Examination Period 3: 2016/17

ECN201417N

Module Title: Data Analysis for Economists
Level: Five
Time Allowed: Two hours

Instructions to students:

- Enter your student number not your name on all answer books.
- Answer three questions:
  - One question from Section A
  - One question from Section B
  - One other question of your choice from Section A or Section B

- All questions are equally weighted. Where a question has more than one part the division of marks is stated.
- Begin each question in a separate answer book; label each answer book clearly with the number of the question you are answering.
- The use of a non-programmable calculator is permitted.
- Formula sheets and statistical tables are provided in a separate document.

<table>
<thead>
<tr>
<th>No. of Pages</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Questions</td>
<td>6</td>
</tr>
</tbody>
</table>
Section A

Answer at least one question and up to one further question. You must answer a total of three questions between Section A and Section B.

Question 1

Data was collected on the British manufacturing sector on the real gross product in millions of British Pounds ($Y_t$), labour input, in millions of days ($X_{2t}$) and capital input, in millions of British Pounds, ($X_{3t}$) for 2001 – 2015. You are given the following information based on 15 observations:

\[
\begin{align*}
\Sigma Y_t &= 362.040 & \Sigma X_{2t} &= 3.980 & \Sigma X_{3t} &= 375.800 & \Sigma Y_t^2 &= 323.707 \\
\Sigma Y_t \times X_{2t} &= 0.905 & \Sigma Y_t \times X_{3t} &= 453.809 & \Sigma X_{2t}^2 &= 0.004 & \Sigma X_{3t}^2 &= 744.550 \\
\Sigma X_{2t} \times X_{3t} &= 0.896
\end{align*}
\]

Where lowercase letters denote deviations from sample mean values [i.e. $y_t = Y_t - \bar{Y}$]. The questions below pertain to a three-variable linear regression model for which the population regression equation can be written in conventional notation as:

\[Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \mu_t\]

a. What are a priori expectations regarding the anticipated signs of the regression coefficients? Justify your answers.  
(5 marks)

b. Use the above information to derive the OLS estimates of the intercept coefficient $\beta_1$ and the slope coefficients $\beta_2$ and $\beta_3$. Show all workings.  
(12 marks)

c. Interpret the regression results and explain in words what they mean.  
(8 marks)

d. Obtain $R^2$ and explain its meaning.  
(8 marks)

Total: 33 marks

Section A continues overleaf
Question 2

Using information in question 1,

a. Compute the estimated variance of $\beta_2$ and $\beta_3$ and the estimated standard error of $\beta_2$ and $\beta_3$.  

(b) 8 marks

b. Estimate the individual t statistics of the independent variables and test the hypotheses that $\beta_2=0$ and $\beta_3=0$ at 95% level. 

(c) 10 marks

c. Test the overall significance of the model at 95% level. 

(d) 10 marks

d. Explain what type I and type II errors are. 

Total: 33 marks

Question 3

A 2015 study in the UK investigated the relationship between house prices, size of house, plot size and number of bedrooms. Equation 1 is the least squares estimated model (standard errors in parentheses):

$$\ln(\text{Price}) = -1.297 + 0.037\text{bedrooms} + 0.7\ln(\text{housesize}) + 0.168\ln(\text{plot})$$  

Equation 1

$$\text{se} = (0.651) (0.028) (0.093) (0.038)$$  
n=88, $R^2= 0.643$

Where Price is house price (thousands of British Pounds), bedrooms is number of bedrooms, housesize is the size of house (square feet), plot is the size of the plot (square feet) and ln denotes the natural logarithm.

a. What are a priori expectations on the anticipated signs of the regression coefficients? Interpret the regression results from Equation 1. 

(13 marks)

b. Conduct appropriate t and F tests for the significance of all independent variables in the model at 95% level of significance. What other independent variables could be added to the regression and why?

