BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
FIRST SEMESTER 2023-24

COURSE NO: ECON F241 **COURSE NAME: ECONOMETRIC METHODS**
DATE: 13 DEC 2023 **TIME: Min 180**
Weightage 35%

COMPREHENSIVE EXAMINATION (REGULAR)

- Write your Name and ID No on your all pages of answer sheets.
- Attempt all questions.
- You may use a calculator if you wish. Cell phones or any other electronic and communication devices are not allowed. Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely.
- This exam paper consists of two parts: PART – A: CLOSED BOOK (20 Marks) and PART –B: OPEN BOOK (15 Marks).
- Write the assumptions, if any, clearly. Neat presentation and careful notation are very important in the answers.
- Answer to the point and show your work. Start answering all parts of a question at one place. Write legibly. Illegible answers carry no weightage. Clearly indicate your final answer to each question.
- After completing the PART-A – Closed Book examination submit the answer sheet and start answering the PART-B (Open Book).
- **Note that you have 180 minutes to complete the exam.**

PART- A (CLOSED BOOK) **Max. Marks.: 20 Marks**

Answer each question as clearly and concisely as possible on the answer sheet.

Questions A1 to A8 each question carries 1.0 mark and question A9 carries 2.0 mark. Question A10 is for 10 marks.

A1. Let X and Z be two independently distributed standard normal distributions and let $Y= X^2+ Z^2$. Compute $E(Y^2)$?

A2. 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?

A3. A researcher runs a regression with a sample of 33 observations and 17 explanatory variables (including the intercept, which is included in k). Her coefficient of determination R^2 is 0.70. What is her adjusted R^2?

A4. Consider the following regression equation for the imports of the INDIA over the period 1970-1983: ln (Imports)t = 0.6754 + 0.3711 ln (GNP)t + 1.5855 ln (CPI)t

	(0.939)	(0.0435)
t =	(3.951)	(36.458)

R^2 = 0.9962 Adjusted R^2 = 0.9955 $r_{\ln(GNP), \ln(CPI)}$ = -0.755

a) Do you suspect that there is severe multicollinearity in the model?
b) Suppose there is severe multicollinearity in the data but estimated coefficients are individually significant at the 5 per cent level and the overall F-test is also significant. In this case should we worry about the collinearity problem?

A5. With reference to the following regression using 45 observations:

$$\hat{y}_t = 0.98 + 0.56x_t$$
$$(0.32) \quad (0.14)$$
$$DW = 1.52, R^2 = 0.4$$

Is there any evidence of 1st order autocorrelation? The Durbin Watson table values are 1.48 and 1.57

A6. Given a sample of data containing Y, X_1, X_2, and X_3 can you use Ordinary Least Squares to estimate the parameters (β1, β2, and β3) of the following production functions? If so, discuss how. If not, discuss why it is not possible.

i) $Y = a\, X_1^{\beta 1}\, X_2^{\beta 2}\, X_3^{\beta 3}$

ii) $Y = (a\, X_1^{\beta 1}\, X_2^{\beta 2}\, X_3^{\beta 3})^{1/3}$

iii) $Y = a + (\beta_1 X_1)^2\, (\beta_2 X_2)^{0.25}$

iv) $Y = [a\, .\, \exp(\beta_1 X_1) * \exp(\beta_2 X_2^3)] / [\exp(\beta_3 X_3^{½})]$

A7. Given the following estimated model (standard errors in parentheses)

$$\underset{}{Y_i} = \underset{(0.94)}{-2.46} + \underset{(2.80)}{6.11}\, X_{2i} - \underset{(1.47)}{1.78}\, X_{3i} \qquad N=40,\ R^2 = 0.43,\ RSS = 287.2$$

And the regression for the test is

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

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

<u>Use the relevant test statistic and test the hypothesis that errors are homoscedastic at 5% level</u>. Table value of test statistic is 11.70.

A8. An investigator estimated the parameters in the equation : $\ln Y_t = \alpha + \beta \ln X_t + u_t$

by Ordinary Least Squares using 52 quarterly observations for 1972 to 1984 inclusive. This resulted in a residual sum of squares (RSS) of 0.78.

