Supplementary Written Examination in Econometrics (B2)

Fall 2016

2016-12-07 08.00-12.00

Bergsbrunnagatan 12, room 1.

Lars Forsberg, Department of Statistics, Uppsala University

Allowed means of assistance:

1. Pen or **pencil** (recommended) and eraser
2. **Calculator**,
   (a) 'programmable' calculator, e.g. calculator with graphing functions is OK.
   (b) Calculators with blue-tooth are not allowed.
   (c) Calculators with access to internet are not allowed.
   (d) Calcuators with which it is possible to send and recieve messages of any kind are not allowed.
3. **Physical (paper) dictionary** (no electronic dictionary allowed).
   (a) Dictionary must contain *no notes* of any kind.
   (b) Each student must have his/her own dictionary. It is not allowed for students to pass a dictionary between them.
4. **Ruler.**
5. Collection of formulae and Statistical Tables named '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*', that the student brings to the exam location.
6. Please note that a collection of critical values for the Student´s t, Normal, Chi-square and F-distributions is given in the Appendix of the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*'.
7. Also note that the '*Test template*', that should be used when performing tests, is given in the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams'.*

That is:

1. NO BOOK (except paper-dictionary) is allowed.
2. NO (student-written) notes are allowed.
3. NO other document than the one 'Collection of Formulae and Statistical Tables for Time Series Exam' is allowed.

Instructions: **Please note the following:**

1. Start with reading through the instructions!
2. Make sure you **follow** the instructions!
3. Start with reading through the exam.
4. You may write your solutions in Swedish or English.
5. Total score is **100** points
   (a) If you want the ECTS grades, please indicate that on the cover page!
   (b) For each task the maximum number of points is given within parenthesis, e.g. (16p in total).
   (c) For each subtask the number of points is given within parenthesis, e.g. (2p)
6. All solutions must be on separate sheets. No solutions on the questionnaire! (If so, they will be disregarded.)
7. Make sure your solutions are: easy to read and easy to understand, that is:
   (a) For each task that you solve, please start with a new sheet: after Task 1, start with a blank sheet for Task 2, etc.
   (b) Write the *task number* at the top of each page, in the

   ....**MIDDLE OF THE PAGE!!!**..

   Like:

   ....**TASK 1**..

   - if you write it in the upper left corner, the staple will cover it, and there is no for way for the examinator to know if the text of that sheet belongs to the previous sub-task or what it is. The Examinators will not make any 'qualified guesses' of what is being displayed on any given page. It is the responsibility of the student to make sure that every task and sub-task is easily identifiable.

(c) If you continue a sub-task on the next sheet of paper - indicate that at the top of the page - IN THE MIDDLE OF THE PAGE, like, for example:

....'Task 1B (cont.)'..

(d) Please separate each subtask A, B etc with a horizontal line across the sheet
____
if they are on the same sheet of paper - that way it will be easy for the examinator to actually see where one subtask ends and next begins.

(e) For examinator readability, it is highly recommended that you use a pencil, (and not a pen), which will allow you to erase and rewrite if you make a mistake. Crossed-over text and corrections using 'tipp-ex' will just cause blurriness and confusion to the examinator.

(f) For examinator readability: Write clearly, that is, letters, mathematical/statistical symbols and numbers should be easy recognizable!! Do not underestimate the correlation between readability and points scored, that is, when readability goes to zero, points scored also goes to zero, no matter your intentions or wheather *you* can read it or not.

(g) Also note that everything that you write will be taken at 'face value'. That is, for example, if you write β_1 the examinator will take that as a β_1 even though you may claim that it is given from the context it should be clear that you meant something else, like β_3. Thus, given this example, writing β_1, and that is not correct in that specific formula or statement, this will lead to subtraction of points, even if you will claim that it is just a typo, and that in another task or subtask, it is clear that you understand the issue.

(h) Please put the sheets in **order**, that is first Task 1, and then Task 2 etc...

8. Please keep the questionaire.

9. Do well!

Task 1

(12 points in total) Consider the following single linear regression

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

A) (6p) Do the following:

1. Draw a Figure representing the Population Regression Function (PRF), draw the regression line, mark out what is displayed on the axes.
2. Mark out what distance is represented by β_1.
3. Mark out what distance is represented by β_2.
4. Mark out an arbitrary observation Y_i, given this observation, mark out the *conditional expected value* given the corresponding X_i, that is, mark out exactly where in the Figure this conditional expected value is 'located',
5. Write down a formula for the conditional expected value of Y.
6. Indicate in the Figure what distance that is represented by u_i.

B) (6p) Do the following:

1. In a SEPARATE FIGURE from the one in Sub-task A, draw a Figure representing the corresponding *Sample* Regression Function (SRF) for the model above, draw the sample regression line. Mark out what is displayed on the axes.
2. Mark out what distance is represented by $\widehat{\beta_1}$.
3. Mark out what distance is represented by $\widehat{\beta_2}$.
4. Mark out an arbitrary observation Y_i, and given this observation, mark out the *estimated conditional expected* value given the corresponding X_i, that is, mark out exactly where in the Figure this estimated conditional expected value is 'located'.
5. Write down a formula for the estimated conditional expected value of Y_i given that value of X_i.
6. Indicate in the Figure what distance that is represented by $\widehat{u}_i$.

