**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**
**FIRST SEMESTER - 2023-2024**

**Course No.: ECON F241** **Course Name: ECONOMETRIC METHODS**
**Date : 12/Oct/2023** **Max Marks: 35**
**MID SEMESTER EXAMINATION (REGULAR)**

**ID NO.** **NAME:**

NOTE: The exam is closed book and closed notes. Calculators are allowed. This exam has two Parts PART A and PART B. Attempt all parts in the same answer sheet. Write legibly. Write the necessary assumptions if any.

Write the Id NO and Name on all pages.

---

**PART –A: Short Answer Questions**

A1 **For the following MULTIPLE CHOICE questions, choose the one correct best answer to each question and put a tick (√) against that letter A/B/C/D. Corrections/overwriting/illegible answers are strictly invalid. (10 x ½ = 5 Marks)**

1) In a linear regression, a residual plot shows a sine pattern. This says that this linear regression model
A. fits data well.
B. does not fits data well.
C. both A) and B) are wrong
D. None of the above

2) You are using a simple classical linear regression model and the OLS estimator to analyze the relationship between the hourly wage rate and years of education. You have reason to believe your error term has non- constant variance. Which of the following would you expect to be the consequence of this situation?
A. Your study will not be internally valid because your estimate of the effect of education will be biased.
B. Your study will not be internally valid because your standard error of the estimate of the effect of education will be incorrect, and therefore any inferences you make or hypotheses you test may be incorrect.
C. Your study will be internally valid but not externally valid, because it will not generalize to the population in which you are interested.
D. Your study will be both internally and externally valid, because non-constant error variance does not affect the validity of your conclusions.

3) You specify a simple classical linear regression model. The dependent variable is a measure of the health status for the tth state in India. The explanatory variable is a measure of medical care spending per capita for the tth state.
Consider the following argument: States with healthy populations spend less on medical care than states with unhealthy populations. If this argument is valid, then which of the following assumptions of the simple classical linear regression model will not be valid?
A. The error term is not correlated with the explanatory variable.
B. The error term has constant variance.
C. The errors are independent.
D. The error term has a normal distribution.

4) The Gauss-Markov Theorem tells us that, under appropriate assumptions, the least squares estimator of β in the usual linear regression model:

A. Is a linear estimator, and therefore is "best"?
B. Has the smallest bias among all possible linear estimators for this parameter vector?
C. Is most efficient among all possible linear and unbiased estimators of this parameter.
D. Is most efficient among all possible unbiased estimators that have a Normal sampling distribution.

5) All of the following are possible effects of multicollinearity EXCEPT

A. the variances of regression coefficients estimators may be larger than expected
B. the signs of the regression coefficients may be opposite of what is expected
C. a significant F ratio may result even though the t ratios are not significant
D. the VIF is zero

6) Which of the following options is not a source of Multicollinearity :

A. Data collection method employed
B. Model specification
C. Over determined model
D. Sampling technique

7) By including another variable in the regression, you will

A decrease the variance of the estimator of the coefficients of interest.
B decrease the regression $R^2$ if that variable is important.
C eliminate the possibility of omitted variable bias from excluding that variable.
D look at the t-statistic of the coefficient of that variable and include the variable only if the coefficient is statistically significant at the 1% level

8) Sample selection bias

A occurs when a selection process influences the availability of data and that process is related to the dependent variable.
B is more important for nonlinear least squares estimation than for OLS.
C is only important for finite sample results.
D results in the OLS estimator being biased, although it is still consistent

9) What is the meaning of the term "heteroscedasticity"?

A. The variance of the errors is not constant
B. The variance of the dependent variable is not constant
C. The errors are not linearly independent of one another
D. The errors have non-zero mean

10) What would be then consequences for the OLS estimator if heteroscedasticity is present in a regression model but ignored?

A. It will be biased
B. It will be inconsistent
C. It will be inefficient
D. All of (a), (b) and (c) will be true.

**A2** **Determine whether the following statements are TRUE or FALSE. Note that Answers without proper reasoning receive no credit. (5 x 1.0 = 5 Marks)**

11) R-squared always increases when we add variables to a regression.

12) I have the estimated slope coefficients from the regression of Y on $X_2$ and from the regression of Y on $X_3$. I also have the slope coefficient from the regression of $X_2$ on $X_3$, so I can compute the slope coefficients that would result from the multiple regression of Y on $X_2$ and $X_3$.

13) "Increasing sample size changes the sampling error variance of a regression and leads to larger standard errors for the estimated coefficients."

14) Heteroscedasticity leads to an unbiased estimator of $\beta_1$ cap and hence the t-test based on $\beta_1$ cap and s.e. of (β1 cap) and is valid. (s. e is Standard Error)

15) Heteroscedasticity is the best ascertained by plotting the residuals.