When 3 dummy variables representing the first 3 quarters of the year were added to the equation the RSS fell to 0.56. Using an appropriate test statistic, test for the presence of seasonality stating what assumptions are being made concerning the form of the seasonality. The critical value of test statistic is 2.76.

A9. Assess the following statements as TRUE or FALSE. Give a short explanation or qualification. If a statement is not true in general, but is true under some conditions, state the conditions.

a) The coefficient of determination (R^2 or r^2) of a two-variable linear regression [$Y_i=\beta_1 + \beta_2 X_i + u_i$] is equal to the square of the sample correlation coefficient between Y and X.

b) The following is the result of regressing the natural logarithm of the quantity of housing consumed per year (*HOUSE*), measured in square feet, on the price of one unit of housing(*P*), measured in rupees in Gurugaon, Haryana state,
ln(*HOUSE*) cap = 4.17 - 0.907P. It follows that when the price of housing increases by Rs.10, the quantity of housing consumed falls by 9.07 square feet.

A10. For the following multiple choice questions <u>choose the correct best answer (A/B/C/D) and put a tick against that letter</u> and also <u>write the corresponding letter in the space provided below. Correction/overwriting/illegible answers are invalid and carry no weightage.</u>

(10.0 Marks)

1. Two random variables X and Y are independently distributed if all of the following conditions hold, with the exception of
   A. $\Pr(Y = y \mid X = x) = \Pr(Y = y)$.
   B. knowing the value of one of the variables provides no information about the other.
   C. if the conditional distribution of *Y* given *X* equals the marginal distribution of *Y*.
   D. $E(Y) = E[E(Y \mid X)]$.

2. Suppose we ran a regression of Y on X 1000 times using 1000 different samples and made a histogram of the resulting slope coefficient values. Which of the following is true about the distribution shown on the histogram?
   A. It would be centered at zero.
   B. It would look like a normal distribution.
   C. All of the observations would be located at the true value of the slope coefficient.
   D. It would be right skewed.

3. Which of the following assumptions are required to show the consistency, unbiasedness and efficiency of the OLS estimator?
   i) $E(u_t) = 0$
   ii) $Var(u_t) = \sigma^2$
   iii) $Cov(u_t, u_{t-j}) = 0 \; \forall \; j$
   iv) $ut \sim N(0, \sigma^2)$

   A. (ii) and (iv) only
   B. (i) and (iii) only
   C. (i), (ii) and (iii) only
   D. (i), (ii), (iii) and (iv) only

4. Which of the following depends on the units variables are measured in?
   A. Correlation.
   B. Coefficient of variation.
   C. Estimated slope coefficient.
   D. t statistic.

5. If the residuals from a regression estimated using a small sample of data are not normally distributed, which one of the following consequences may arise?
   A. The coefficient estimates will be unbiased but inconsistent
   B. The coefficient estimates will be biased but consistent
   C. The coefficient estimates will be biased and inconsistent
   D. Test statistics concerning the parameters will not follow their assumed distributions.

6. 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.

7. Suppose we use an F test after running a multivariate regression to test the null hypothesis that $\beta_3 = \beta_4 = 0$ and get an F statistic that is larger than the critical value for a 5% significance level. We would conclude that:
   A. $\beta_3 \neq \beta_4$
   B. $\beta_3 > 0$ or $\beta_4 > 0$.
   C. $\beta_3 \neq 0$ and $\beta_4 \neq 0$.
   D. None of the above.

8. Which of the following would make you more likely to reject the hypothesis that an individual slope coefficient is equal to zero?
   A. A larger standard error for that slope coefficient.
   B. A smaller t statistic for that slope coefficient.
   C. A smaller F statistic for the regression.
   D. A larger value for the ratio of the coefficient to its standard error.

9. Suppose that the value of R^2 for an estimated regression model is exactly zero. Which of the following are true?
   (i) All coefficient estimates on the slopes will be zero
   (ii) The fitted line will be horizontal with respect to all of the explanatory variables
   (iii) The regression line has not explained any of the variability of *y* about its mean value
   (iv) The intercept coefficient estimate must be zero.

