2023

ECONOMICS — HONOURS

Paper : CC-10

(Introductory Econometrics)

Full Marks : 65

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Group - A

1. Answer ***any ten*** questions : 2×10

(a) Determine the type of data of the following variables :

(i) Defence expenditure in each of the Asian countries in 2020.

(ii) Yearly automobile production in India, Japan and China in the time period 1990 to 2022.

(b) State two reasons behind inclusion of error term in an econometric model.

(c) What is the most salient difference between economic model and econometric model?

(d) Mention two assumptions of CLRM.

(e) Distinguish between ex-ante forecast and ex-post forecast.

(f) Which of the following statements is *false*?

(i) Heteroskedasticity does not cause bias in OLS estimator

(ii) Heteroskedasticity does not cause inconsistency in OLS estimator

(iii) Heteroskedasticity does not cause inefficiency in OLS estimator.

(g) State whether the following statements are *true* or *false* :

(i) If the correlation coefficient between two variables is zero, it means that there is no relationship between the two variables.

(ii) In presence of multicollinearity the regression model exhibits very low coefficient of determination and very high individual slope coefficients.

(h) On the basis of a sample survey, the 95% confidence interval for the mean systolic blood pressure for the employees of a firm is computed to be (122, 138). Which of the following statements gives a valid interpretation of this interval?

(i) 95% of the sample employees of a firm have a systolic blood pressure between 122 and 138.

(ii) 95% of the employees in the company have a systolic blood pressure between 122 and 138.

(iii) If the sampling procedure were repeated 100 times, then approximately 95 of the sample means would be between 122 and 138.

(iv) If the sampling procedure were repeated 100 times, then approximately 95 out of resulting 100 confidence intervals would contain the true mean systolic pressure for all employees of the company.

Which distribution will the sample mean in this case follow?

(i) Consider the following regression results and fill in the blanks :

$\text{Mean-wage}_i = 0.7437 + 0.6416\ \text{Education}_i$

SE = (0.8355) (?)

t = (?) (16.6)

(j) Show that the least square estimate of the slope coefficient of a two variable CLRM is unbiased.

(k) What do you mean by BLUE property of an estimator?

(l) What is dummy variable trap?

(m) What are the limitations of simple regression model?

(n) The following ANOVA table is provided :

ANOVA Table

Source of variation	Sum of Squares	Degrees of Freedom
ESS	139023	1
RSS	236894	53

Determine the value of F-statistic.

(o) Suppose we plot the Consumer Price Index (CPI) on the vertical axis and Wholesale Price Index (WPI) on the horizontal axis. *A priori* what kind of relationship do you expect between the two indices and why?

Group - B

Answer ***any three*** questions.

2. (a) State one reason behind heteroskedasticity.

(b) What are its consequences?

(c) Mention a test for heteroskedasticity.

(d) Which distribution does this test statistics follow? 1+2+1+1

3. The correlation between the number of nesting birds and the birth rate is $r = 0.62$. However, the area per inhabitant correlates both with the number of nesting birds ($r = 0.58$) and with the birth rate ($r = 0.92$). Find the correlation between number of nesting birds and the birth rate eliminating the effect of area per inhabitant. 5

4. Consider the following regression equation :

$$\log C_t = -0.4677 + 0.8049 \log DPI_t + 0.2013 \log W_t - 0.0027\, r_t$$

p-value : (0.2) (0.04) (0.51) (0.001)

where C = consumption expenditure

DPI = Real Disposable Personal Income

W = Real Wealth

r = Interest Rate

R squared = 0.9995; Adjusted R squared = 0.999533; n = 55.

Durbin-Watson statistics = 1.2892

Corresponding to 5% level of significance, n = 55 and no. of Regressor (k) = 3, the critical d-values are d_L = 1.452 and d_U = 1.681.

(a) How would you interpret the slope coefficient of interest rate (r) from the estimated equation?

(b) On the basis of the given information comment whether autocorrelation is present and if so, what is the nature of autocorrelation?