Task 2

(42 points in total) Consider the following model

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + u_i \tag{1}$$

See Eviews output in the figures below.

A) (6p) Do a formal test of the model to see if it has any explanatory power at all. Choose the significance $\alpha = 0.05$. Make sure all the steps and calculations you make are easy to follow and understand. Fully document the test procedure as outlined in the test-templete.

B) (6p) Test the hypothesis that for a one unit change in X_3, the conditional mean of Y changes by at least two units, against the alternative that the change in the conditional mean of Y is less than two. Choose the significance $\alpha = 0.05$. Make sure all the steps and calculations you make are easy to follow and understand. Fully document the test procedure as outlined in the test-templete.

C) (6p) Starting with a relevant probabilistic statement, derive a confidence interval for a general confidence level $(1 - \alpha) 100\%$ for β_3. State each and every assumption that you (have to) make in order to derive this interval. Write out each step explicitly so that it is clear what happens from one step to another.

D) (4p) For model (1), calculate and interpret a 90% confidence interval for β_3. Explicitly state the lower and upper bounds of the interval.

E) (6p) There is a theory that suggests that X_1 and X_2 have equal effect on Y. Using the Eviews ouput in this task, design and perform a test of this theory. Choose the significance $\alpha = 0.05$. Make sure all the steps and calculations you make are easy to follow and understand. Fully document the test procedure as outlined in the test-templete.

F) (6p) A collegue of yours claim that credible inference about the population cannot be made for model (1) since the sample size is small. Perform a formal test prove him wrong. Choose the significance $\alpha = 0.05$. Make sure all the steps and calculations you make are easy to follow and understand. Fully document the test procedure as outlined in the test-templete.

G) (4p) For the estimated model (1), interpet the coefficient of determination, that is, interpet the actual number.

H) (4p) For the estimated model (1), interpet the adjusted coefficient of determination, that is, interpet the actual number.

Dependent Variable: Y
Method: Least Squares
Date: 11/07/16 Time: 17:26
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.040130	2.906944	2.077828	0.0481
X1	7.205017	0.669640	10.75954	0.0000
X2	8.239843	0.656590	12.54945	0.0000
X3	2.175709	0.532902	4.082756	0.0004
X4	-1.638648	0.924089	-1.773258	0.0884

R-squared	0.961890	Mean dependent var	-4.227556
Adjusted R-squared	0.955792	S.D. dependent var	70.21633
S.E. of regression	14.76350	Akaike info criterion	8.373205
Sum squared resid	5449.024	Schwarz criterion	8.606738
Log likelihood	-120.5981	Hannan-Quinn criter.	8.447914
F-statistic	157.7469	Durbin-Watson stat	1.432643
Prob(F-statistic)	0.000000		

¶

Figure 2.1 Eviews output.

Dependent Variable: Y
Method: Least Squares
Date: 11/07/16 Time: 17:12
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.178412	2.730859	1.896258	0.0691
X1+X2	7.729081	0.319860	24.16392	0.0000
X3	2.259990	0.522388	4.326267	0.0002
X4	-1.681727	0.919187	-1.829580	0.0788

R-squared	0.960677	Mean dependent var	-4.227556
Adjusted R-squared	0.956140	S.D. dependent var	70.21633
S.E. of regression	14.70529	Akaike info criterion	8.337858
Sum squared resid	5622.388	Schwarz criterion	8.524684
Log likelihood	-121.0679	Hannan-Quinn criter.	8.397625
F-statistic	211.7303	Durbin-Watson stat	1.608808
Prob(F-statistic)	0.000000		

¶

Figure 2.2 Eviews output.

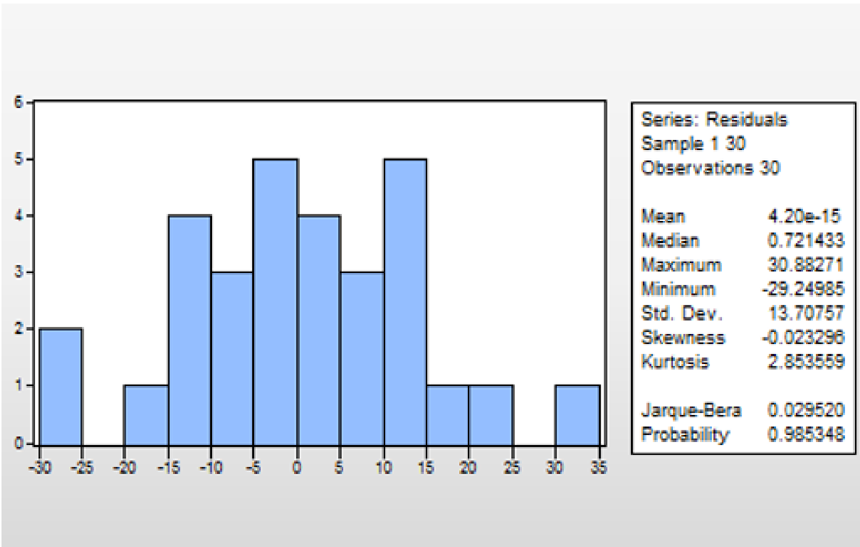

Figure 2.3: Residual analysis from the estimation of model 1

Task 3

(22 points in total)

NOTE: There is NO estimation output for this task!