   A. (ii) and (iv) only
   B. (i) and (iii) only
   C. (i), (ii) and (iii) only
   D. (i), (ii), (iii) and (iv) only

10. Omitted variable bias
   A will always be present as long as the regression $R^2 < 1$.
   B is always there but is negligible in almost all economic examples.
   C exists if the omitted variable is correlated with the included regressor but is not a determinant of the dependent variable.
   D exists if the omitted variable is correlated with the included independent variable(s) and is a determinant of the dependent variable.

11. Consider the following regression equation: $Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}^2 + u_i$
   A. This equation is nonlinear in the parameters.
   B. This equation is nonlinear in the variables.
   C. This equation is linear in the variables.
   D. None of the above

12. With heteroskedastic errors, the weighted least squares estimator is BLUE. You should use OLS with heteroscedasticity-robust standard errors because
   A the Gauss-Markov theorem holds.
   B this method is simpler.
   C Both A and B
   D the exact form of the conditional variance is rarely known.

13. Which test for serial correlation is still valid when the regressors in the original equation include a lagged value of the dependent variable—e.g., yt = β1 + β2 xt + β3 yt-1 + εt ?
    A. A t-test from a regression of the least-squares residual on its lag (without an intercept).
    B. The Durbin-Watson test.
    C. An alternative test," based on an auxiliary regression.
    D. Goldfeld-Quandt Test

14. Suppose that the variance of a regression model's stochastic error term is given by: $\mathrm{VAR}(\varepsilon_i) = \sigma^2\sqrt{Z_i}$, where Z_i is some variable (possibly one of the models independent variables). Then, to make the models error term homoskedastic, it would be necessary to
    A. divide the equation for the model through by $Z_i^{0.25}$.
    B. divide the equation for the model through by Z_i^2.
    C. divide the equation for the model through by $Z_i^{0.5}$.
    D. divide the equation for the model through by Z_i^4.

15. A student considers applying the Durbin-Watson d statistic to test $H_0 : \rho = 0$ versus $H1 : \rho \neq 0$ in a MLR model where the error terms may potentially exhibit an AR(1) structure:
    $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$. Denote the test statistic as d. Let d_L and d_U be usual lower and upper critical values. Which of the following statements about the rejection rule is right?
    A If $d > d_U$, then we can reject the null.
    B If $d < d_L$, then we can't reject the null.
    C If $d_L \leq d \leq d_U$, we can't reach any conclusions about serial correlation.
    D If $d_U \leq d \leq 4 - d_U$, we can't reach any conclusions about serial correlation

16. Consider the following 2 regression models:
    Model 1: $y_t = \beta_1 + \beta_2 x_{2t} + u_t$
    Model 2: $y_t = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + u_t$
    Which of the following statements are true?
    i) Model 2 must have an R^2 at least as high as that of model 1
    ii) Model 2 must have an adjusted R^2 at least as high as that of model 1
    iii) Models 1 and 2 would have identical values of R^2 if the estimated coefficient on α_3 is zero
    iv) Models 1 and 2 would have identical values of adjusted R^2 if the estimated coefficient on α_3 is zero.
    A. (ii) and (iv) only
    B. (i) and (iii) only
    C. (i), (ii) and (iii) only
    D. (i), (ii), (iii) and (iv) only

17. Which of the following models can be estimated using OLS, following suitable transformations if necessary? (Note that "e" denotes the exponential).
    i. $y_t = \alpha + \beta x_t + u_t$
    ii. $y_t = \alpha + \beta e^{x_t} + u_t$
    iii. $\ln(y_t) = \alpha + \beta \ln(x_t) + u_t$
    iv. $y_t = \alpha + \beta x_t^2 + u_t$.
    A. (i) only
    B. (i) and (iii) only
    C. (i), (iii) and (iv)
    D. (i), (ii), (iii) and (iv) only

18. The RAF at the SWD, BITS Pilani wants to estimate how often students tend to go to the movies each semester. They interview a sample of 25 students and ask how many times they go out to the movies each semester. The average response is 14, and the standard deviation of the responses is 2.8. The table t-value for 95% CI (based on 24 df): 2.064

Construct a 95% confidence interval based on the data given.