(20 marks)

Total: 33 marks

End of Section A

Section B follows overleaf
Section B

Answer at least one question and up to one further question. You must answer a total of three questions between Section A and Section B.

Question 4

Production function studies that are testing for constant returns to scale sometimes specify a restricted model, as shown in equation 2 below, which can then be considered, alongside the unrestricted model (equation 1).

\[ \ln Q = \beta_0 + \beta_1 \ln L + \beta_2 \ln K + u \]  \hspace{1cm} \text{Equation 1}

\[ \ln \left( \frac{Q}{K} \right) = \beta_0 + \beta_1 \ln \left( \frac{L}{K} \right) + u \]  \hspace{1cm} \text{Equation 2}

Where Q is a measure of gross domestic product (millions of dollars), L is employment in thousands of people and K is capital (millions of dollars), and ln denotes the natural logarithm.

Based on annual data for the United States between 1978 and 2016, the above equations have been estimated by least squares with the following results (standard errors in parentheses):

\[ \ln Q = -3.845 + 1.522 \ln L + 0.423 \ln K \]  \hspace{1cm} \text{Equation 3}

\[ \text{se} = (0.245) \ (0.089) \ (0.052) \]

\[ \text{RSS} = 0.045, \ n=39, \ \text{sample is between 1978-2016} \]

\[ \ln Q = -4.045 + 1.634 \ln L + 0.231 \ln K \]  \hspace{1cm} \text{Equation 4}

\[ \text{se} = (0.342) \ (0.211) \ (0.252) \]

\[ \text{RSS}=0.032, \ n=20, \ \text{sample is between 1978-1997} \]

\[ \ln Q = -2.521 + 1.011 \ln L + 0.582 \ln K \]  \hspace{1cm} \text{Equation 5}

\[ \text{se} = (0.532) \ (0.143) \ (0.055) \]

\[ \text{RSS}=0.003, \ n=19, \ \text{sample is between 1998-2016} \]

\[ a. \] Show how it is that equation 2 incorporates the constant returns to scale restriction. \hspace{1cm} (10 marks)

\[ b. \] What are a priori expectations on the anticipated signs of the regression coefficients? Interpret the regression results from Equation 3. \hspace{1cm} (8 marks)

Question 4 continues overleaf
c. Suppose that we estimate Equation 3 for the two periods 1978-1997 (Equation 4) and 1998-2016 (Equation 5). Use the results from regressions for Equations 3, 4 and 5 to conduct an appropriate test of parameter constancy. Discuss your results. 

(15 marks)

Total: 33 marks

Question 5

A model, with parameters \( \beta_0 \), \( \beta_1 \) and \( \beta_2 \) is estimated by least squares with the following results:

\[
\hat{Y}_t = 1.1 + 0.7X_{1t} + 1.4X_{2t} \tag{Equation 1}
\]

The residual sum of squares, RSS=15; the total sum of squares, TSS=95; and the total number of observations is n=25.

A revised model, adding two extra variables with parameters \( \beta_3 \) and \( \beta_4 \), gave;

\[
\hat{Y}_t = 1.5 + 0.8X_{1t} + 0.6X_{2t} - 1.09X_{3t} + 2.9X_{4t} \quad \text{RSS}=4 \tag{Equation 2}
\]

a. Calculate \( R^2 \) for each model. 

(10 marks)

b. For model (1) test the hypothesis that \( \beta_1 = \beta_2 = 0 \). 

(10 marks)

c. Test the hypothesis that \( \beta_3 = \beta_4 = 0 \) using the RSS values for the two models. 

(13 marks)

Total: 33 marks
Question 6

a. What is meant by heteroskedasticity? (8 marks)

b. Explain its effects on
   i. OLS estimators and their variances
   ii. The use of t and F tests of significance (10 marks)

c. What are the methods for detecting heteroskedasticity? (10 marks)

d. How can we remedy heteroskedasticity? (5 marks)

Total: 33 marks