BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2017-18

Course No.: ECON F241 **Course Title: ECONOMETRIC METHODS**
Date: 03 MAY 2018 **Max. Marks: 40**

COMPREHENSIVE EXAMINATION

NOTE: Attempt all questions. Answer to the point. This paper consists of two parts: PART- A (Closed Book: 20 Marks) and PART-B (Open Book: 20 Marks). Attempt PART-A in the same answer sheet and after completing submit this part and attempt PART-B (Open Book) in the separate answer sheet. Write assumptions if any clearly.

PART- A (CLOSED BOOK) 20 Marks

A1) **For multiple choice questions Choose the correct best alternative and write the corresponding letter (A/B/C/D in the space provided below. Corrections/Overwriting/illegible answers are invalid. Each question carries equal marks (0.50 mark).**

1	2	3	4	5

6	7	8	9	10

11	12	13	14	15

16	17	18	19	20

1 If an estimator is unbiased, then its mean square error (MSE)
A must be greater than its variance.
B must be equal to its variance.
C must be less than its variance.
D must be identically zero.

2 The central limit theorem
A. states conditions under which a variable involving the sum of $Y_1,...,Y_n$ i.i.d. variables becomes the standard normal distribution.
B. postulates that the sample mean $\bar{Y}$ is a consistent estimator of the population mean μ_Y.
C. only holds in the presence of the law of large numbers.
D. states conditions under which a variable involving the sum of $Y_1,...,Y_n$ i.i.d. variables becomes the Student t distribution.

3 If the computed p-value for a test statistic is less than the size of the test, we
A can reject the null hypothesis.
B cannot reject the null hypothesis.
C cannot compute the test statistic.
D answer cannot be determined from the information given.

4 When testing joint hypothesis, you should
A use *t*-statistics for each hypothesis and reject the null hypothesis if all of the restrictions fail.
B use the *F*-statistic and reject all of regression coefficients if p-value (Sig./Prob.) $\leq \alpha$.
C use *t*-statistics for each hypothesis and reject the null hypothesis if p-value (Sig./Prob.) $\leq \alpha$.
D use the *F*-statistics and reject at least one of the regression coefficient if p-value (Sig./Prob.) $\leq \alpha$.

5 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. $E(Y) = E[E(Y \mid X)]$.

6 Suppose we use the least-squares predictor $(\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1})$ to predict y_{n+1} . The variance of the prediction error ($y_{n+1} - \hat{y}_{n+1}$) is smaller,
A the smaller the variance of the error term σ^2.
B the farther x_{n+1} is from $\bar{x}$, the mean of the sample from which the coefficients were estimated..
C the smaller the variation of the x-values in our sample around the sample mean $\bar{x}$.
D All of the above.

7 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.

8 The variance of the least-squares slope estimator $\hat{\beta}_j$ is larger, and thus the true value of β_j is estimated less precisely,
A. the larger the sample size.
B. the smaller the variance of the error term σ^2.
C. the larger the variation of the X_{ij} values around the sample mean $\bar{x}_j$.
D. the more closely correlated X_{ij} is with the other regressors.

9 According to which model does a one-unit change in X cause approximately a five percent increase in Y?
A y = 6 + 0.05 X .
B y = 6 + 0.05 ln(X) .
C ln(Y) = 6 + 0.05 x .
D ln(Y) = 6 + 0.05 ln(X) .

10 Suppose a production function is estimated of the form $Y_i = \beta_1 + \beta_2 X_i$, where Y_i denotes output in kilograms and X_i denotes labor input. Now suppose the output data are converted to pounds (there are about 2.2 pounds to a kilogram) and the equation is re-estimated. Which of the following are true?

A $\hat{\beta}_1$ will increase by a factor of 2.2 .

B $\hat{\beta}_2$ will increase by a factor of 2.2 .

C The sum of squared residuals will increase by a factor of $(2.2)^2$.

D All of the above.

11 If two regressors X_{i2} and X_{i3} are closely but not perfectly correlated, then the least-squares estimators of their coefficients

A will be biased.

B will be inconsistent.

C will have large standard errors.

D will be zero.

12 Suppose Q = quantity demanded, P = price of the good, and I = consumer income. In which specification does β_2 equal the price elasticity of demand?

A $Q_i = \beta_1 + \beta_2 P_i + \beta_3 I_i$.

B $Q_i = \beta_1 + \beta_2 (P_i/I)$.