A. 14 ± 0.23
B. 14 ± 0.96
C. 14 ± 1.16
D. 14 ± 1.40

19. Which of the following issues does not cause the error term to be correlated with a regressor?

A. A regressor, which happens to be correlated with the included regressors, is omitted from the equation.
B. An independent (X) variable is measured with error.
C. The dependent (Y) variable is measured with error.
D. The wrong variable is used as the dependent variable.

20. Simultaneous causality bias

A. results in biased estimators if there is heteroscedasticity in the error term.
B. happens in complicated systems of equations called block recursive systems.
C. is also called sample selection bias.
D. arises in a regression of Y on X when, in addition to the causal link of interest from X to Y, there is a causal link from Y to X

1	2	3	4	5

6	7	8	9	10

11	12	13	14	15

16	17	18	19	20

--After completing PART –A (Closed Book) submit it and Answer PART-B (Open Book) examination.--

COMPREHENSIVE EXAMINATION (REGULAR)
(PART- B: OPEN BOOK) (15.0 Marks)

NOTE: Attempt all questions. Answer to the point. Show your complete work to get full credit. Attempt all parts of the question at one place.

B1) Consider the following regression model, which we will refer to as "Model A":

$$BEEFCONS_t = \beta_1 + \beta_2 INCOME_t + \beta_3 PBEEF_t + \beta_4 PPORK_t + \beta_5 PCHICKEN_t + \varepsilon_t$$

where: BEEFCONS is per capita India's beef consumption in year t (in Grams)
INCOME is real per capita disposable personal income in year t (in Rupees)
PBEEF is the price of beef (Rupees per Gram) in year t
PPORK is the price of pork (Rupees per Gram) in year t
PCHICKEN is the price of chicken (Rupees per Gram) in year t

Suppose we collect a sample of the 22 years from 1980 to 2001 and use OLS to estimate the parameters of this model and several variants of it:

Variable	Model A	Model B	Model C
Constant	104.06 (9.81)	134.28 (13.16)	70.75 (61.47)
INCOME	0.001 (1.72)	-0.001 (-4.02)	
PBEEF	-15.03 (-3.10)	-21.19 (-3.16)	
PPORK	-9.04 (-4.51)		
PCHICKEN	2.46 (1.79)		
Residual SS	57.23	148.88	612.15

A) For Model A, interpret the estimated coefficients clearly and completely.
B) In Model A, how many Grams of beef would you expect the average person to consume if his or her income were Rs. 30,000, the price of beef were Rs.1.50 per Gram, the price of pork were Rs.3.00 per Gram, and the price of chicken were Rs.4.00 per Gram?
C) Compute and evaluate the general F-statistic for Model A. Be sure to do the following:
Write out the null and alternative hypotheses of the test in terms of the betas
Calculate the numerical value of the test statistic
Specify the degrees of freedom and the critical value for a 99% level of confidence
Explain your conclusion about the test.

B2) The simple linear regression model, $Y_t = \beta_1 + \beta_2 t + u_t$ was estimated on the **log India GDP** quarterly series, 2000-2009 (40 observations). The following table shows the regression output, unfortunately with some values missing:

ANOVA				
Source	Sum of Squares	Degrees of Freedom	Mean Square	F_{Obs}
Regression	**?**	**?**	**?**	?
Residuals	**0.0361**	**?**	**?**	
Coefficients Table				
Variable	Coefficient	SE (βi Cap)	t_{obs}	p-value
Intercept	**9.2255**	**?**	**931.87**	**0.0000**
t	**?**	**0.0004**	**26.50**	**0.0000**

A) Complete the Table (missing values). Compute R^2, Adjusted R Square, $\hat{\sigma}^2$ and variance of (Y_t).

B) It was shown that Corr ($\hat{u}_t$, $\hat{u}_{t-1}$) = **0.8.** Use a suitable test to decide whether the residuals are autocorrelated. If so - what would you do to resolve this problem?

