Stockholm University
Department of Statistics
Oskar Gustafsson

HOME EXAM ECONOMETRICS I
2021-06-04

Time for examination: 9.00-14.00. The exam shall be submitted electronically via the department's web site no later than **15.00**. The system does not allow submission after deadline. Therefore, start the submission well in advance. The last hour of the exam time is intended for arranging the electronic submission.

All necessary information about submission, anonymous code, extended writing time, etc. can be found in a separate file. 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. Exams sent in by email after deadline will not be corrected.

For questions regarding the submission, email to: expedition@stat.su.se. Practical help is only available during the first hour of the exam.

For questions regarding the content of the exam, email to:
edgar.bueno@stat.su.se.
Incoming email questions are answered continuously 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.

Note: The exam should be written individually. All types of collaborations and/or help from others are strictly forbidden. Suspected cheating is reported to the Disciplinary Board and can lead to suspension

Allowed tools: Pocket calculator, computer, course books and lecture notes.

The exam consists of 4 independent problems. Well motivated and clear solutions are required for full scoring.

Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description. Answers may be given in English or Swedish.

Good luck!

If not indicated otherwise, the disturbance terms u_i in the models are supposed to fulfill the usual requirements of normality, homoscedasticity and independence.

1. (25p) The average yearly price for a particular type of television set during the last six years (2009–2014), coded as $1, \ldots, 6$) have been:

Year X_i	Average price (SEK) Y_i
1	10510
2	9580
3	8450
4	7530
5	6420
6	5500

Summary statistics: $\sum x_i = 21$, $\sum x_i^2 = 91$, $\sum y_i = 47\,990$, $\sum y_i^2 = 401\,806\,300$, $\sum x_i y_i = 150\,240$

   (a) Suggest (with motivation) and write down an appropriate regression model with Y as dependent variable based on this data set.

   (b) Compute a point and interval estimate of the expected value for Y in the year 2015 (confidence level 95%).

   (c) Also without performing it, suggest a test for heteroscedasticity. Make a short comment on the validity of the distributional assumption of the test statistic in this case.

   (d) Suppose we believe that the variance $V(Y_i) = \frac{\sigma^2}{x_i}$, implying heteroscedasticity. (This could be motivated by assuming that the yearly variation decreases with time.) Using this information, rewrite your original model, so that we obtain a new model which is homoscedastic.

2. (25p) Suppose we have the following model based on quarterly data from the first quarter of the year 2013 to the last in 2020:

$$Y_t = \beta_1 + \beta_2 t + \beta_3 D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + u_t, \tag{1}$$

where Y_t is the yield for a certain stock at time point t, t is coded as $1, 2, \ldots, 32$ and

$$D_{1i} = \begin{cases} 1 & \text{for quarter 1} \\ 0 & \text{otherwise} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{for quarter 2} \\ 0 & \text{otherwise} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{for quarter 3} \\ 0 & \text{otherwise} \end{cases}$$

Result:

`The regression equation is`

$y = 1.04 + 0.0368t - 1.04D_1 - 0.726D_2 - 0.911D_3$

ANOVA		
Source	Sum of Squares	df
Regression (SSE)	9.7685	4
Residuals (SSR)	0.8412	27
Total (SST)	10.6097	31
Coefficients Table		
Variable	Coefficient	SE($\hat{\beta}_i$)
Constant	1.04190	0.08746
t	0.036804	0.003404
D_1	-1.04459	0.08884
D_2	-0.72639	0.08851
D_3	-0.91132	0.08832

We also have the result using the following model:

$$Y_t = \beta_1' + \beta_2' t + u_t'$$

`The regression equation is`

$y = 0.300 + 0.0411t$

ANOVA		
Source	Sum of Squares	df
Regression (SSE)	4.6145	1
Residuals (SSR)	5.9952	30
Total (SST)	10.6097	31
Coefficients Table		
Variable	Coefficient	SE($\hat{\beta}_i'$)
Constant	0.300	0.1618
t	0.041128	0.008559

(a) How do you interpret β_1 in model (1)?

(b) Using model (1), compute predicted values for Y for all quarters in 2021.

(c) A friend of yours claims that the coefficient of determination is here 92.1% and that this means that we get a VIF-values which is 12.6, implying that we have multicollinearity problems. What do you reply to your friend? (If we altogether disregard the unlikely event that any friend of yours would be interested in this issue.)

3. (25p) A friend of yours who is studying economics is confused, because of the result of his regression analysis. He has used two X-variables, measuring income and wealth and has consumption as Y-variable.

   Result:

   `The regression equation is`

   $y = 24.8 + 0.94X_2 - 0.042X_3$

ANOVA		
Source	Sum of Squares	df
Regression (SSE)	8565.5541	2
Residuals (SSR)	324.4459	7
Total (SST)	8890	9
Coefficients Table		
Variable	Coefficient	SE($\hat{\beta}_i$)
Constant	24.7747	6.7525
X_2	0.9415	0.8229
X_3	-0.0424	0.0807

   $R^2 = 96.3\%$, but according to his t-tests, none of the X-variables should be included in the model, even though each X-variable is highly correlated with Y. Furthermore, he does not like the fact that if one increases the value of the wealth-variable X_3, $\hat{Y}$ decreases.

   Help your friend to explain these effects and show him by a test that (at least) one the explanatory variables should be included. Also try to explain to him what could be done instead of using this particular model.

4. (25p) Suppose we have a simple linear regression model for a monetary variable Y, using some quantitative X-variable:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

   Person A och and person B use the same model and the same "X-data". However, A measures Y in millions of kronor and B measures Y in thousands of kronor.

   The result was for A: $\hat{\beta}_1 = 0.031$, $\hat{\beta}_2 = 0.056$, $\hat{\sigma} = 0.072$.

   (a) What should the resulting estimates be for B?

   (b) However, B did not get the estimates that we expected and in particular B got that $\hat{\sigma} = 83.0$. What could be the reason for this? Try to formulate a model for B that "explains" this result and also compute estimates of additional parameters (one or more) if possible.

(c) Assume now that we should have added another X-variable which is highly correlated with Y (but not with the first X-variable). Mention with short explanation one possible negative consequence of omitting an important X-variable in a linear regression model.