Supplementary Written Examination in Econometrics (B2)

Spring 2016

2016-04-20 08.00-12.00

Bergsbrunnagatan 15, 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 $\widehat{u}_i$.

For the following assumptions for the linear regression model, explain in words what they actually mean, or how they can be interpreted in terms of the 'idea of the model'.

C) (3p) $Cov(u_i, u_j) = 0 \; i \neq j$
D) (3p) $Cov\left(X_i, u_i\right) = 0$

Task 2

(38 points in total)

Consider the following models for corporate executive pay:

$$Salary_i = \beta_1 + \beta_2 Age_i + \beta_3 Tenure_i + \beta_4 \text{Profits}_i + u_i \tag{1}$$

$$Salary_i = \beta_1 + \beta_2 Age_i + \beta_3 Tenure_i + \beta_4 Sales_i + u_i \tag{2}$$

and

$$Salary_i = \beta_1 + \beta_2 \text{Profits}_i + u_i \tag{3}$$

$$Salary_i = \beta_1 + \beta_2 Tenure_i + \beta_3 \text{Profits}_i + \beta_4 Sales_i + u_i \tag{4}$$

For Eviews output see figures.

A) (3p) For model (1) interpret the parameter β_2.

B) (3p) For model (1) interpret the estimated parameter $\widehat{\beta}_3$.

C) (6p) For model (2) test if Sales has any affect on salary whatsoever, use significance level 5%. Make sure to dokument the test as outlined in the test-template.

D) (6p) For model (1) test the null that Profits has a positive impact on Salary, use significance level 5%. Make sure to dokument the test as outlined in the test-template.

E) (4p) Interpret the coefficient of determination for models (1) and (2) . Given this measure, and this measure alone, which model is the best?

F) (6p) Compare the explanatory power of model (2) and (3) using a relevevant measure. Explain why you use this measure and not any other in this situation.

G) (6p) Derive a 95% confidence interval for the parameter tenure in model (2) .

H) (4p) Calculate and interpret a 95% confidence internal for the parameter tenure in model (2) .

Equation: MODEL1 Workfile: CEO::Ceo\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALARY
Method: Least Squares
Date: 04/18/16 Time: 11:56
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	956.7280	654.0031	1.462880	0.1442
AGE	9.206902	12.01398	0.766349	0.4439
TENURE	33.41977	9.904277	3.374277	0.0008
PROFITS	0.412659	0.048451	8.517105	0.0000

R-squared	0.168780	Mean dependent var	2027.517
Adjusted R-squared	0.163151	S.D. dependent var	1722.566
S.E. of regression	1575.794	Akaike info criterion	17.57181
Sum squared resid	1.10E+09	Schwarz criterion	17.60853
Log likelihood	-3923.301	Hannan-Quinn criter.	17.58629
F-statistic	29.98391	Durbin-Watson stat	1.920464
Prob(F-statistic)	0.000000		

Figure 2.1: Eviews output

Equation: MODEL2 Workfile: CEO::Ceo\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALARY
Method: Least Squares
Date: 04/17/16 Time: 20:03
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	836.5745	652.8046	1.281508	0.2007
AGE	7.999714	11.99824	0.666741	0.5053
TENURE	35.51031	9.894298	3.588967	0.0004
SALES	0.039886	0.004618	8.637038	0.0000

R-squared	0.172084	Mean dependent var	2027.517
Adjusted R-squared	0.166478	S.D. dependent var	1722.566
S.E. of regression	1572.659	Akaike info criterion	17.56783
Sum squared resid	1.10E+09	Schwarz criterion	17.60454
Log likelihood	-3922.410	Hannan-Quinn criter.	17.58231
F-statistic	30.69286	Durbin-Watson stat	1.929326
Prob(F-statistic)	0.000000		

