QP Code : 24/PT/11/XXB(P)

POST-GRADUATE DEGREE PROGRAMME
Term End Examination — December, 2024
ECONOMICS
Paper-XXB : ECONOMETRIC APPLICATION (PRACTICAL)

Time : 2 hours] [Full Marks : 50

Weightage of Marks : 100%

Special credit will be given accuracy and relevance in the answer. Marks will be deducted for spelling, untidy work and illegible handwriting. The weightage for each question has been indicated in the margin.

Use of scientific calculator is strictly prohibited.

1. Answer any *four* of the following questions : $2\frac{1}{2} \times 4 = 10$

   a) After estimating a regression, you find the residuals are not normally distributed. What impact does this have on inference ? What tests would you use to check for normality ?

   b) A multiple regression output gives $R^2 = 0 \cdot 95$ but, some of the coefficients are statistically insignificant. What does this indicate ? Should you rely on this model ?

   c) The estimated model gives a Durbin-Watson statistic of 0·92. Interpret the results. What step should you take ?

   d) A correlogram of a time series variable shows a sinusoidal pattern. What does this imply about the underlying process ? How would you proceed with modeling it ?

   e) In PCA, explain the difference between eigenvalues and factor loadings. How does one decide the number of principal components to retain ?

   f) Interpret the coefficient of a dummy variable in a regression model. What does the sign and significance imply ?

2. Answer any *four* of the following questions : $5 \times 4 = 20$

   a) Distinguish between fixed effects and random effects models in panel data analysis.

   b) In a study estimating the effect of education on earnings, why might the education variable be endogenous ?

c) A regression output shows as follows :

$\log(\text{wage}) = 5{\cdot}8 + 0{\cdot}35\ lm(\text{education}) + 0{\cdot}6(\text{experience}) - 0{\cdot}01\ \text{experience}^2$

i) Interpret each coefficient

ii) Why is the square of experience included ? 4 + 1

d) You are analysing GDP data over the past 30 years and want to use it in regression model. Why is it important to test for stationarity and how can the Augmented Dickey-Fuller [ADF] test help in this context ?

e) You are conducting factor analysis on a mental health survey measuring stress, anxiety and well-being. After extraction, you find that some items have very low communalities. What does this indicate and how should you address it in your analysis ? 3 + 2

f) While analysing household consumption data across income levels you notice that the variance of residuals, increases with income. What econometric issue does it suggest ? How would you test for it ? 4 + 1

3. Answer any *two* of the following questions : $10 \times 2 = 20$

a) A researcher is studying monthly Consumer Price Index (CPI) inflation data for India from January 2018 to December 2023 to forecast inflation for the next 6 months. After ensuring stationarity, the following ARIMA model is selected :

ARIMA (1, 1, 1) = $\Delta \text{CPI}_t = \phi_1 \Delta CPI_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$.

The estimated model output is

Parameter	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
AR (1)	0·52	0·11	4·73	0·00
MA (1)	– 0·44	0·10	– 4·40	0·00
Sigma^2	1·22	—	—	—

The AIC = 245·8, and residual diagnostics (Ljung – Box Q – statistics) suggest no significant autocorrelation left in residuals up to log 12.

Given the following information answer the following :

i) Explain the steps the researcher would have taken to arrive at this ARIMA (1, 1, 1) model. Include discussion on stationarity testing and model selection. 6

ii) Interpret the meaning and sign of the AR(1) and MA (1) coefficients in the context of inflation behaviour in India. 2

iii) What does the "1" in the middle of ARIMA (1, 1, 1) signify ? Why is differencing needed ? 2

b) Given the data for 10 students study hour (X) and test score (Y).

***X* (Study Hour) :**	4	6	5	7	9	3	8	2	10	11
***Y* (Test Score) :**	60	66	63	70	78	57	74	53	82	85

i) Estimate the regression line $Y = a + bX$

ii) Find R^2

iii) Calculate Standard Errors (a & b).

[Note $\sum X^2 = 505$; $\sum Y^2 = 4800$ and $\sum XY = 4890$ 5 + 2 + 3

c) A researcher uses NSSO data on salaried workers in urban India to estimate the following regression.

$$ln\,(\text{wages}) = \beta_0 + \beta_1 \text{Experience} + \beta_2 \text{Female} + \beta_3 (\text{Experience} \times \text{Female}) + \mu$$

Where *ln* (wage) = Natural logarithm of monthly wage.

Experience = Years of job experience.

Female : Dummy variable (1 is female, 0, otherwise)

The estimated output is as follows:

Variable	**Coefficient**	**Std. Error**	***t*-statistic**	***p*-value**
Intercept	8·100	0·200	40·50	0·000
Experience	0·045	0·008	5·63	0·000
Female	– 0·120	0·050	– 2·40	0·018
Experience × Female	– 0·015	0·006	– 2·50	0·013

Answer the following :

i) Interpret each of the coefficients in the context of urban labour market in India.

ii) What does the interaction term (Experience × Female) represent in the model ?

iii) Compute the marginal effect of experience on *ln*(wage) for both male and female.

iv) What does the negative sign of both β_2 and β_3 imply about gender disparities in wages and experience returns ?

2 + 2 + 3 + 3

d) A study investigates the probability that an individual uses digital banking (1 = Yes; 0 = No) based on their income (in Rs '000 per month). A logit model is estimated as follows :

$$\text{logit}(P) = ln\left(\frac{P}{1-P}\right) = -3{\cdot}5 + 0{\cdot}08 \text{ Income}$$

Where P is the probabilities of using digital banking.

Answer the following :

i) Explain why a logit model is used instead of OLS in this case.

ii) Interpret the coefficient of income in this model.

iii) Compute the probabilities of using digital banking for an individual earning Rs. 40,000 per month.

iv) What happens to this probability when income increases to Rs. 50,000 per month ? Comment on the result. 2 + 2 + 3 + 3