PRIFYSGOL ABERYSTWYTH UNIVERSITY

SEMESTER TWO 2026

ABERYSTWYTH BUSINESS SCHOOL

AB23420: ECONOMETRICS

Time allowed: 3 HOURS

You must answer FOUR questions in total. Answer TWO questions from SECTION A and TWO questions from SECTION B.

All questions carry equal marks. Marks for separate parts of each question are provided in brackets.

A Help sheet and Statistical tables are provided.

CALCULATORS ARE PERMITTED (Casio FX83 or FX85 models only)

STUDENTS ARE EXPECTED TO SUBMIT ANSWERS TO THIS PAPER IN ENGLISH

Once seated, students may complete the front cover of the answer book(s) and the attendance slip ONLY.

DO NOT TURN THE EXAMINATION QUESTION PAPER OVER UNTIL INSTRUCTED TO DO SO.

Section A

1.

a) Consider the following non-linear time series model: $Y_t = Y_0 e^{(\beta_1 t + u_t)}$, where Y_t is a time series variable, t is time measured in annual years, and Y_0 is the value that Y takes in the first period. Explain why simple linear regression techniques cannot be used to the above model?

(12%)

b) Transform the model in part I above so that OLS can be used to estimate the parameters of the model. State the population regression function and interpret the parameters. Roughly sketch the population regression line providing appropriate labels to the axes and the parameters of the regression line.

(28%)

c) Regression analysis deals with the dependence of one variable on another. Does such an analysis necessarily imply causation? Explain with an example.

(20%)

d) In the context of a simple linear regression model, where the population regression function is given by $Y_i = \beta_1 + \beta_2 X_i + u_i$, and the sample regression function is given by $Y_i = \widehat{Y_i} + \widehat{u_i} = \widehat{\beta_1} + \widehat{\beta_2} X_i + \widehat{u_i}$, prove the following:

(20 x 2 = 40%)

i. The mean value of the estimated residuals is zero, i.e. $\widehat{u_i} = 0$

ii. The total sum of squares is equal to the sum of the explained sum of squares and the residual sum of squares, i.e. TSS=ESS+RSS.

2.

a) With reference to OLS estimation in regression analysis, explain what is meant by:

(16%)

i. Coefficient of determination
ii. Standard error of the regression

b) A health economist attempts to study the demand for marijuana among regular users of the drug (addicts) and considers the following estimated equation:

$$\underset{}{log(weeds)_i} = \underset{}{-5.76} - \underset{(0.09)}{0.08\, pric_i} + \underset{(0.01)}{0.05 inc_i} + \underset{(0.001)}{0.04 pcoke_i} + \underset{(0.02)}{}$$

n=120, $R^2 = 0.3783$, $Adj - R^2 = 0.3783$, $F(3, 116) = 80.00$

(Figures in brackets are the estimated standard errors of the coefficients, n is the sample size, R^2 is the coefficient of determination, Adj-R^2 is the adjusted R^2, and F is the F-statistic)

Please turn over ....

where, *weeds* is the total amount of marijuana consumed (in ounces) by an individual in a year, *pric* is the average price (in pound sterling) paid by an individual for a bag containing 1 ounce of marijuana, *inc* is the annual income of the individual and *pcoke* is the average price (in pound sterling) paid by an individual for a bag containing 1 ounce of cocaine. Now answer the following questions:

i. Interpret the estimated coefficients on pric, inc, and pcoke. Are the signs and magnitudes of the coefficients in line with your a priori expectations? (20%)

ii. Using the test of significance approach, test the hypothesis that price has no effect on consumption of marijuana, controlling for income of individuals and price of cocaine in the population against the alternative that it has a positive effect. (NB: Show all hypothesis, test procedure and conclusions. Use 5% significance level). (32%)

iii. Using the F-test, carry out the test for overall significance of the model estimated above. (NB: Show all hypothesis, test procedure and conclusions. Use 5% significance level). (32%)

3.

a) Distinguish between Type I and Type II errors in the context of hypothesis testing in regression analysis. Also, explain what is meant by confidence coefficient and level of significance and how they are related to Type I and/or Type II errors. (20%)

b) Consider the following regression model:

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

Discuss and explain how an F–test may be employed to test the hypothesis $H0 : \beta_3 = \beta_4 = 0$ against the alternative hypothesis $H1: at\ least\ one\ \beta_j \neq 0\ (\ j = 3,4)$. (20%)

c) Suppose you are interested in estimating an econometric model aimed at analyzing whether housing rental rates are influenced by the student population in a university town. Let rent be the average monthly rent paid on rental units in a University town in the UK. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for such a relationship is:

$$ln\ (rent_i) = \beta 0 + \beta 1\ ln\ (pop_i) + \beta 2\ ln\ (avginc_i) + \beta 3\ (pctstu_i) + u_i$$

i. What signs do you expect for β_1 and β_2? Explain. (8%)

ii. Now if the equation estimated using data for 70 University towns for a given year is:

$$\underset{}{\ln\ (rent_i)} = \underset{(0.01)}{0.05} + \underset{(0.04)}{0.07 \ln\ (pop_i)} + \underset{(0.09)}{0.51 \ln\ (avginc_i)} + \underset{(0.002)}{0.01\ (pctstu_i)}$$

Please turn over ....

$n = 70, \quad R^2 = 0.512,$

Where numbers in parenthesis are standard errors, n is the total number of observations, R^2 is the coefficient of determination, and the variables are as defined earlier.

What is wrong with the statement: "a 10% increase in the average city income is associated with about 51% increase in rent"? (8%)

iii. State the null hypothesis that the size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect (without being specific whether the effect is positive or negative).

(12%)

iv. Using the confidence interval approach to hypothesis testing, test the hypothesis stated in part (c) above at the 5% level of significance.

(32%)

4.

a) In the context of regression analysis, distinguish between an intercept dummy and a slope dummy.

(20%)

b) Outline what is meant by the dummy variable trap.

(20%)

c) The following consumption equation was estimated using OLS on data for 31 years from 1960 to 1990 for a developed economy:

$$\widehat{Y_t} = \underset{(15.21)}{4.2} + \underset{(0.02)}{0.08X_t} + \underset{(10.76)}{12.3\,D_t} - \underset{(0.001)}{0.05\,(DX)_t}$$

$n=31,\ R^2 = 0.837,\ \text{Adj-}R^2 = 0.832,\ F(3, 27) = 72.31$

(Figures in brackets are the estimated standard errors of the coefficients, n is the sample size, R^2 is the coefficient of determination, Adj-R^2 is the adjusted R^2, and F is the F-statistic)

where, Y is consumption (in billions of pounds), X is income (in billions of pounds), D is a dummy variable that takes the value 1 for 1960-1977, and 0 for 1978-1990, and (DX) is an interaction term (i.e. variable X multiplied by variable D). Now answer the following questions:

i. Comment on whether the estimated model is an ANOVA or an ANCOVA model. (8%)

ii. What is the benchmark period in the above estimated equation? (8%)

iii. By how much is the marginal propensity to consume in the period 1960-77 different from the marginal propensity to consume in the period 1978-1990?

(12%)

iv. Using the above results, write down the estimated consumption equation for 1960-1977 and the estimated consumption equation for 1978-1990. Also, sketch out the estimated consumption equation for the two time periods: 1960-1977 and 1978-1990 (NB: plots

v. don't have to be exact but should be approximate; make sure that you provide clear labels).

(20%)

vi. Explain how one can test if there is a structural break in the relationship between consumption and income for the given developed economy in 1977 based on the estimation results presented above.

(20%)

SECTION B

5.

Write short notes on any THREE of the following:

(NB: Clearly outline null and alternate hypotheses, any assumptions, steps involved in the procedure or test and implications/consequences, wherever applicable. Also provide examples and/or diagrams, wherever possible, to support your explanations)

(100%)

a) White's Test
b) Weighted Least Squares Estimation
c) Test for normality of errors
d) Breusch – Godfrey Test
e) Breusch – Pagan Test

6.

a) State with brief explanations whether the following statements are true, false or uncertain: (20%)

i. In the presence of autocorrelation, OLS estimators are unbiased but efficient.

ii. If autocorrelation is present, the conventional t and F-tests are valid for carrying out hypothesis testing.

iii. If residuals estimated from an OLS regression exhibit a systematic pattern, it always means that autocorrelation is present in the data.

iv. The Durbin-Watson test statistic (d-statistic) can be used to detect any order of autocorrelation.

v. The Breusch-Godfrey LM test can be used to detect first order autocorrelation only.

b) The following consumption function was estimated using data for the US economy:

$$\widehat{log(CONS)}_t = -0.4677 + 0.8049log(DPI)_t + 0.2013log(W)_t$$
$$(0.017498) \qquad (0.017593)$$