C $\ln(Q_i) = \beta_1 + \beta_2 P_i + \beta_3 I_i$.

D $\ln(Q_i) = \beta_1 + \beta_2 \ln(P_i) + \beta_3 \ln(I_i)$.

13 Under the null hypothesis of no heteroskedasticity, the Goldfeld-Quandt test statistic is close to

A zero.

B one.

C two.

D four.

14 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

15 Suppose we want to estimate the effect of the number of police officers per population on the crime rate. Poverty also has an effect on the crime rate, but we omit poverty from the equation for lack of data. Suppose police officers have a negative effect and poverty has a positive effect on crime, and police officers and poverty are positively correlated with each other in our data. Then omitting poverty from the equation will cause the least-squares estimator of the coefficient of the number of police officers

A to be biased up (toward zero).

B to be biased down (away from zero).

C to be unbiased.

D Cannot be determined from information given.

16 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.

17 Let u_i be the stochastic error term for observation i and assume: VAR ($u_i = \sigma^2 Z_i^2$. Further, assume that Zi= 19.56 for all i. Then:
A the stochastic error terms are homoskedastic.
B the stochastic error terms are heteroscedastic.
C we don't have enough information to determine whether the stochastic error terms are heteroscedastic or homoskedastic.
D the stochastic error terms definite display positive first-order serial correlation.

18 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., $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$?
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

19 Serial correlation is a violation of the Classical Assumption that:
A observations of the stochastic error term are homoskedastic.
B the regression model is linear in the coefficients.
C observations of the stochastic error term are heteroskedastic.
D observations of the stochastic error term are uncorrelated with one another.

20 Suppose that the variance of a regression model's stochastic error term is given by:

$$\mathrm{VAR}(\varepsilon_i) = \sigma^2 \sqrt{Z_i}, \text{ where}$$

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.

A2) Answer each question as clearly and concisely as possible on the answer sheet. Attempt all questions. Each question carries equal marks (1.0)

1) Soumya has a high school GPA=4.0. Given the additional information that the standard error of prediction for her observation is .04 and that the 95% critical value from the relevant normal distribution is 1.65, calculate a 95% confidence interval on her predicted college GPA

2) 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?

3) In a recent paper Sreya estimated the following equation:

$$\ln(PRICE) = \underset{}{0.152} + \underset{(0.0914)}{1.272} \ln(TAX) \qquad R^2 = 0.890,\ N = 2385$$

where, *PRICE* = average price/pack of cigarettes in a given state in a given year, and
TAX = average excise tax/pack of cigarettes in a given state in a given year

Standard errors are in parentheses.

If the tax increases by 10%, what is the expected change in price?

4) You estimated the demand for Oranges in the country over the period 1962 to 1991 and obtain the following results (standard errors in parentheses):

$$(1)\quad \ln Y_t = \underset{(1.38)}{6.18} - \underset{(0.29)}{0.52} \ln PO_t + \underset{(0.25)}{0.61} \ln X_t - \underset{(0.18)}{0.23} \ln PK_t \qquad R^2 = 0.74 \quad RSS = 23.19$$

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

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

$$(2)\quad \ln Y_t = \underset{(1.23)}{6.62} - \underset{(0.30)}{0.71} \ln(PO_t/PK_t) + \underset{(0.19)}{0.49} \ln(X_t/PK_t) \qquad R^2 = 0.68 \quad RSS = 28.54$$

A. What is it about equation (1) that leads you to suspect multicollinearity? Write clearly.

B. Test the hypothesis that the restriction is statistically valid at the significance level of 5%? Show all calculations. The critical value at 5% is 4.23.

5) In a simple regression model (with one explanatory variable X), show in a diagram that the relationship between actual and predicted values in the case that (Total Sum of Squares) TSS = ESS (Explained Sum of Squares). Consider dependent variable as Y.

6) Consider the following regression equation for the imports of the INDIA over the period 1970-1983: $\ln(\text{Imports})_t = 0.6754 + 0.3711 \ln(\text{GNP})_t + 1.5855 \ln(\text{CPI})_t$

$$\begin{array}{lcc} & (0.939) & (0.0435) \\ t = & (3.951) & (36.458) \end{array}$$

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

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?

7) With reference to the following regression using 45 observations:

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

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

8) Given the following estimated model (standard errors in parentheses)

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

And the regression for the test is

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

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

Use the relevant test statistic and test the hypothesis that errors are homoskedastic at 5% level. Table value of test statistic is 11.70.

