Econometrics I

Written examination

Monday June 10, 2019, 10:00 - 15:00

Examiner: Andreas Rosenblad, Department of Statistics, Stockholm University

Instructions

Allowed tools:

- Pocket calculator
- Text book: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach.* Cengage Learning, Boston.
- 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 given in the right margin. 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.

Solutions to the exam questions will be uploaded to Athena after the exam is finished. The corrected exams will be available at the student office of the Department of Statistics three weeks after the date of the exam.

Question 1 (40 points)

The R package **wooldridge** contains the data set WAGE1, which gives data on wages and associated characteristics for a number of individuals in the U.S., obtained from the 1976 Current Population Survey. We are interested in estimating the multiple linear regression model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \log(exper) + \beta_3 tenure + \beta_4 tenure^2 + \beta_5 female + \beta_6 married + u$$

where, for each individual, $wage$ gives the average hourly earnings in U.S. dollars, $educ$ and $exper$ give years of education and experience, respectively, $tenure$ gives the number of years working with the current employer, while $female$ and $married$ are dummy variables indicating whether the individual is a female or is married, respectively. The error term u is assumed to fulfill the usual requirements of normality, homoskedasticity, and independence. The R code and parts of the output for estimating this model using the sample of $n = 526$ observations with complete cases are given below.

```
> library(wooldridge)
> out.wage1 <- lm(log(wage) ~ educ + log(exper) + tenure + I(tenure^2)
+ female + married, data = wage1)
> summary(out.wage1)

Call:
lm(formula = log(wage) ~ educ + log(exper) + tenure + I(tenure^2) +
    female + married, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.82739 -0.25647 -0.02142  0.24139  1.18864

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.3253461  0.1038424   3.133  0.00183
educ         0.0859426  0.0066289
log(exper)   0.0892101  0.0212931   4.190 3.28e-05
tenure       0.0324068  0.0068035
I(tenure^2) -0.0006929  0.0002327
female      -0.2938684  0.0363661  -8.081 4.55e-15
married      0.0561446  0.0413728
---

Residual standard error: 0.4021 on 519 degrees of freedom
Multiple R-squared:              Adjusted R-squared:  0.4276
F-statistic: 66.37 on 6 and 519 DF,  p-value: < 2.2e-16
```

Note that the standard error of regression (SER) is called residual standard error in the R output.

(a) What is the value of R^2 for this model? (8)

(b) Are those who are married having higher average hourly earnings than those who are not married, ceteris paribus? State the formal null and two-sided alternative hypotheses for testing this research question using the estimated model, and perform the test using a significance level of 5%. What is your conclusion? (8)

(c) Approximately, how many percent higher is a person's average hourly earnings estimated to be if his/hers experience increases with 1%, ceteris paribus? (6)

(d) Approximately, how many percent lower average hourly earnings are women estimated to have compared to men, ceteris paribus? (6)

(e) After how many years of working with the current employer is it estimated that the average hourly earnings of an individual start to decrease? (6)

(f) A researcher is interested in testing the null hypothesis that years of education has no effect on average hourly earnings against the alternative hypothesis that it has a positive effect. State the formal null and alternative hypotheses for testing this research question using the estimated model, and perform the test using a significance level of 1%. What is your conclusion? (6)

Question 2 (12 points)

Suppose that the results in the Swedish General Election 2018 (SGE2018) for the Swedish Social Democratic Party (S) and the Swedish Moderate Party (M) are determined by the two-equation system

$$vote_S = \alpha_0 + \alpha_1 vote_M + \alpha_2 education + \alpha_3 employ + u_1$$
$$vote_M = \beta_0 + \beta_1 vote_S + \beta_2 education + \beta_3 income + u_2$$

where $vote_S$ and $vote_M$ are the percentage of votes in a constituency obtained by S and M, respectively. The variables *education*, *employ*, and *income* are assumed to be exogenous, with α_1, α_2, α_3, β_1, β_2, and β_3 all being different from zero.

(a) Which of these two equations is/are identified? (8)

(b) What type of regression method is appropriate for estimating this model? (4)

Question 3 (12 points)

Suppose that one wants to test whether students who attend charter schools get higher grades than students attending public schools. Let *grade* denote the average final grade a student gets when finishing upper secondary school, *chartschool* denote a dummy variable indicating whether the student attends a charter school, *totinc* denote the total income of the student's parents, *meduc* denote the education level of the student's mother, and *feduc* denote the education level of the student's father. An equation relating *grade* to *chartschool*, *totinc*, *meduc*, and *feduc* is

$$grade = \beta_0 + \beta_1 chartschool + \beta_2 totinc + \beta_3 meduc + \beta_4 feduc + u_1$$

We are interested in using *numchs*, the number of charter schools in a municipality, as an instrumental variable (IV) for *chartschool*.

