UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

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

Date of exam: Thursday, June 4, 2015

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

The problem set covers 4 pages of text (not 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 ECON4150: Introductory Econometrics
Spring 2015

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 $1/12^{th}$ of the total grade.

Question 1

An economist wants to investigate the effect of family size on the educational attainment of children. He performs a regression of years of education (when the child has finished his or her education) $Education_i$ on the variable $More2kids$ which equals 1 if a child has at least two siblings (his mother had more than 2 children) and zero if the child has one sibling (his mother had 2 children). The economist estimates the following regression by OLS

$$Education_i = \beta_0 + \beta_1 More2kids_i + u_i$$

and obtains the following OLS estimates

```
. regress Education More2kids, r

Linear regression                                     Number of obs =      30000
                                                        F(  1, 29998) =  6759.41
                                                        Prob > F      =   0.0000
                                                        R-squared     =   0.1834
                                                        Root MSE      =   1.0208
```

Education	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
More2kids	-.9871441	.0120068	[redacted]	[redacted]	[redacted]	[redacted]
_cons	12.69732	.0076388	1662.20	0.000	12.68235	12.71229

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

b) Test the null hypothesis that the coefficient on $More2kids_i$ is equal to zero using a 5 percent significance level.

c) Describe one potential threat to the internal validity of the current regression results.

d) The economist thinks that he can obtain a consistent estimate of the effect family size on the educational attainment of children by performing 2SLS. He decides to use the incidence of twins at second birth as instrument for family size. He estimates the following first stage regression

$$More2kids_i = \pi_0 + \pi_1 Twins2_i + v_i$$

where $Twins2$ equals one if the mother had twins at second birth and zero otherwise. He obtains the following OLS estimates.

```
. regress More2kids Twins2, r

Linear regression                                      Number of obs =     30000
                                                          F(  1, 29998) =     79.40
                                                          Prob > F      =    0.0000
                                                          R-squared     =    0.0027
                                                          Root MSE      =    .48947

------------------------------------------------------------------------------
             |               Robust
   More2kids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Twins2 |   .1974412   .0221584     8.91   0.000     .1540097    .2408726
       _cons |   .3977492   .0028496   139.58   0.000     .3921638    .4033346
------------------------------------------------------------------------------
```

Do you think that the instrument relevance condition holds? Is $Twins2_i$ a weak instrument?

e) Do you think that $Twins2_i$ satisfies the instrument exogeneity condition? Explain why or why not.

f) The following table shows the averages of $Education_i$ and $More2kids_i$ for the individuals whose mother had twins at second birth ($Twins2_i = 1$) and for the individuals whose mother did not have twins at second birth ($Twins2_i = 0$). Use the results in the table below to obtain the instrumental variable estimate of the effect of $More2kids_i$ on years of education ($Education_i$).

	$Twins2_i = 1$	$Twins2_i = 0$
$\widehat{E}\left[Education_i \vert Twins_i = x\right]$	12.20	12.30
$\widehat{E}\left[More2kids_i \vert Twins_i = x\right]$	0.60	0.40

Question 2

A researcher decides to build a forecasting model for the annualized rate of inflation. She has quarterly data on the inflation rate (`inflation`). Let `d_inflation` be the change in the inflation rate from period $t-1$ to period t.

a) The researcher estimates the following AR(2) model

$$\triangle inflation_t = \beta_0 + \beta_1 \triangle inflation_{t-1} + \beta_2 \triangle inflation_{t-2} + u_t$$

and obtains the following estimation results

```
. regress d_inflation L1.d_inflation L2.d_inflation if tin(1962q1,1994q4)
```

Source	SS	df	MS
Model	73.3612493	2	36.6806246
Residual	347.325089	129	2.69244255
Total	420.686339	131	3.2113461

Number of obs =	132
F(2, 129) =	13.62
Prob > F =	0.0000
R-squared =	0.1744
Adj R-squared =	0.1616
Root MSE =	1.6409

d_inflation	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_inflation						
L1.	-.2873615	.0818998	-3.51	0.001	-.4494022	-.1253208
L2.	-.3692611	.0818178	-4.51	0.000	-.5311396	-.2073826
_cons	.0245487	.1428358	0.17	0.864	-.2580554	.3071529

Compute a 99% confidence interval for β_1.

b) The previous regression is based on $\triangle inflation_t$, because the researcher is worried that $inflation_t$ has a stochastic trend. The researcher estimates the following regression model

$$\triangle inflation_t = \gamma_0 + \delta \cdot inflation_{t-1} + \gamma_1 \triangle inflation_{t-1} + \gamma_2 \triangle inflation_{t-2} + u_t$$

and obtains the following estimation results.

```
. regress d_inflation L1.inflation L1.d_inflation L2.d_inflation if tin(1962q1,1994q4),
```

d_inflation	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
inflation						
L1.	-.0993034	.0474993	-2.09	0.039	-.1932889	-.0053179
d_inflation						
L1.	-.2307487	.0852646	-2.71	0.008	-.3994594	-.062038
L2.	-.3306665	.0828522	-3.99	0.000	-.4946038	-.1667293
_cons	.5055666	.2698531	1.87	0.063	-.0283838	1.039517

Use the estimation results to test for the presence of a stochastic trend in $inflation_t$. Use a 5% significance level.

c) The researcher wants to know how many lags of $\triangle inflation_t$ to include in the autoregression. She estimates $AR(p)$ models for $p = 1, 2, 3$ and 4 over the sample period 1962:1 to 1994:4 (first quarter of 1962 through the fourth quarter of 1994) and obtains the following sum of squared residuals (SSR) for each of the estimated models.

p	**1**	**2**	**3**	**4**
SSR	402.168	347.3251	331.970	331.383

Use the Akaike Information Criterion (AIC) to estimate the number of lags that should be included in the autoregression.

d) The researcher augments the AR(2) model of part (a) with four lagged values of the annualized unemployment rate. The researcher computes the Granger-causality F-statistic on the four lags of the unemployment rate and obtains the following results.

```
. test L1.unemployment=L2.unemployment=L3.unemployment=L4.unemploymen=0

 ( 1)  L.unemployment - L2.unemployment = 0
 ( 2)  L.unemployment - L3.unemployment = 0
 ( 3)  L.unemployment - L4.unemployment = 0
 ( 4)  L.unemployment = 0

       F( [redacted],    125) =    11.07
          [redacted]       [redacted]
```

Do the unemployment rates help to predict the inflation rate (at a 5% significance level)?

Question 3

Consider the following population regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^3 + u_i$ with $E\left[u_i|X_i, X_i^3\right] = 0$. A 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 $Cov\left(X_i, v_i\right)$?

b) Is the OLS estimator of β_1 consistent?