UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON3150/4150 – Introductory Econometrics**

Date of exam: Friday, May 25, 2018 **Grades are given:** June 14, 2018

Time for exam: 14.30 – 17.30

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

Resources allowed:

- Open book examination, where all written and printed resources, in addition to some calculators, are allowed. Calculators allowed for examination:

- Aurora HC106
- Casio FX-85EX

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.
25 May 2018; 14:30h-17.30h.

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 parents' participation in a welfare program increases the probability that their child will also participate in a welfare program as an adult. She has a data set with information on 10 000 children and their parents. The dependent variable $Wchild_i$ is a binary variable that equals 1 if the child receives welfare benefits when he is between 18 and 30 years old. The explanatory variable $Wparent_i$ equals 1 if the parents received welfare benefits when the child was between 12 and 18 years old.

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

$$Wchild_i = \beta_0 + \beta_1 \cdot Wparent_i + u_i \tag{1}$$

and obtains the following estimation result

```
. regress Wchild Wparent, robust

Linear regression                               Number of obs     =     10,000
                                                F(1, 9998)        =      10.00
                                                Prob > F          =     0.0016
                                                R-squared         =     0.0013
                                                Root MSE          =     .21783

------------------------------------------------------------------------------
             |               Robust
      Wchild |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Wparent |   .0219976   .0069551     3.16   0.002     .0083641    .0356311
       _cons |   .0467356   .0022875    20.43   0.000     .0422516    .0512195
------------------------------------------------------------------------------
```

Give an interpretation, in words, of the two estimated coefficients, $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

b) Is the coefficient on $Wparent_i$ significantly different from zero at a 1 percent significance level?

c) Do you think that the OLS estimator of β_1 is an unbiased estimator of the causal effect of parents' welfare participation on child's welfare participation as an adult? Explain why or why not.

d) The data set also includes the variable $edu\,parent_i$ which contains the average number of years of education completed by the parents. The variable $edu\,parent_i$ is negatively correlated with parents' welfare participation ($Wparent_i$) and has a negative effect on child's welfare participation ($Wchild_i$). Explain what will happen with the estimated coefficient on $Wparent_i$ when $edu\,parent_i$ is included as control variable in the OLS regression of $Wchild_i$ on $Wparent_i$?

e) Since the dependent variable $Wchild_i$ is a binary variable, the researcher decides to estimate a probit model and obtains the following estimation results

```
Probit regression                                 Number of obs     =     10,000
                                                  Wald chi2(2)      =     376.66
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -1716.3527                 Pseudo R2         =     0.1354
```

Wchild	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Wparent	.1411273	.0596087	2.37	0.018	.0242964	.2579582
edu_parent	-.2019093	.010536	-19.16	0.000	-.2225594	-.1812591
_cons	.6865118	.117618	5.84	0.000	.4559849	.9170388

What is the estimated effect of parents' welfare participation on the probability that the child participates in a welfare program, given that the parent has obtained 12 years of education?

f) Construct a 90 percent confidence interval around the coefficient on $Wparent_i$ in the probit regression model.

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

```
Logistic regression                               Number of obs     =     10,000
                                                  Wald chi2(2)      =     418.46
                                                  [redacted]        [redacted]
Log pseudolikelihood = -1718.0542                 Pseudo R2         =     0.1345
```

Wchild	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Wparent	.2732244	.1199389	2.28	0.023	.0381485	.5083004
edu_parent	-.4377319	.0217626	-20.11	0.000	-.4803858	-.3950781
_cons	2.040748	.2336515	8.73	0.000	1.5828	2.498697

```
. test Wparent edu_parent,[redacted]

 ( 1)  [Wchild]Wparent = 0
 ( 2)  [Wchild]edu_parent = 0

       F(  [redacted]  ) =  209.23
```

What is the estimated effect of parents' welfare participation on the probability that the child participates in a welfare program, given that the parent has obtained 12 years of education?