(c) State two drawbacks of Durbin-Watson test. 1½+1½+2

5. Consider the following regression equation :

$$\text{Consumption}_i = \beta_1 + \beta_2\ \text{Income}_i + \beta_3\ \text{Wealth}_i + u_i$$

where $\text{Income}_i = a + b\ \text{Wealth}_i$

(a) What problem will you face in estimating the regression equation?

(b) State any one method that solves this problem. 3+2

6. Suppose you want to study the behaviour of sales of automobile over a number of years and somebody suggest you to try the following models:

$Y_t = \beta_0 + \beta_1 t$ ..... Model I

$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$ ..... Model II

where Y_t = sales at time t and t = time, measured in years

(a) Comment on the linearity of the two models.

(b) How does the slope coefficients of the two models differ from one another?

(c) How would you decide between the two models? 1+2+2

Group - C

Answer *any three* questions.

7. (a) Consider the following estimated regression equation :

$\hat{Y} = 233621.5 + 47.99 X_1 + 9.95 X_2$

$SE = (1250364) \quad (7.06) \quad (0.98)$

$R^2 = 0.981065$; $n = 151$

where Y = output; X_1 = labour; X_2 = capital

(i) Interpret the regression coefficients.

(ii) Which of the individual slope coefficients are significant at 1% level of significance?

(iii) Test the null hypothesis that all the coefficients are simultaneously zero against the alternative hypothesis that they are not simultaneously zero. [Given $F_{0.01,(2,\ 148)} = 4.71$]

(b) State the Gauss Markov Theorem. (3+2+3)+2

8. (a) Consider the following model :

$\hat{Y}_i = -0.2610 - 2.3606\ D_{2i} + 0.8028 X_i$

$SE = (1.1073) \quad (0.4302) \quad (0.08101)$

$R^2 = 0.2032$; $n = 528$

where Y = hourly wage; X = Education (years of schooling)

$D_{2i} = 1$ if female, 0 otherwise.

(i) Interpret the meaning of the coefficients in the above model.

(ii) Test in the context of above model, whether there is any difference in hourly wage between males and females.

(iii) Test in the context of above model, whether hourly wage is significantly related to years of schooling. (Given $t_{0.01,525} = 2.576$)

(b) If $r_{12} = 0.80$; $r_{13} = 0.64$; $r_{23} = 0.79$, find the value of $R^2_{1.23}$ (3+2+2)+3

9. (a) Suppose the following model was estimated $Y_i = \alpha + \beta X_i + U_i$

where the variance error term takes the form $E(U_i^2) = \sigma^2 X_i^2$

(i) Does this model satisfy the classical linear regression assumptions? Why?

(ii) What remedial measure do you suggest?

(b) Why is the adjusted R^2 a better measure of goodness of fit compared to R^2? (2+4)+4

10. (a) The following equation relates housing price (price) to the distance from a recently built garbage processing unit (dist):

$\log(\text{price}) = 8.05 + 0.365 \log(\text{dist})$

$n = 142$; $R^2 = 0.180$

(i) Interpret the coefficient of log (dist)

(ii) Does the sign of the estimated slope coefficient match with what you expect it to be? Why?

(iii) Can you state two more factors that may affect price of a house?

(b) Consider the following two variable CLRM :

$$Y_i = \alpha + \beta X_i + U_i,$$

where $n = 7$; $\overline{X} = 4$; $\overline{Y} = 11$; $\sum X_i^2 = 140$; $\sum Y_i^2 = 875$; $\sum X_i Y_i = 334$

Obtain the estimated regression equation. (2+2+2)+4

11. (a) What do you mean by functional form misspecification? Give an example.

(b) Is it better to include irrelevant variables than to omit the relevant ones? Justify your answer.

(c) Mention a test used to detect presence of specification error in estimating regression equations. What is the null hypothesis associated with this test? Which distribution is followed by this test statistics? (2+1)+4+3
