2.362111

a) Interpret the sign and magnitude of the coefficient on $Union_{it}$.

b) Test the null hypothesis that the coefficients on $Union_{it}$ and $Education_i$ are both equal to zero using a 1 percent significance level.

c) The researcher is mainly interested in the effect of union membership on wages. Describe one potential threat to the internal validity of the current regression results.

d) The researcher decides to augment the model by including individual fixed effects and obtains the following estimation results.

```
. xtreg ln_Wage Union Education, fe i(id) robust
```

```
Fixed-effects (within) regression               Number of obs      =         55000
Group variable:  id                             Number of groups   =          5000

R-sq:  within  =  0.0045                        Obs per group: min =            11
       between =  0.3087                                       avg =          11.0
       overall =  0.0375                                       max =            11

                                                F(1,4999)          =        245.03
corr(u_i, Xb)  =  0.2693                        Prob > F           =        0.0000

                                    (Std. Err. adjusted for    5000 clusters in id)
```

ln_Wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Union	.0232507	.0014853	15.65	0.000	.0203388	.0261626
Education	0	(omitted)				
_cons	2.070391	.00034	6089.28	0.000	2.069724	2.071057
sigma_u	.06462443					
sigma_e	.12884386					
rho	.20100616	(fraction of variance due to u_i)				

Explain why the variable $Education_i$ is omitted from the regression.

Question 2

A researcher wants to know whether a preschool program for immigrant children has an effect on their future educational attainment. In 1990 the government set up an experiment where 1000 (5-year old) immigrant children were randomly assigned to a treatment group (the preschool program) and to a control group (no preschool program), 25 years later the researcher uses data on these 1000 immigrant children (who are now 30 years old) to investigate the effect of participation in the preschool program ($preschool_i$) on years of completed education ($education_i$). The researcher decides to estimate the following regression model by OLS

$$education_i = \beta_0 + \beta_1 \cdot preschool_i + u_i \tag{1}$$

and obtains the following regression results

```
. regress education preschool, robust

Linear regression                               Number of obs =        1000
                                                F(  1,   998) =      169.11
                                                Prob > F      =      0.0000
                                                R-squared     =      0.1449
                                                Root MSE      =      1.1155
```

education	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
preschool	.9187161	.0706472	13.00	0.000	.7800821	1.05735
_cons	12.76565	.0512387	249.14	0.000	12.6651	12.8662

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

b) The researcher wants to analyze whether there is a difference in the effect of the preschool program between boys and girls. Describe in detail how you would extend model (1), such that you can test the null hypothesis that the effect of the preschool program on years of education does not depend on gender.

c) The researcher also wants to know whether participating in the preschool program increases the likelihood of having a job at age 30. The data set contains two additional variables; $employed_i$ which is equal to one if an individual has a job in 2015 and is zero otherwise and $girl_i$ which is equal to one for girls and zero for

boys. The researcher estimates the following regression model

$$employed_i = \beta_0 + \beta_1 \cdot preschool_i + \beta_2 \cdot girl_i + u_i \tag{2}$$

and obtains the following estimation results

```
. regress employed preschool girl, robust noheader
```

employed	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
preschool	.1371742	.0290887	4.72	0.000	.0800921	.1942564
girl	.3726898	.0289697	12.86	0.000	.3158413	.4295384
_cons	.2655838	.0251674	10.55	0.000	.2161966	.314971

On the basis of these estimation results, what is the change in the probability of being employed that is associated with participating in the preschool program *for boys*?

d) The researcher also estimates a logit model and obtains the following estimation results

```
. logit employed preschool girl, robust noheader

Iteration 0:   log pseudolikelihood =   -692.80914
Iteration 1:   log pseudolikelihood =   -613.31915
Iteration 2:   log pseudolikelihood =   -613.13556
Iteration 3:   log pseudolikelihood =   -613.13555
```

employed	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
preschool	.6451726	.1396817	4.62	0.000	.3714015	.9189436
girl	1.609523	.1398102	11.51	0.000	1.3355	1.883546
_cons	-1.030281	.1265701	-8.14	0.000	-1.278354	-.7822083

On the basis of these estimation results, what is the change in the probability of being employed that is associated with participating in the preschool program *for boys*?

e) Some children who were assigned to the control group did participate in the preschool program, because their parents managed to convince the preschool teacher to let their child participate. Do you think that the OLS estimator of β_1 in model (2) (estimated in part (c)) is a consistent estimator of the causal effect of participating

in the preschool program on the probability of being employed at age 30? Explain why or why not.

f) The researcher decides to use an instrumental variable approach to estimate the effect of participating in the preschool program on the probability of being employed at age 30. He uses the assignment to the treatment ($assignment_i = 1$) or control group ($assignment_i = 0$) as an instrument for whether or not a child participated in the preschool program. The researcher obtains the following first stage OLS estimates.

```
. regress preschool assignment girl, robust

Linear regression                               Number of obs =        1000
                                                F(  2,   997) =       30.38
                                                Prob > F      =      0.0000
                                                R-squared     =      0.0571
                                                Root MSE      =      .48567
```

preschool	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
assignment	.2349113	.0307472	7.64	0.000	.1745745	.295248
girl	-.0362907	.0308057	-1.18	0.239	-.0967422	.0241607
_cons	.4236373	.0267232	15.85	0.000	.3711972	.4760774

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

Question 3

Consider the following population regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ with $Cov(X_i, u_i) = 0$. The researcher observes Y_i but does not observe X_i, instead he observed a noisy measure $X_i^* = X_i + \varepsilon_i$, where $\varepsilon_i = \varepsilon$ (it is identical for all i). The researcher has a large sample with i.i.d observations on Y_i and X_i^* and estimates the following equation by OLS

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

a) What is $Cov(X_i^*, v_i)$?

b) Is the OLS estimator of β_1 consistent?