**A3** **For the following Short Questions write answer to the point.**

**(5 x 2.0: 10 marks)**

1. Are the following valid null hypotheses in statistical testing? Explain.
   (A) $\mu_X = 100$;
   (B) X bar = 100.

2. What do the no perfect multicollinearity and zero conditional mean assumption mean in the basic assumption of linear regression model? What are the consequences if they are violated?

3. What are the four factors that affect the variance of individual slope coefficient estimators under OLS for a multiple regression model? How do they affect the variance?

4. Because omitted variables cause bias, it is always recommended to include all available explanatory variables in a regression.

5. Given the following estimated model (standard errors in parentheses)

$Y_i = -2.46 + 6.11\ X_{2i} - 1.78\ X_{3i}$

(0.94) (2.80) (1.47) N=40, $R^2 = 0.43$, RSS = 287.2

$$\hat{u}_i^2 = 4.2 + 1.24X_{2i} + .862X_{3i} + .743X_{2i}^2 + 3.86X_{3i}^2 + .065X_{2i}X_{3i},$$
(0.44) (0.31) (0.55) (0.71) (1.23)

And the regression for the test is

N=40, $R^2 = 0.25$, RSS = 127.2

Use the relevant test statistic and test that the hypothesis errors are homoscedastic at 5% level (The critical value is 11.7 at 5% level of significance)

**PART – B: Long Answer Questions** **(15.0 Marks)**

For the following Long Questions, read the question carefully and attempt all parts of the question. write answer to the point.

B1 Following Leamer's example, you estimate the demand for tangerines (substitute one for oranges) in the country China over the period 1962 to 1991 and obtain the following results (standard errors in parentheses):

(1) $\ln Y_t$ = 6.18 − 0.52 $\ln PT_t$ + 0.61 $\ln X_t$ − 0.23 $\ln PO_t$ $R^2 = 0.74$

(1.38) (0.29) (0.25) (0.18) RSS = 23.19

where Y is the consumption of tangerines in thousands of bags, PT is the price of tangerines in Rs. per bag, X is per capita income in Rs.1000, and PO is the price or oranges in Rs. per bag. (Note that oranges are a substitute for tangerines.)

Using the same data, you also get the following results:

(2) $\ln Y_t$ = 6.62 − 0.71 $\ln(PT_t/PO_t)$ + 0.49 $\ln(X_t/PO_t)$ $R^2 = 0.68$

(1.23) (0.30) (0.19) RSS = 28.54

**(4.0)**

a) What is it about equation (1) that leads you to suspect multicollinearity?

b) Test the hypothesis that the restriction is statistically valid at the significance level of 5%? Show all calculations. (The critical or table value at 5% level of significance is 4.23.)

B2 In a study of population density as a function of distance from the central district, based on a sample of 39 census tracts in the Gujarat State area in 1980, gives (standard errors in parentheses):

(1) $\ln Y_i = \underset{(0.185)}{10.093} \; \underset{(0.019)}{- 0.239 X_i}$ $\qquad R^2 = 0.803$

(2) $\ln Y_t/(\sqrt{X_t}) = \underset{(0.208)}{9.923 (1/\sqrt{X_t})} \; \underset{(0.015)}{- 0.2253 \sqrt{X_t}}$ $\qquad R^2 = 0.860$

where Y is population density and X is distance from the central district in miles.

a) What assumption is the author making about heteroscedasticity?
b) How would you decide if the transformation in equation (2) removed the heteroscedasticity? Briefly explain your answer.
c) How do you interpret the regression results? Do they make sense?

**(4.0)**

B3 You are investigating the relationship between output growth and investment in the Delhi magisterial district across industrial sectors. You have measured the average annual growth rate for the years 1980-1996 and you have calculated new investment in 1980 as a proportion of the overall value of the company's capital in 1980. A regression of growth on investment yields the following (edited) regression output by a computer package:

```
. fit growth invest

  Source |       SS                                  Number of obs =      17
---------+----------------                           F(  ,    ) =
   Model | .010085059                                Prob > F      =
Residual |                                           R-squared     =
---------+----------------                           Adj R-squared =
   Total |   .55067797                               Root MSE      =

------------------------------------------------------------------------------
  growth |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  invest |  -.2247175   .4248027
   _cons |   .2638587                2.838   0.012       .0656945    .4620229
------------------------------------------------------------------------------
```

You are also told that the average investment level in the industrial sectors was 0.15.

**(7.0)**

a) Interpret the regression coefficients.

b) From the data given above calculate the standard error of the constant in the regression.

c) Determine whether the regression coefficients are significant. (the critical value of 2.056 at the 5% level of significance)

d) Calculate the $R^2$ the coefficient of determination for this regression and also F- Statistic. Comment.

e) Estimate $\sigma^2$ (the variance).

f) According to the model, what would you expect the growth rate of an industrial sector to have been if its rate of investment was 0.1 (i.e. 10%)?

g) Calculate a 95% confidence interval for the prediction you made in question 1.f.

(Hint use the definition of var($\hat{\beta}_2$) ). (the critical value of 2.056)