STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics I, Regression analysis, ST223G
Autumn Semester 2020

Written Re-examination in Econometrics I

Date	2021-01-11
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.

For questions about the content of the exam, contact the course coordinator on jorgen.save-soderbergh@stat.su.se. Incoming e-mail questions are answered between 10.00 and 11.00 during the exam.

If the course coordinator needs to send out information to all students during the exam, this is done to your registered email address. Therefore, check your email during the exam.

Please note that practical help is only available during the first hour of the exam by email expedition@stat.su.se. Carefully read the enclosed instructions for exam submission. There you find all the necessary information about submission, anonymous code, etc. If you, despite the instructions have problems submitting the exam, email the exam to tenta@stat.su.se. However, this is only done in exceptional cases.

Good luck!

The Data Set

This data set concern the performance of top 500 companies in the U.S. economy and consists of twenty observations.

- 1990 rank = company's market value (share price on March 16, 1990 multiplied by available common shares outstanding)
- 1989 rank = company's market value 1989
- P-E ratio = price-to-earnings ratio, based on 1989 earnings and March 16, 1990 share price
- Yield = Annual dividend rate as a percentage of March 16, 1990 share price

1. A financial analyst wishes to analyze the data set concerning the performance of top 500 companies in the U.S. economy. After some preliminary analysis the following model is estimated

$$\ln(\text{Yield}) = \beta_0 + \beta_1 \text{P-E ratio} + \beta_2 1990 \text{ rank} + \varepsilon.$$

   (a) From the estimation above, it is found that the total sum of squares (SST) equals 48.0803 and that the explained sum of squares (SSE) equals 28.62285. Use this information to test

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

   Use 5% significance level. (15 p)

   (b) Another model is tried and estimated

$$\ln(\text{Yield}) = \beta_0 + \beta_1 \text{P-E ratio} + \beta_2 1990 \text{ rank} + \beta_3 1989 \text{ rank} + \varepsilon,$$

   where the analyst finds that the explained sum of squares (SSE) equals 29.97726. Use this information to test

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

   Use 5% significance level. (15 p)

2. Our financial analyst wishes further to check whether the residuals are heteroskedastic or not. The analyst collected the residuals $\hat{u}$ from the model in b) of question one. Then the analyst squared the residuals $\hat{u}^2$ and fitted the same model as in b), but used the squared residuals as the dependent variable instead of $\ln(\text{Yield})$. From this regression the analyst saved $R^2_{\hat{u}^2}$ which turned out to be 0.07472829. The analyst thought of using this determination coefficient in conducting tests about heteroskedasticity.

   (a) Formulate both the null hypothesis and the alternative hypothesis. Give some intuition behind this formulation. (5 p)

   (b) Compute the F test statistic. Formulate the critical region. What is the result of the F test? (10 p)

   (c) Compute the LM test statistic. Formulate the critical region. What is the result of the LM test? (10 p)

   (d) What is the overall conclusion from the use of these tests? (5 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. Your boss asks you to test the null hypothesis $H_0 : \beta_1 + a\beta_2 = 0$, where a is a constant that you can choose as you like (you must of course let a assume a certain number before you can actually test this null hypothesis, the upside of this is that you in this way can treat several hypotheses at the the samt time). Show how this can be done.

Hint: Define $\theta = \beta_1 + a\beta_2$ and write a regression equation involving β_0, θ and β_2 that allows you to directly obtain $\hat{\theta}$ 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)