$R^2 = 0.999$, n=54, SER=0.011934, RSS=0.0.007121, DW =0.91013

(Figures in brackets are the standard errors of the coefficients, R^2 is the coefficient of determination, n is the sample size, SER is the standard error of the regression, RSS is the residual sum of squares and DW is the Durbin Watson statistic.)

where, log is natural logarithm, *CONS* is real consumption expenditure, *DPI* is real disposable personal income, and *W* is real wealth. Answer the following questions:

i. Explain why the above model may be affected by autocorrelation. (8%)

ii. What are the consequences for OLS estimation if the above model suffers from autocorrelation? (8%)

iii. Test for the presence of first order autocorrelation in the above estimated equation, clearly stating the null hypothesis being tested and its alternative, the steps involved in the test procedure and the conclusions that you reach. (24%)

c) In estimating a regression equation given by $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ using quarterly data, suppose we know that autocorrelation is present and the error term u_t is generated by the following mechanism:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \epsilon_t$$

where ε_t is a white noise error term. Now, answer the following questions:

i. What is the order of autocorrelation suggested by the above equation? (8%)

ii. Why can one not use the Durbin-Watson test to test for autocorrelation in the above model? What are the main limitations of the Durbin-Watson test statistic in detecting autocorrelation? (16%)

iii. On carrying out the Breusch-Godfrey LM test to test for autocorrelation of the order identified above, suppose you obtain a value of 5.461 for the test statistic. Write down the null and alternate hypothesis of the test and the conclusions you draw based on this given value of the test statistic. (16%)

7.

a) Discuss what is meant by multicollinearity. Outline at least three possible causes of multicollinearity. What are the implications of multicollinearity for OLS estimation? Discuss some possible remedial measures when you encounter the problem of multicollinearity. (48%)

Please turn over ....

b) Now consider the following three OLS regressions based on time series data:

Model A: $\widehat{Y} = 24.7747 + 0.9415X_{2i} - 0.0424\,X_{3i} \quad R^2 = 09635; n = 35$
(6.7525) (0.8229) (0.0807)

Model B: $\widehat{Y} = 24.4545 + 0.5091X_{2i} \quad R^2 = 09621; n = 35$
(6.4138) (0.0357)

Model C: $\widehat{Y} = 24.4110 + 0.049X_{3i} \quad R^2 = 09567; n = 35$
(6.8740) (0.0037)

(Figures in brackets are the standard errors of the coefficients, R^2 is the coefficient of determination, and n is the sample size.)

Suppose you know that both X_2 and X_3 are trending upwards over the sample period. Is there any evidence in the econometric output stated above that would lead you to suspect that Model A is affected by multicollinearity? (NB, report all t-ratios/statistics, F-ratios/statistics to support your answer)

(52%)

8.

a) Explain what is meant by heteroscedasticity. Describe at least three possible causes of heteroscedasticity. Discuss the consequences of heteroscedasticity and the implication for OLS estimation. Outline some possible remedial measures when you encounter the problem of heteroscedasticity.

(48%)

b) Consider the following model that attempts to explain weekly food expenditure:

$$Food_Exp = \beta_0 + \beta_1 inc + u$$

where, *inc* is weekly income (in pounds sterling). Now, if it is known that the conditional variance of the error term in the above model is proportional to the square of income, i.e. $var(u|inc) = \sigma^2 inc^2$, is it possible to transform the model so that it has a homoscedastic error term?

(20%)

c) A sociologist, to explain crime rates across 50 (n) US states runs the following regression model:

$$CRIMEi = \beta_0 + \beta_1 Educ_i + \beta_2 Income_i + u_i$$

where, *CRIME* is the number of crimes committed per thousand people, *Educ* is years of schooling, and *Income* is GDP per capita (in dollars). Now answer the following questions:

i. Explain why might this model be affected by heteroscedasticity?
(12%)

Please turn over ....

ii. In order to test for heteroscedasticity, the sociologist estimates an auxiliary regression where she regresses the squares of the estimated residuals ($\hat{u}_i^2$) from the above model on Educ, Educ^2, Income, Income^2, and (Educ*Income)$_i$. Based on this, she obtains an R^2 value of 0.3094.

What is the name of this test? Outline the null and alternate hypotheses and the steps involved in carrying out this test. What conclusions would the sociologist draw about heteroscedasticity in the original model on crime rates?

(20%)

END OF PAPER