C) In the regression: $\boldsymbol{\hat{u}_t^2 = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + v_t}$, the R^2 was found to be **0.62**. Use a suitable test to decide whether the residuals are heteroscedastic. If so - what would you do to resolve this problem?

D) An analyst estimates two separate simple linear regression models to the first and second part of this decade (2000-2004 and 2005-2009). The RSS for each part were found to be $RSS_{2000\text{-}2004}$ **= 0.0125** and $RSS_{2005\text{-}2009}$ **= 0.0092**; respectively. Use a suitable test to decide whether this indicates a structural change or not.

B3) The following equation describes workers' earnings:

$$W = \beta_1 + \beta_2 Age + \beta_3 Age^2 + \beta_4 Male + \beta_5 Dropout + \beta_6 College + \beta_7 Master + u$$

where *W* is annual earnings (in rupees), *Age* is the worker's age (in years), *Male*, *Dropout*, *College*, and *Masters* are dummy variables, and *u* is a residual.

The dummy variables are defined as: *Male=1* if the worker is male; 0 otherwise.

Dropout=1 if the worker has 0-11 years of schooling; 0 otherwise.

College =1 if the worker has 16-17 years of schooling; 0 otherwise.
Masters =1 if the worker has 18+ years of schooling; 0 otherwise.
In addition, the data set contains a dummy for high school graduates without a completed college degree, *HSgrad* = 1 if the worker has 12 - 15 years of schooling; 0 otherwise.

The estimated equation over a cross-section sample of 560 workers is:

$$\widehat{W} = -16000 + 2346 Age - 23 Age^2 + 12000 Male - 10000 Dropout + 16000 College + 46000 Master$$

A) What is the predicted annual earnings of a 30-year-old man who went to college but not to graduate school?

B) If we replace Dropout with Hsgrad, what would the coefficient estimates be?

C) If we have both Dropout and Hsgrad in the regression, what would we find?

D) Someone states: "If Age increases, so does Age^2. Age and Age^2 are thus strongly positively correlated, and the OLS estimates are biased, inconsistent, and inefficient." Comment.

E) This specification implies that the effects of schooling on wage earnings are the same by sex. Explain in detail how you would explore whether returns to schooling differ by sex. State how you would re-specify this equation and the specific null hypothesis in terms of that equation. Provide the test statistic and its degrees of freedom.

B4) The following model was estimated with time series data using the Ordinary Least Squares (OLS) procedure (with OLS standard errors reported in parentheses):

$$y_t = -6.29 \underset{(0.70)}{} + \underset{(0.07)}{1.45x_t} + \hat{u}_t$$

where y and x are expressed in natural logarithms. The sequence of fifteen residuals, $\hat{u}_t$, obtained from the regression model, ranked from 1990 through to 2004, is given by:

Time (T)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
$\hat{u}_t$	+ 0.016	+ 0.041	- 0.029	- 0.017	- 0.071	+ 0.019	- 0.001	- 0.005	+ 0.002	+ 0.011	+ 0.058	+ 0.002	+ 0.003	+ 0.021	- 0.009

where T is the observation year and $\hat{u}_t$ is the estimated residual for time period t.
From the above outline the possible sources and consequences of autocorrelated errors in a linear regression model. <u>Use the reported information to implement a parametric test for autocorrelated errors in this regression model.</u>
Use a significance level of 0.05 and state clearly the null and alternatives under test. Draw the inference.

B5) Consider the simultaneous-equation model:

$Y_2 = \alpha_1 Y_1 + \alpha_2 X + U_1$,

$Y_2 = \alpha_3 Y_1 + U_2$,

where the exogenous variable X is independent of the disturbances U_1 and U_2, and all variables have zero expectations (for convenience).

A) Solve the equations and derive the algebraic reduced form: $Y_1 = \pi_1 X + V_1$, $Y_2 = \pi_2 X + V_2$
B) You are told that $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$, and asked to predict the value of Y_1 that will occur when X = 1. What would be your prediction? Explain briefly.

****