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

Written Re-examination in Econometrics I

Date	2020-06-08
Hour:	15.00-20.00
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!

The Data Set

We will utilize the data set *FERTIL1*. The data in FERTIL1 are a pooled cross section on more than a thousand U.S. women for the even years between 1972 and 1984. Exactly, there $n = 1129$ women in the sample.

1.	year	72 to 84, even
2.	educ	years of schooling
3.	meduc	mother's education
4.	feduc	father's education
5.	age	in years
6.	kids	# children ever born
7.	black	= 1 if black
8.	east	= 1 if lived in east at 16
9.	northcen	= 1 if lived in nc at 16
10.	west	= 1 if lived in west at 16
11.	farm	= 1 if on farm at 16
12.	othrural	= 1 if other rural at 16
13.	town	= 1 if lived in town at 16
14.	smcity	= 1 if in small city at 16
15.	y74	= 1 if woman gave birth to a child in year = 74
16.	y76	
17.	y78	
18.	y80	
19.	y82	
20.	y84	
21.	agesq	age^2
22.	y74educ	y74*educ
23.	y76educ	
24.	y78educ	
25.	y80educ	
26.	y82educ	
27.	y84educ	

1. A researcher has eighteen observations on three variables that are interesting in her field of knowledge. One of them is the dependent variable y and the other two are one explanatory variable x and one dummy/indicator variable that has values in three categories. Thus, we will need two dummy variables to take care of the three different categories. Let d_1 denote the first one and d_2 the second (we don't need to know which category is the base category for this problem).

   (a) Assume that the researcher models the connection between the variables as

   $$y = \beta_0 + \beta_1 \, x + \beta_2 \, d_1 + \beta_3 \, d_2 + \varepsilon$$

   She finds that the total sum of squares (SST) equals 85260.44444 and that the residual sum of squares (SSR) equals 5078.71318. Use this information to test

   $$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{versus} \quad H_1 : \neg H_0.$$

   Use 5% significance level. (15 p)

   (b) The next step for the researcher is to estimate the model

   $$y = \beta_0 + \beta_1 \, x + \beta_2 \, d_1 + \beta_3 \, d_2 + \beta_4 \, xd_1 + \beta_5 \, xd_2 + \varepsilon,$$

   where she finds that the explained sum of squares (SSE) equals 82707.77658195. Use this information to test

   $$H_0 : \beta_4 = \beta_5 = 0 \quad \text{versus} \quad H_1 : \neg H_0.$$

   Use 5% significance level. (15 p)

2. In this question we bring example 13.1 in Wooldridge (page 428 in 7th) a little further. We are seeking to estimate a model for the total number of children ever born to a woman. What are the determinants of this? Since our dependent variable is a count variable we are interested in using Poisson regression.

   An analysis of the data material FERTIL1 was undertaken with R where we relate the number of kids to a number of explanatory variables.

```
Call:
glm(formula = kids ~ educ + age + I(age^2) + black + east + northcen +
    west + farm + othrural + town + smcity + y74 + y76 + y78 +
    y80 + y82 + y84, family = poisson, data = fertil1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.91598  -0.67884  -0.04123   0.55625   2.49302

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.0604626  1.2106974  -2.528 0.011476 *
educ        -0.0482027  0.0072302  -6.667 2.61e-11 ***
age          0.2044553  0.0547527   3.734 0.000188 ***
I(age^2)    -0.0022290  0.0006171  -3.612 0.000304 ***
black        0.3603475  0.0610748   5.900 3.63e-09 ***
east         0.0878001  0.0526729   1.667 0.095535 .
northcen     0.1417221  0.0475056   2.983 0.002852 **
west         0.0795427  0.0656991   1.211 0.226006
farm        -0.0148484  0.0575534  -0.258 0.796412
othrural    -0.0572939  0.0691574  -0.828 0.407412
town         0.0306807  0.0485793   0.632 0.527675
smcity       0.0741129  0.0615484   1.204 0.228535
y74          0.0932809  0.0630849   1.479 0.139232
y76         -0.0287888  0.0675828  -0.426 0.670123
y78         -0.0156856  0.0686754  -0.228 0.819334
y80         -0.0196524  0.0689821  -0.285 0.775727
y82         -0.1926076  0.0674991  -2.853 0.004324 **
y84         -0.2143735  0.0694641  -3.086 0.002028 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1331.1  on 1128  degrees of freedom
Residual deviance: 1184.3  on 1111  degrees of freedom
AIC: 4176.5
```

```
Number of Fisher Scoring iterations: 5
```

(a) Interpret the coefficient on $y82$. (10 p)

(b) What is the estimated percentage difference in fertility between a black woman and a nonblack woman, holding other factors fixed? (10 p)

(c) What is the estimated value of the standard deviation? Is there evidence of over- or underdispersion? (10 p)

3. Assume that you have decided that the model

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

fits your data material. An editor in a famous research magazine asks you to test the null hypothesis $H_0 : \beta_1 + \beta_2 = 1$. Show how this can be done.

Hint: Define $\theta_1 = \beta_1 + \beta_2 - 1$ and write a regression equation involving β_0, θ_1 and β_2 that allows you to directly obtain $\hat{\theta}_1$ and its standard error. (20 p)

4. Consider the simple linear regression $y = \beta_0 + \beta_1 x + u$, where the assumptions SLR.1 - SLR.4 are assumed to be true. Let the OLS estimator of β_1 be denoted (as usual) by $\hat{\beta}_1$. Show that

$$E\left(\hat{\beta}_1\right) = \beta_1$$

(20 p)