Figure 2.2: Eviews output

Equation: MODEL3 Workfile: CEO::Ceo\

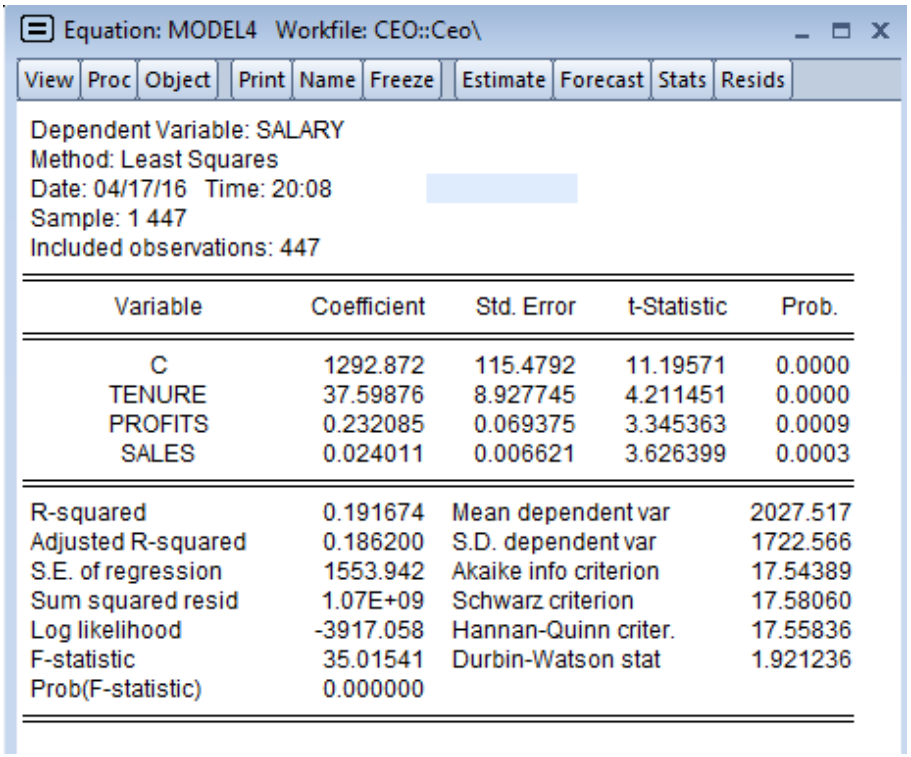
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALARY
Method: Least Squares
Date: 04/17/16 Time: 20:04
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1737.852	83.22906	20.88035	0.0000
PROFITS	0.413534	0.049174	8.409685	0.0000

R-squared	0.137133	Mean dependent var	2027.517
Adjusted R-squared	0.135194	S.D. dependent var	1722.566
S.E. of regression	1601.899	Akaike info criterion	17.60023
Sum squared resid	1.14E+09	Schwarz criterion	17.61859
Log likelihood	-3931.652	Hannan-Quinn criter.	17.60747
F-statistic	70.72280	Durbin-Watson stat	1.876498
Prob(F-statistic)	0.000000		

Figure 2.3: Eviews ouput



Equation: MODEL4 Workfile: CEO::Ceo\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALARY
Method: Least Squares
Date: 04/17/16 Time: 20:08
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1292.872	115.4792	11.19571	0.0000
TENURE	37.59876	8.927745	4.211451	0.0000
PROFITS	0.232085	0.069375	3.345363	0.0009
SALES	0.024011	0.006621	3.626399	0.0003

R-squared	0.191674	Mean dependent var	2027.517
Adjusted R-squared	0.186200	S.D. dependent var	1722.566
S.E. of regression	1553.942	Akaike info criterion	17.54389
Sum squared resid	1.07E+09	Schwarz criterion	17.58060
Log likelihood	-3917.058	Hannan-Quinn criter.	17.55836
F-statistic	35.01541	Durbin-Watson stat	1.921236
Prob(F-statistic)	0.000000		

Figure 2.4: Eviews ouput

Task 3

(32 points in total) (For Eviews output for this task, refer to task 2).

A) (6p) Given that Profits is already in the model, perform a test to test if Tenure and Sales contribute to the explanatory power of the model. Perform the test at 5% significance level and make sure to dokument the test as outlined in the test-template.

B) (6p) For model (2) perform a test of the model. Perform the test at 5% significance level and make sure to dokument the test as outlined in the test-template.

C) (5p) One of the assumptions for the (multiple) linear regression model is that we have *no exact (or perfect) multicollinearity*. What does it mean that we have *exact* multicollinearity in our data? Explain in words what it means, and give one example using a formula.

D) (5p) What does it mean that we have a *high degree* of (but not perfect) multicollinearity? Explain using words, (no formulae needed) in terms of explanatory power of the regressors.

E) (5p) When doing empirical work, *before* we do an actual estimation of *any* model, how can we know if there is a possibility that we are facing the problem of multicollinearity?

F) (5p) Let us say you have estimated a multiple regression model model using Eviews or any other software. In terms of the standard output from that estimation, what would you look at to give you an idea about if multicollinearity could be a problem? Is multicollinearity potentially a problem for model (4)?

Task 4

(12 points in total)

Consider the following model

$$Y_i = \beta_2 X_{i,2} + u_i$$

A) (6p) Derive the OLS estimator for β_2. State any assumptions that you make as you make them.

B) (6p) Under the assumption of $Var(u) = \sigma_i^2$ and $Cov\left(u_i, u_j\right) = 0$, derive the variance of the OLS estimator. State any assumptions that you make as you make them.