BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2015-16

Course No.: ECON F241 **Course Title: ECONOMETRIC METHODS**
Date: 13 MAY 2016 **Max. Marks: 40**

COMPREHENSIVE EXAMINATION

NOTE: Attempt all questions. Answer to the point. This paper consists of two parts: PART- A (Closed Book: 15 Marks) and PART-B (Open Book: 25 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) **15 Marks**

- **Answer each question as clearly and concisely as possible on the answer sheet. Attempt all questions. Question A2 Carries 2.0 Mark and all other questions carry equal marks.**

A1) What is asymptotic normality? What is its implication for regression analysis? (in other words, why is asymptotic normality important?)

ID.NO:	NAME:

A2) **For multiple choice questions Choose the correct best alternative and put a tick (√) against that letter. Corrections/Overwriting/illegible answers are invalid. Each question carries equal marks.**

1) In econometrics, we typically do not rely on exact or finite sample distributions because

A. we have approximately an infinite number of observations (think of re-sampling).
B. variables typically are normally distributed.
C. the covariances of Y_i, Y_j are typically not zero.
D. asymptotic distributions can be counted on to provide good approximations to the exact sampling distribution.

2) The central limit theorem states that

A. the distribution for $\frac{\bar{Y} - \mu_Y}{\sigma_{\bar{Y}}}$ becomes arbitrarily well approximated by the standard normal distribution.
B. $\bar{Y} \xrightarrow{p} \mu_Y$.
C. the probability that $\bar{Y}$ is in the range $\mu_Y \pm c$ becomes arbitrarily close to one as n increases for any constant $c > 0$.
D. the t distribution converges to the F distribution for approximately $n > 30$.

3) An estimator $\hat{\mu}_Y$ of the population value μ_Y is more efficient when compared to another estimator $\tilde{\mu}_Y$, if

A. $E(\hat{\mu}_Y) > E(\tilde{\mu}_Y)$.
B. it has a smaller variance.
C. its c.d.f. is flatter than that of the other estimator.
D. both estimators are unbiased, and var($\hat{\mu}_Y$) < var($\tilde{\mu}_Y$).

4) Among all unbiased estimators that are weighted averages of $Y_1,...,Y_n$, $\bar{Y}$ is

A. the only consistent estimator of μ_Y.
B. the most efficient estimator of μ_Y.
C. a number which, by definition, cannot have a variance.
D. the most unbiased estimator of μ_Y.

5) An estimator $\hat{\mu}_Y$ of the population value μ_Y is more efficient when compared to another estimator $\tilde{\mu}_Y$, if

A. $E(\hat{\mu}_Y) > E(\tilde{\mu}_Y)$.
B. it has a smaller variance.
C. its c.d.f. is flatter than that of the other estimator.
D. both estimators are unbiased, and var($\hat{\mu}_Y$) < var($\tilde{\mu}_Y$).

6) Two variables are uncorrelated in all of the cases below, with the exception of

A. being independent.
B. having a zero covariance.
C. $|\sigma_{XY}| \leq \sqrt{\sigma_X^2 \sigma_Y^2}$.
D. $E(Y \mid X) = 0$.

7) Interpreting the intercept in a sample regression function is

A. not reasonable because you never observe values of the explanatory variables around the origin.
B. reasonable because under certain conditions the estimator is BLUE.
C. reasonable if your sample contains values of X_i around the origin.
D. not reasonable because economists are interested in the effect of a change in X on the change in Y.

8) The slope estimator, β_1, has a smaller standard error, other things equal, if

A. there is more variation in the explanatory variable, X.
B. there is a large variance of the error term, u.
C. the sample size is smaller.
D. the intercept, β_0, is small.

ID.NO:	**NAME:**

A3) We wish to estimate the average level of some protein in the blood of an entire population. We assume the protein level follows a normal distribution across the entire population with unknown mean μ and unknown variance σ^2. Measuring the protein level in any individual is costly, so we must base our estimates on a small sample. Accordingly, we have measured the protein level in a random sample of **n=12** individuals. Denote each individual's protein level measurement by x_i, for i = 1, …, n, and denote the sample mean by $\bar{x}$. The following statistics have already been calculated: $\sum_{i=1}^{n} x_i = 396$ and $\sum_{i=1}^{n} (x_i - \bar{x})^2 = 2112$.

Test the null hypothesis H_0: μ=20 against the **two-sided** alternative hypothesis H_1: $\mu \neq 20$ at **5%** significance. Given the table value of the test statistic 2.201 whether the null hypothesis should be accepted or rejected at 5% significance.