9) Given a sample of data containing Y, X_1, X_2, and X_3 can you use Ordinary Least Squares to estimate the parameters (β_1, β_2, and β_3) 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) $Y = [a \,.\, \exp(\beta_1 X_1) * \exp(\beta_2 X_2^3)] / [\exp(\beta_3 X_3^{½})]$

10) Use the following list of assumptions for the linear regression model.

1 The regression model is linear in unknown coefficients α and β_i.

$$Y_t = \alpha + \beta_1 X_{1t} + ... + \beta_k X_{kt} + \mu_t \text{ for } t=1,2,.....,n$$

2 Not all observations on **X** are the same, i.e. $\mathbf{Var(X) > 0}$

3 The error term μ_t is a random variable with $\mathbf{E(\mu_t \mid X_t) = 0}$

4 $\mathbf{X_t}$ is given and nonrandom, implying that it is uncorrelated with ms that is $\mathbf{Cov(X_t, \mu_s) = 0}$

5 Given $\mathbf{X_t}$, μ_t has constant variance. $\mathbf{Var(\mu_t \mid X) = \alpha^2}$

6 Given $\mathbf{X_t}$, μ_t and μ_s are independently distributed for all **t ,es** so $\mathbf{Cov(\mu_t, \mu_s \mid X) = 0}$

7 The number of observation (n) must be greater than the number of regression coefficients estimate

8 For a given **X**, μ_t is normally distributed. $\mathbf{\mu_t \mid X \sim N(0, \alpha^2)}$

a) A failure of assumption 5 is called:

____

b) A failure assumption 6 is called:

____

c) If error terms are heteroscedastic, OLS (ordinary least squares) estimates of the parameter of the regression equation will be (check all that apply): (Biased /unbiased/efficient/not efficient/consistent) :

____

d)Assuming that all 8 assumptions hold, the nR^2 from the OLS regression) is distributed as:

____

End of PART –A

Attempt PART-B (Open Book)

Space for Rough Work

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER 2017-2018

Course No.: ECON 241 **Course Name: ECONOMETRIC METHODS**

Date : 03/MAY/2018 **COMPREHENSIVE EXAMINATION (PRT-B – OPEN BOOK)** **Weightage: 20%**

NOTE: This exam is open book examination. This exam contains 5 questions. You should try to do all 5 questions. Please provide explanations for all your answers. Right answers with no explanation will receive no credit. As it is an open book examination *No Partial credit* will be given for the numerical/ mathematical problems. Highlight your final answer and correct all answers to four decimal places.

B1) **(3.0)**

a) Consider the following model:

$Y_i = \beta_0 + u_i$. What is the OLS estimator for β_0.

b) A researcher is interested in how well ones height predicts ones weight. Hence, he collected data on 110 people in order to run the simple regression.

$Y_i = \beta_0 + \beta_1 X_i + u_i$, i = 1, ..., 110;

$\sum_{i=1}^{110} Yi = 17375$; $\sum_{i=1}^{110}(Yi - \bar{Y}) = 94228.8$; $\sum_{i=1}^{110} Xi = 7665.5$; $\sum_{i=1}^{110}(Xi - \bar{X}) = 1248.9$,

$\sum_{i=1}^{110}(Xi - \bar{X})(Yi - \bar{Y}) = 7625.9$ where $\bar{Y}$ and $\bar{X}$ denote the respective sample means.

I. From the information given above, calculate the OLS estimates for β_0 and β_1. Show your Work

II. Calculate the (unadjusted) R^2 measure for this regression and explain its meaning.

B2)Your Instructor estimated the following equation. Standard errors are in parentheses.

$$WKLYEARNINGS = -449.53 + \underset{(1.668)}{72.14}\ EDUCATION + \underset{(0.429)}{8.23}\ EXPERIENCE$$

a) Derive a 95% confidence interval for the returns to experience.

b) Suppose again he estimated:

$$WKLYEARNINGS = 170.95 \underset{(6.701)}{-65.145}\ EDUCATION + \underset{(0.279)}{5.73}\ EDUCATION^2 + \underset{(1.156)}{27.698}\ EXPERIENCE - \underset{(0.024)}{0.460}\ EXPERIENCE^2$$

Carefully interpret the coefficients on *EXPERIENCE* and *EXPERIENCE*^{2}.

c) Using the second equation, what are the weekly expected earnings for someone with a college degree (16 years of education) and no experience?

d) Test the null hypothesis that the returns to education are constant.

