Summer Examinations 2016

ECN201416N

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 from Section A, one from Section B and 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.
- Mathematical formulae sheets and statistical tables will be provided.

<table>
<thead>
<tr>
<th>No. of Pages</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Questions</td>
<td>6</td>
</tr>
</tbody>
</table>
Section A

1. Data was collected on the Taiwanese manufacturing sector on the real gross product in millions of Taiwanese dollars ($Y_t$), labour input, in millions of days ($X_{2t}$) and capital input, in millions of Taiwanese dollars, ($X_{3t}$) for 2000 – 2014. You are given the following information based on 15 observations.

$$
\Sigma Y_t = 371.030 \quad \Sigma X_{2t} = 4.310 \quad \Sigma X_{3t} = 382.600 \quad \Sigma y_t^2 = 332.606
$$

$$
\Sigma y_t x_{2t} = 0.835 \quad \Sigma y_t x_{3t} = 443.602 \quad \Sigma x_{2t}^2 = 0.003 \quad \Sigma x_{3t}^2 = 753.330
$$

$$
\Sigma x_{2t} x_{3t} = 0.944
$$

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:

$$
Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \epsilon_t
$$

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

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

(Total 33 marks)

3. The equation below is a least squares estimated relationship between teachers’ annual earnings, their teaching experience, ethnicity and gender in the UK for year 2013 (standard errors in parentheses).

$$\hat{Y}_i = 15.374 + 1.383X_{1i} + 2.159D_{1i} + 1.962D_{2i}$$

se = (0.345) (0.842) (0.532) (0.723)

where $Y$= teachers’ annual salary in thousands of GBP (£);  
$X_1$= years of teaching experience;  
$D_1$=1 if male teacher  
=0 otherwise;  
$D_2$=1 if Caucasian  
=0 otherwise  
n=30

a. Indicate a priori expectations and interpret the regression results. Test the significance of the independent variables in the model.  

b. Explain the role of dummy variables in regression models.  

(13 marks)

(8 marks)

Question 3 continues overleaf
c. Estimate the mean salaries for the following:

1. Caucasian female teachers
2. Non-Caucasian female teachers
3. Caucasian male teachers
4. Non-Caucasian male teachers

(12 marks)

(Total 33 marks)

End of Section A
Section B follows overleaf
Section B

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).

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

\[
\ln (Q/L) = \beta_0 + \beta_2 \ln (K/L) + u \quad \text{Equation 2}
\]

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, the above equations have been estimated by least squares with the following results (standard errors in parentheses):

\[
\ln Q = -1.652 + 0.341 \ln L + 0.846 \ln K \quad \text{Equation 3}
\]

\[
\text{se} = (0.606) \quad (0.188) \quad (0.093)
\]

\[
\text{RSS} = 0.014, \quad n = 20
\]

\[
\ln (Q/L) = -0.495 + 1.0153 \ln (K/L) \quad \text{Equation 4}
\]

\[
\text{se} = (0.122) \quad (0.016)
\]

\[
\text{RSS} = 0.017
\]

a. Show how it is that Equation 2 incorporates the constant returns to scale restriction. 

(10 marks)

b. What are a priori expectations on the anticipated signs of the regression coefficients? Interpret the regression results from Equation 3.

(8 marks)

c. Indicate how you would use the results from regressions for both Equations 3 and 4 to test the constant returns to scale hypothesis.

(15 marks)

(Total 33 marks)
5. A recent study has investigated the relationship between the number of employees and the sales revenue of 249 small to medium sized companies in the USA for the year 2010. **Equation 1** is the least squares estimated model (standard errors in parentheses):

\[
Sales_i = 199.543 + 5.439 \cdot Employees_i \quad (Equation 1)
\]

\[
\text{se} = (99.838) \quad (1.554)
\]

\[
R^2 = 0.654
\]

\[
\ln \hat{U}_i = 38.817 - 2.654 \ln Employees_i \quad (Equation 2)
\]

\[
\text{se} = (38.345) \quad (4.215)
\]

\[
R^2 = 0.080
\]

where Sales= total company sales in $1000
Employees= number of FTEs employed by the company

a. What is a priori expectation of the sign of \( Employees \) coefficient? Interpret the regression results from **Equation 1**. Test the significance of 'Employees' in the model.  

(8 marks)

b. Consider **Equation 2** to test for heteroskedasticity in the model using Park test. 

(15 marks)

c. Apart from Park test, which other tests can detect heteroskedasticity? 

(5 marks)

d. What are the remedies to heteroskedasticity? 

(5 marks)

(Total 33 marks)
6.

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

b. Explain its effect on OLS estimators and their variances 
(10 marks)

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

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

(Total 33 marks)

End of Section B
End of Paper