ID.NO:	NAME:

A4) Discuss the effect of the following phenomena on the unbiasedness and the variance of the OLS estimators, $\hat{\beta}_j$

Multicollinearity:

Including an irrelevant variable:

A5) You run a regression using four explanatory variables. Your adjusted R^2 (R bar square) is very high at 0.99 , but the t statistic suggests that none of the explanatory variables are significant at 5% level of significance using a 2-tail test. What type of problem might you suspect ? How do you solve it?

ID.NO:	NAME:

A6) 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

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

$$Y_i = -2.46 + 6.11\,X_{2i} - 1.78\,X_{3i}$$
$$(0.94) \quad (2.80) \quad (1.47)$$

N=40, R^2 = 0.43, RSS = 287.2

And the regression for the test is

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

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. The table value of test statistic is 11.70.

ID.NO:	NAME:

A8) 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) $Y = a + (\beta_1 X_1)^2\ (\beta_2 X_2)^{0.25}$

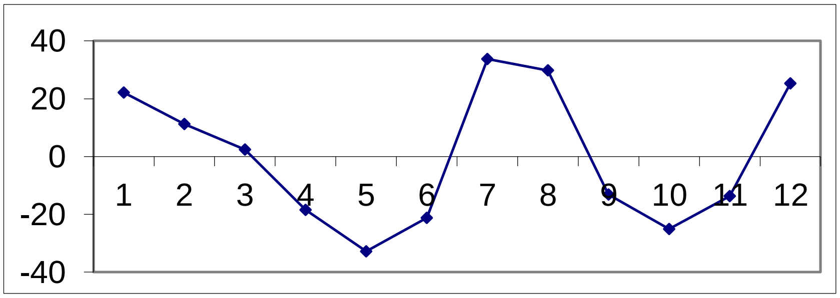
iv) $Y = [a\ .\ \exp(\beta_1 X_1) * \exp(\beta_2 X_2^{3})] / [\exp(\beta_3 X_3^{½})]$

ID.NO:	NAME:

A9) Below you will find a regression model that compares the relationship between the average utility bill (Y, in Rupees) for homes of a particular size and the average monthly temperature (X, in Fahrenheit). The data represents monthly values for the past year. Also, the value for the Durbin-Watson statistic = 1.244, and a residual plot is shown below.

Summary measures	
Multiple R	0.0295
R-Square	0.0009
StErr of Estimate	24.8184

ANOVA table					
Source	df	SS	MSS	F	p-value
Explained	1	5.3575	5.3575	0.0087	0.9275
Unexplained	10	6159.5125	615.9512		
Regression coefficients	Coefficient	Std Err	t-value	p-value	
Constant	112.547	28.815	3.9059	0.0029	
Average Monthly Temp	0.0403	0.4316	0.0933	0.9275	



Is there a linear relationship between X and Y? Explain how you arrived at your answer

In looking at the graph of the residuals, do you see any evidence of any violations of the assumptions regarding the errors of the regression model?

Giving the Durbin-Watson value presented above, what would you conclude about the data?

A10) How the Durbin-Watson ‘d statistic’ is related to the Von Neumann ratio? Briefly explain.

A11) Consider the case that the regression function does not have an intercept. If we know that the population regression function is $Yi = \beta_1 X_i + u_i$. The intercept $\beta_0 = 0$.

What is $E(Y_i/X_i)$? What is $E(Y_i/X_i = 0)$?

Derive the OLS estimator for β_1 using the moment condition $E(u_i X_i) = 0$:

ID.NO:	NAME:

A12) Consider a regression with time series data measured at quarterly intervals and let three seasonal dummy variables be defined as follows:

$$X_{1t} = \begin{cases} 1, \text{if observation t is in the first quarter} \\ 0, \text{otherwise} \end{cases}$$

$$X_{2t} = \begin{cases} 1, \text{if observation t is in the second quarter} \\ 0, \text{otherwise} \end{cases}$$

$$X_{3t} = \begin{cases} 1, \text{if observation t is in the third quarter} \\ 0, \text{otherwise} \end{cases}$$

Next consider the following estimated regression model:

$$\hat{S}_t = 15{,}600 - 4{,}500 \cdot X_{1t} + 300 \cdot X_{2t} + 62{,}500 \cdot X_{3t},\ t = 1, \ldots, n, \text{ where}$$

S_t is the rupee amount of sales at a souvenir shop at a city resort in period t.