h) Test the null hypothesis that both the coefficients on $Wparent_i$ and $edu\,parent_i$ are zero using a 5 percent significance level.

i) A reform took place that made it more difficult to participate in a welfare program. This reform affected about half of the parents. The researcher decides to use this reform as an instrument for parent's welfare participation and estimates the following first stage regression by OLS

$$Wparent_i = \pi_0 + \pi_1 \cdot reform_i + \varepsilon_i$$

She obtains the following estimation results

```
. regress Wparent reform, robust

Linear regression                               Number of obs     =     10,000
                                                [redacted]        =     [redacted]
                                                [redacted]        =     [redacted]
                                                R-squared         =     0.0431
                                                Root MSE          =     .34778
```

Wparent	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
reform	-.1476753	.0069603	-21.22	0.000	-.1613189	-.1340316
_cons	.2223558	.005886	37.78	0.000	.210818	.2338935

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

j) The following table shows the fraction of children and the fraction of parents that participated in a welfare program separately for the children with parents that were affected by the reform ($reform_i = 1$) and for the children with parents that were not affected by the reform ($reform_i = 0$) . Use the results in the table below to obtain the instrumental variable estimate of the effect of $Wparent_i$ on $Wchild_i$.

	$reform_i = 1$	$reform_i = 0$
$\widehat{E}\left[Wchild_i \middle\| reform_i = x\right]$	0.049	0.050
$\widehat{E}\left[Wparent_i \middle\| reform_i = x\right]$	0.075	0.222

Question 2

A policy maker wants to know whether the inflow of immigrants affects the wages of native workers. The country is divided into two regions, region A and region B. There was a sudden influx of immigrants into region A but not in region B. The policy maker has information about wages of native workers in regions A and B both before and after the influx of immigrants. The following Stata output shows the averages of the logaritm of wages of native workers (*lnwage*):

```
. bys region time: sum lnwage
```

-> region = A, time = after

Variable	Obs	Mean	Std. Dev.	Min	Max
lnwage	**3,040**	**2.890215**	**.0553958**	**2.699678**	**3.059546**

-> region = A, time = before

Variable	Obs	Mean	Std. Dev.	Min	Max
lnwage	**2,942**	**2.994545**	**.0489649**	**2.797889**	**3.160841**

-> region = B, time = after

Variable	Obs	Mean	Std. Dev.	Min	Max
lnwage	**1,984**	**3.064744**	**.0463179**	**2.87966**	**3.239686**

-> region = B, time = before

Variable	Obs	Mean	Std. Dev.	Min	Max
lnwage	**2,034**	**3.090116**	**.0460191**	**2.919806**	**3.215129**

a) Compute the difference-in-differences estimate of the effect of the inflow of immigrants on the logarithm of wages of native workers.

b) Interpret the sign and magnitude of the difference-in-differences estimate obtained in part (a).

c) Explain the common trend assumption in the context of the application in this exercise.

Question 3

A researcher wants to estimate the effect of an additional year of schooling (S_i) on yearly earnings (E_i). Consider the following population regression model $E_i = \beta_0 + \beta_1 S_i + u_i$ with $Cov(S_i, u_i) = 0$. The researcher has a large data set with i.i.d observations on years of schooling S_i and on yearly earnings reported to the tax authority E_i^*. According to a colleague of the researcher, individuals under-report their earnings to the tax authority to reduce the amount of taxes they have to pay. This means that the observed taxable earnings differ from true earnings, more specifically $E_i^* = \gamma \cdot E_i$ with $0 < \gamma < 1$. The researcher wants to estimate the causal effect of an additional year of schooling on *true* earnings. He estimates the following equation by OLS

$$E_i^* = \beta_0 + \beta_1 S_i + v_i$$

a) Express v_i in terms of β_0, β_1, γ, S_i, u_i and show that $Cov(S_i, v_i) = (\gamma - 1)\beta_1 Var(S_i)$

b) Is the OLS estimator of β_1 a consistent estimator of the causal effect of an additional year of schooling on *true* earnings? Show why or why not.