Consider the following model for quarterly data for Swedish GDP

$$GDP_t = \beta_1 + \beta_2 RDisp_t + \gamma_2 D_{2,t}^{Q2} + \gamma_3 D_{3,t}^{Q3} + \gamma_4 D_{4,t}^{Q4} + e_t \tag{2}$$

where

1. $RDisp_t$ is (aggregate) real disposable income and
2. $D_{2,t}^{Q2}$ is a dummy variable taking on the value 1 for Quarter 2, zero otherwize
3. $D_{3,t}^{Q3}$ is a dummy variable taking on the value 1 for Quarter 3, zero otherwize
4. $D_{4,t}^{Q4}$ is a dummy variable taking on the value 1 for Quarter 4, zero otherwize

A) (5p) Draw a scetch for the PRF for model (2) where you clearly mark out the four regression lines that corresponds to each quarter. Also, mark out the distances that are represented by all the different parameters of the model, for the purpose of this scetch - assume that $\beta_1 > 0$, $\beta_2 > 0$, $\gamma_j > 0$ for $j = 2, 3, 4$ and that $\gamma_2 < \gamma_3 < \gamma_4$. You need to mark out β_2 only once. Do not mark out any observation Y_i and thus there is no need to mark out any error term distance from any regression line.

B) (6p) Without actually performing the test, outline the steps for a test of seasonality in GDP. That is

1. Write down the two models that you would have to estimate in order to do this test, name them in an intuitive way.
2. Write down the null and the alternative hypotheses.
3. Write down the test statistic for the test and its distribution under the null. Clearly define all quantities that constitutes the test statistic.

Again, note that there is no Eviws output in this task - you cannot and should not perform the actual test.

C) (5p) Rewrite the model in (2) so that it

1. has no intercept, but still

2. does not suffer from perfect multicollinearity.
3. Denote this model (2^{**}) for ease of reference.

You may or may not have to define one or more new variable(s). If so, write out the definition(s) clearly. For the purpose of next subtask, denote all parameters in your newly defined model (2^{**}) with superscript **, that is, for instance β_2^{**} so that there will be no confusion regarding the next subtask.

D) (6p) Now, given the model in (2) and model (2^{**}), write down the relationship between the parameters in the two models, on the format, for instance

$$\beta_2^{**} = f\left(\text{parameter(s) of model } (2)\right)$$

where you explicitly state the function $f()$ for each and every parameter in model (2^{**}) as functions of the parameter in model (2).

Task 4

(24 points in total)

Consider the following (true) model

$$Y_i = \beta_0 + \beta_1 X_i + \gamma W_i + u_i \tag{3}$$

where the regressors X_i and W_i are to be considered random (i.e. not fixed in repeated sampling) and where

$$\begin{aligned} Cov(X_i, W_i) &\neq 0, \\ Cov(X_i, u_i) &= 0, \\ Cov(W_i, u_i) &= 0. \end{aligned}$$

A) (4p) Now a researcher is not aware of the nature of the true model, and instead he runs the following regression

$$Y_i = \beta_0 + \beta_1 X_i + e_i. \tag{4}$$

Given that (3) is indeed the true model, calcuate $Cov(X_i, e_i)$ for model (4).

B) (2p) Is there any problem with the OLS estimator of β_1 when estimating the model (4)? If so, what are these problems?

C) (6p) Now, for model (3) assume that $\beta_0 = \gamma = 0$ and derive the OLS estimator for β_1. State any assumptions, if you make any, as you make them, that you need for this derivation. Any assumptions, that are not nessecary for this derivation, will result in reduction of points.

D) (6p) Derive $E\left(\widehat{\beta_1}\right)$ for the OLS estimator above derived in the previous subtask. You do not have to restate any assumptions that you stated in subtask A. But over and above those, state any assumptions, if you make any, as you make them, that you need for this derivation. Any assumptions, that are not nessecary for this derivation, will result in reduction of points.

E) (6p) Assume that the error term e_i is uncorrelated and homoscedastic, derive $V\left(\widehat{\beta_1}\right)$ (start with the definition of the variance). You do not have to restate any assumptions that you stated in previous subtasks. But over and above those, state the assumptions of uncorrelated and homoscedastic error terms *when* (at what stage of the derivation) you actually use these assumptions.