(a) Write the reduced form equation for *chartschool*. (6)

(b) To be a valid IV, *numchs* must be partially correlated with *chartschool*. Explain how this assumption can be tested from the reduced form equation for *chartschool*. (6)

Question 4 (36 points)

Give the correct answer for the following multiple-choice questions. No motivation is needed.

(a) If a simultaneous equations model is represented by the two equations (3)

$$y_1 = \gamma_1 + \alpha_1 y_2 + \beta_1 z_1 + u_1$$
$$y_2 = \gamma_2 + \alpha_2 y_1 + \beta_2 z_2 + u_2$$

what are y_1 and y_2 called?

A. Structural parameters.
B. Reduced form parameters.
C. Exogenous variables.
D. Endogenous variables.

(b) In the equation (3)

$$y = \beta_0 + \beta_1 x + u$$

which one of the following four formulas gives the estimated value of β_0?

A. $\bar{y} - \beta_1 \bar{x}$
B. $\bar{y} - \hat{\beta}\bar{x}$
C. $y_i - \hat{\beta}_1 x_i$
D. $\bar{y} - \hat{\beta}_1 \bar{x}$

(c) Using a significance level $\alpha = 0.05$, what would you conclude about a regression model if the Breusch-Pagan resulted in a P-value > 0.05? (3)

A. The null-hypothesis of heteroskedasticity is not rejected.
B. The null-hypothesis of homoskedasticity is not rejected.
C. The null-hypothesis of non-homoskedasticity is not rejected.
D. The null-hypothesis of homoskedasticity is rejected.

(d) For the regression equation (3)

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \beta_3 x_3 + u$$

what does β_1 imply?

A. β_1 measure the ceteris paribus effect of y on x_1
B. β_1 measure the ceteris paribus effect of x_1 on x_2
C. β_1 measure the ceteris paribus effect of x_2 on x_1
D. β_1 measure the ceteris paribus effect of $x_1 + x_2$ on y

(e) Which one of the following four statements describes an advantage of using $\bar{R}^2$over R^2? (3)

A. $\bar{R}^2$ adjusts for the bias of R^2 in multiple linear regression models.

B. $\bar{R}^2$ but not R^2 may be used to choose between nonnested models.

C. $\bar{R}^2$ is easier to calculate than R^2 for multiple linear regression models.

D. $\bar{R}^2$ but not R^2 can be calculated for models with logarithmic functions.

(f) A group of individuals becomes unemployed on January 1, 2018 and are followed either until they get a new job or until end of follow-up occurs on January 1, 2019, whichever cames first. On January 1, 2019 some but not all of the individuals have got a new job. The outcome of interest is the time until an individual gets a new job. What time of regression model is the most appropriate to use in this case? (3)

A. A censored regression model.

B. A truncated regression model.

C. A Poisson regression model.

D. A logistic regression model.

(g) If $\hat{\beta}_j$ is an unbiased and consistent estimator of β_j, which one of the following four statements is true when the sample size tends to infinity? (3)

A. The distribution of $\hat{\beta}_j$ collapses to the single point x_j.

B. The distribution of $\hat{\beta}_j$ collapses to the single point $\bar{x}_j$.

C. The distribution of $\hat{\beta}_j$ collapses to the single point zero.

D. The distribution of $\hat{\beta}_j$ collapses to the single point β_j.

(h) Suppose that z is an instrument for x in the simple linear regression model (3)

$$y = \beta_0 + \beta_1 x + u$$

For which of the following four situations is z said to be a poor instrument for x?

A. If there is low correlation between z and u.

B. If there is high correlation between z and u.

C. If there is low correlation between z and x.

D. If there is high correlation between z and x.

(i) Which one of the following four terms may be used as a synonym for non-experimental data? (3)

A. Cross-sectional data.

B. Observational data.

C. Time series data.

D. Panel data.

(j) What is a sample selection method that is based on the dependent variable in a regression model called? (3)

A. Exogeneous sample selection.
B. Endogeneous sample selection.
C. Stratified sample selection.
D. Non-endogeneous sample selection.

(k) For the equation (3)

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

what does the null hypothesis $H_0 : \beta_2 = 0$ state?

A. x_2 has no effect on the expected value of β_2.
B. x_2 has no effect on the expected value of y.
C. β_2 has no effect on the expected value of y.
D. y has no effect on the expected value of x_2.

