STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics I, Regression analysis, ST223G
Spring semester 2020

Written Examination in Econometrics I

Date	2020-04-28
Hour:	9.00-14.00
Examiner:	Jörgen Säve-Söderbergh
Allowed tools:	1) Textbook: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, Cengage. 2) Pocket calculator 3) Notes written in the text book are allowed.

- Note that no formula sheet is provided.
- Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.
- The maximum number of points for each problem is stated after each question. If not indicated otherwise, to obtain the maximum number of points on each problem, detailed and clear solutions are required. Answers may be given in English or Swedish.

Good luck!

1. This question concerns a data set that includes information on wages, education, parents' education and several other variables for 1230 working men in 1991. The model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

was estimated using R

```
c11out <- lm(educ~motheduc + fatheduc + abil + I(abil^2),data=htv)
```

The printout became

```
Coefficients:
            Estimate   Std. Error t value
(Intercept) 8.240226   0.287410  28.671
motheduc    0.190126   0.028096   6.767
fatheduc    0.108939   0.019601   5.558
abil        0.401462   0.030288  13.255
I(abil^2)   0.050599   0.008304   6.093

Residual standard error: 1.758 on 1225 degrees of freedom
Multiple R-squared:  0.4444,Adjusted R-squared:  0.4425
F-statistic: 244.9 on 4 and 1225 DF,  p-value: < 2.2e-16
```

Although the estimated model does not seem too bad, it was suggested that some test for functional misspecification should be completed. The following code was used

```
resettest(c11out, power = 2:3)
```

which resulted in

```
RESET test

data:  c11out
RESET = 4.411, df1 = 2, df2 = 1223, p-value = 0.01234
```

(a) Formulate the model that the RESET test specifies. (10 p)

(b) What null hypothesis is being tested? What is the alternative hypothesis (10 p)

(c) What conclusion can we draw from the prinout. Why? Use 5% significance level. Draw a conclusion based on the decision made using the RESET test. (10 p)

2. We are interested in estimating the return to education. We will do so through OLS and the Heckman model (Heckit). We will use the data material called MROZ which is part of 1976 Panel Study of Income Dynamics and consists of a sample of 753 married women aged 30-60, of whom 428 worked at some time in 1975. (This is the same data material that is used in chapter 17).

   We will focus on the Heckit estimation, but we will also report an OLS regression on the 428 women that had worked.

   The following was printed from R when estimating the Heckman model:

```
--------------------------------------------
Tobit 2 model (sample selection model)
2-step Heckman / heckit estimation
753 observations (325 censored and 428 observed)
19 free parameters (df = 735)
Probit selection equation:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.270077   0.508593   0.531  0.59556
educ         0.130905   0.025254   5.183 2.82e-07 ***
exper        0.123348   0.018716   6.590 8.37e-11 ***
I(exper^2)  -0.001887   0.000600  -3.145  0.00173 **
nwifeinc    -0.012024   0.004840  -2.484  0.01320 *
age         -0.052853   0.008477  -6.235 7.63e-10 ***
kidslt6     -0.868328   0.118522  -7.326 6.24e-13 ***
kidsge6      0.036005   0.043477   0.828  0.40786
Outcome equation:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.5602851  0.4587672  -1.221 0.222370
educ         0.1187171  0.0340507   3.486 0.000518 ***
```

```
exper         0.0598358  0.0336730   1.777 0.075987 .
I(exper^2)   -0.0010523  0.0006381  -1.649 0.099566 .
nwifeinc      0.0038434  0.0044919   0.856 0.392492
age          -0.0111580  0.0134792  -0.828 0.408054
kidslt6      -0.1880451  0.2308275  -0.815 0.415533
kidsge6      -0.0122255  0.0296063  -0.413 0.679775
Multiple R-Squared:0.1649,Adjusted R-Squared:0.1489
Error terms:
Estimate Std. Error t value Pr(>|t|)
invMillsRatio   0.2885     0.4636   0.622    0.534
sigma           0.6896         NA      NA       NA
rho             0.4183         NA      NA       NA
--------------------------------------------
```

a) Which test statistic is used for testing the null hypothesis $H_0 : \rho = 0$? (10 p)

b) Is there a significant sample selection correction? (10 p)

c) We are interested in the return to education, so let's compare the OLS and Heckit estimates What effect of the result from (b) can we see in table 1? (Found on the next page). (10 p)

Table 1: Comparison of OLS and Heckit estimates

	Dependent variable:	
	log(wage)	
	OLS	*selection*
	(1)	(2)
educ	0.100***	0.119***
	(0.015)	(0.034)
exper	0.041***	0.060*
	(0.013)	(0.034)
I(exper^2)	−0.001*	−0.001*
	(0.0004)	(0.001)
nwifeinc	0.006*	0.004
	(0.003)	(0.004)
age	−0.004	−0.011
	(0.005)	(0.013)
kidslt6	−0.056	−0.188
	(0.089)	(0.231)
kidsge6	−0.018	−0.012
	(0.028)	(0.030)
Observations	428	753
Note:	*p<0.1; **p<0.05; ***p<0.01	

3. Assume that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

is a model that you are interested in estimating. At a recent seminar you heard of the idea of testing the null hypothesis $H_0 : \beta_1 = 2\beta_2$. Show how we can test this null hypothesis H_0 by rewriting the model so that an ordinary t-test can be utilized. (20 p)

4. The test statistic for testing that q variables have parameters zero in our multiple regression model is given by

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)}.$$

This is equation [4.37] on page 141 in the 7th edition.

Show that this test statistic can be written

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1-R^2_{ur})/(n-k-1)}.$$

This is equation [4.41] on page 145 in the 7th edition.

Hint: Use the fact that $SSR_r = SST(1-R^2_r)$ *and* $SSR_{ur} = SST(1-R^2_{ur})$. (20 p)