(4.0)

B3) Suppose you are given the following results for a time series of 23 annual Indian aggregate economic data (standard errors in parentheses):

$$\hat{C}t = 8.133 + 0.95W_t + 0.452\ P_t + 0.121A_t$$

$$(8.91) \quad (0.95) \quad (.066) \quad (1.09) \quad R^2 = 0.95 .... (1)$$

Where C_t = annual Indian domestic consumption (in billions of Indian Rupees);
W_t= annual Indian wage income (in billions of Indian Rupees);
P_t= annual Indian nowage-nofarm income (in billions of Indian Rupees);
A_t = annual Indian farm income (in billions of Indian Rupees).

a How would you interpret the coefficients of Wt, $P_{t,}$ $A_{t.}$ Test the null hypotheses that the coefficients of W_t, $P_{t,}$ A_t are individually indifferent from zero respectively at a 5% significance level using a one- sided test. Table value of test statistic is 1.729

b Interpret the coefficient of determination. Specify the hypothesis to test the overall significance of the regression model. Use the same coefficient of determination value and construct the appropriate test statistic and test the hypothesis.

c What can conclude from (a) and (b)? In this case what are the properties of the OLS estimator? Do you need to be concerned with the problem from the perspective of hypothesis testing? Why or Why not?

d Briefly explain how to use an auxiliary regression to detect the problem you identified in (C). Suggest a way to solve the problem.

e The following results are obtained based on the same data by another researcher when he estimated the model:

$$\hat{C}t = 8.94 + 0.61\ P_t$$

$$(1.67) \quad (0.20) \quad R^2 = 0.80 .... (2)$$

Which model, (1) or (2), do you prefer? Give your reasoning.

(5.00)

B4) A researcher estimated the simple linear regression model, $Y_t = \beta_1 + \beta_2 t + u_t$ on the **log India GDP** quarterly series, 2007-2016 (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.

E). Another analyst included the BSE index and the Inflation as independent variables in the original model. RSS and ESS for this new model was found to be **0.0344** and **0.6688** respectively. Use a suitable test to decide whether the inclusion of the BSE index and the Inflation significantly improves the model. Would you expect strong multicollinearity in this model? If so - what can be done to resolve this problem?

(5.00)

B5)

a) A researcher investigating whether government expenditure tends to crowd out investment fits the regression (standard errors in parentheses) is given below. The notations are government recurrent expenditure, G, investment, I, gross domestic product, Y, and population, P, for 30 countries in 1997 (source: 1999 International Monetary Fund Yearbook). G, I, and Y are measured in US$ billion and P in million

$$\hat{I} = \underset{(7.79)}{18.10} - \underset{(0.14)}{1.07}G + \underset{(0.02)}{0.36}Y \qquad R^2 = 0.99.$$

She arranges the observations by size of *Y* and runs the regression again for the 11 countries with smallest *Y* and the 11 countries with largest *Y*. *RSS* for these regressions is 321 and 28101, respectively. Perform a Goldfeld–Quandt test for heteroscedasticity.

b) The researcher again runs the following regressions as alternative specifications of the model (standard errors in parentheses):

$$\hat{I}/P = \underset{(0.28)}{-0.03}(1/P) \underset{(0.16)}{-0.69}(G/P) \underset{(0.03)}{+0.34}(Y/P) \qquad R^2 = 0.97 \quad ....(1)$$

$$\hat{I}/Y = \underset{(0.04)}{0.39} + \underset{(0.42)}{0.03}(1/Y) - \underset{(0.22)}{0.93}(G/Y) \qquad R^2 = 0.78 \quad ....(2)$$

$$\widehat{\log I} = \underset{(0.26)}{-2.44} - \underset{(0.12)}{0.63}\log G - \underset{(0.12)}{1.60}\log Y \qquad R^2 = 0.98 \quad ....(3)$$

In each case the regression is run again for the subsamples of observations with the 11 smallest and 11 greatest values of the sorting variable, after sorting by Y/P, G/Y, and log Y, respectively. The residual sums of squares are as shown in the following table.

	11 Smallest	**11 largest**
1	1.43	12.63
2	0.0223	0.0155
3	0.573	0.155

Perform a Goldfeld – Quandt test for each model specification and discuss the merits of each specification. Is there evidence that investment is an inverse function of government expenditure?

(3.00)