(l) In a linear model of the annual savings of an individual as a function of his/hers education and annual income, education is a categorical variable measuring the individual's highest achieved education level, divided into the four categories No education, Mandatory school, Secondary school, and College/University. How many dummy variables for highest achieved education level should be included in the regression model? (3)

A. 3
B. 1
C. 5
D. 4

Department of Statistics

Stockholms universitet

Correction sheet

Date: 10/06/2019

Room: Uggleviksalen

Exam: Econometrics 1

Course: Econometrics

Anonymous code: 0023-JUW

☒ I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	X	X						3
Teacher's notes 40	12	10	30						

MW

Points	Grade	Teacher's sign.
92	A	AR

SU, DEPARTMENT OF STATISTICS

Room: Ugglevikssalen Anonymous code: 0023-JUW Sheet number: 1/3

Question 1.

a. We know that $n = 526$ $(\rightarrow n-1 = 525)$, $n-k-1 = 519$ and $\bar{R}^2 = 0{,}4276$. We can obtain R^2 from the relationship between R^2 and $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

$$\Leftrightarrow \quad R^2 = 1 - \frac{(1-\bar{R}^2)(n-k-1)}{n-1}$$

$$= 1 - \frac{0{,}5724 \times 519}{525}$$

$$= \underline{0{,}4341}.$$

/8

b. We want to know if those who are married have higher average hourly earnings than those who are not.

We test $H_0: \beta_6 = 0$ vs $H_A: \beta_6 \neq 0$.

$$t = \frac{\hat{\beta}_6}{se(\hat{\beta}_6)} \sim t_{0{,}025,\infty} = 1{,}96 \text{ if } H_0 \text{ is true}.$$

We reject H_0 if $|t_{obs}| > 1{,}96$.

$$t_{obs} = \frac{0{,}0561446}{0{,}0413728} \simeq \underline{1{,}36}$$

We cannot reject the null hypothesis that being married has no impact on the wage

/8

c: The log-log model implies $\% \Delta y = \beta_2 \% \Delta x$.
Therefore, if a person's experience increases with 1%, their wage is expected to increase by 0,089 %, holding all other factors constant.

/6

d. The log-level model implies $\% \Delta y = 100 \beta_3 \Delta x$.
Therefore, women are expected to have 29,39 % lower average hourly earning than men (other factors held constant).

/6

e. To know how many years of tenure it takes on average for the wage of an individual to start decreasing, we calculate the turning point of the quadratic model.

$$x^* = \left| \frac{\hat{\beta}_3}{2\hat{\beta}_4} \right| = \left| \frac{0{,}0324068}{-2 \times 0{,}0006929} \right| \simeq 23{,}4.$$

The wage starts to decrease after 23 years on average, holding all other factors constant.

/6

f. We want to know if education has a positive impact on the wage:

We test: $H_0 : \beta_1 \neq 0$ vs $H_A : \beta_1 > 0$.

$$t_{obs} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t_{0{,}01, \infty} = 2{,}326 \text{ if } H_0 \text{ is true.}$$

We reject H_0 if $t_{obs} > 2{,}326$.

$$t_{obs} = \frac{0{,}0859426}{0{,}0066289} \simeq 12{,}96.$$

Hence we can reject the null hypothesis, and we conclude that education has a positive, significant impact on wage.

/6

SU, DEPARTMENT OF STATISTICS

Room: Vg Anonymous code: 0023-JUW Sheet number: 2/3

Question 2.

a. In a two-equation system, the first-equation is identified "if and only if the second equation contains at least one exogenous variable (with a non-zero coefficient), that is excluded from the first equation" (Wooldridge, p. 541).

* We see that this condition holds for the first equation (where $votes_s$ is the dependent variable) since the other equation contains the exogenous variable income (income has non-zero coefficient).

* The rank condition also holds for the second equation ($votes_H$ is the dependent variable) since the first equation contains the variable employ (also with non-zero coefficient)

Conclusion: Both equations are identified.

/8

b. Since we have identified both equations in the model, we can estimate them by two stage least squares (2SLS), where the instrumental variables are the exogenous variables.

/4

Question 3.

a. Assuming that numchs is exogenous ($Cov(numchs, u_1) = 0$) we can write the reduced form equation for chartschool in terms of numchs.

$$chartschool = \pi_0 + \pi_1 numchs + v_1.$$

(We don't know if the other variables in the model are exogenous so we do not add them to the reduced form equation).

/4

b. In addition to being exogenous, the instrument ~~be~~ must be relevant, that is $Cov(chartschool, numchs) \neq 0$.
This implies that the coefficient π_1 in the reduced form equation should be significantly different from 0.

* Hence we test $H_0: \pi_1 = 0$ vs $H_A: \pi_1 \neq 0$ and we compute the t-statistic based on the values of $\hat{\pi}_1$ and $se(\hat{\pi}_1)$. If the observed t-value is larger than the critical value, we reject H_0.

/6

SU, DEPARTMENT OF STATISTICS

Room: Ug Anonymous code: 0023-JUW Sheet number: 3/3

Question 4.

a. D. /3
b. D. /3
c. B. /3
d. D. /3
e. A /0
f. C /0
g. D. /3
h. C /3
i. B. /3
j. B /3
k. B /3
l. A. /3

Σ 30