According to this model,

a) Compute the expected sales in the first quarter?

b) Compute the expected sales in the third quarter?

ID.NO:	NAME:

A13) Suppose you have the model : price = $\beta_0 + \beta_1$sqrft + β_2 bdrms + u.

where price is the house price measured in thousands of rupees, sqrft is the size of the house in square feet and bdrms is the number of bedrooms in the house.

The OLS regression line for this model is

Estimated price = -19:32 + 0.128sqrft + 15.20 bdrms n = 88, R^2 = :632

Interpret the estimated coefficient on sqrft.

Now if you run the regression $\widetilde{price} = \widetilde{\beta_0} + \widetilde{\beta_1} sqrft$

What would be the likely sign and magnitude of $\widetilde{\beta_1}$ relative to $\widehat{\beta_1} = .128?$ briefly explain your answer.

ID.NO:	NAME:

A14) Suppose that from 24 yearly observations on the quantity demand of a commodity in kilograms per year Y, its price in rupee X1, consumer's income in thousands of rupees X2, and the price of a substitute commodity in rupees X3, the following estimated regression is obtained, where the numbers in parentheses represent standard errors:

$$\hat{Y} \text{ cap} = 13 - 7X_1 + 2.4\,X_2 - 4\,X_3$$
(2) (0.8 (18)

F = 126.68, P-value = 0.0000, $R^2 = 0.9425$
t(0.025;20) = 2.086, t(0.05;20) = 1.725, F(0.05;3,20) = 3.10

Indicate whether the signs of the parameters conform to those predicted by demand theory. Are the estimated slope parameters significant at the 5% level? Explain in details.

End of PART –A

Attempt PART-B (Open Book

ID.NO:	NAME:

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER 2015-16

ECON C342/FIN C332: ECONOMETRICS | **ECON F241: ECONOMETRIC METHODS**
DATE: 13 MAY 2016 | **MAX.MARKS: 25**

COMPREHENSIVE EXAMINATION – PART: B (OPEN BOOK)

Note: Attempt all questions. Start answering each question on a fresh page. Attempt all parts of question at one place. You must show all your work to receive full credit. Any assumptions you make and intermediate steps should be clearly indicated. Read the questions carefully, answering what is asked.

B1) Suppose you are given the following results for a time series of 23 annual Indian aggregate economic data (standard errors in parentheses):

$$\hat{C}t = 8.133 + 0.95W_t + 0.452\ P_t + 0.121A_t$$
$$(8.91)\quad (0.95)\quad (.066)\qquad (1.09)\qquad R^2 = 0.95.........\ (1)$$

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). If we measure the dependent variable and all the independent variables by millions of Indian Rupees instead of by billions of Indian Rupees, without re-estimating the model, what are the new estimated coefficients?

The following results are obtained based on the same data:

$$\hat{C}t = 8.94 + 0.61\ P_t$$
$$(1.67)\quad (0.20)\qquad R^2 = 0.80.........\ (2)$$

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

(6.00)

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**

ID.NO:	**NAME:**

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: $\hat{u}_t^2 = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + v_t$, 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?

(8.00)

B3) 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.

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

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

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

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

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.

ID.NO:	NAME:

	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)

B4) In a study of costs in the banking industry, data has been collected for 85 banks in India. Some banks specialize in lending money to households; other banks specialize in lending money to businesses. Similar issues hold with respect to their depositors. We are interested in investigating whether these bank characteristics affect their costs. Accordingly, we define the dependent and explanatory variables as follows:

Y = total costs per employee (in thousands of rupees per year).

X_1= proportion of total loans which go to businesses

(measured as a percentage so that a value of, say, 20 means 20% of loans are made to businesses).

X_2= proportion of total deposits which come from households

(measured as a percentage so that a value of, say, 20 means 20% of deposits come from households).

D = a dummy variable which equals 1 if the bank is a big bank (has more than 100 employees),

= 0 otherwise.

We also constructed another variable, $Z = X_2 \times D$.

A). We ran a regression of Y on X_1, X_2, D and Z. Results from this regression are given below in the following fitted regression line:

Y =	**960**	**- 109 x X_1**	**+ 120 x X_2**	**- 1.49 x D**	**- 23 x Z**
	(6×10^{-7})	(0.008)	(0.04)	(0.03)	(0.002)

where the numbers in parentheses are P-values for testing the hypothesis that the coefficient equals zero.

i) How would you interpret (in words) the estimated coefficients in this model? What is the OLS estimate of the marginal effect of X2 on Y?

ii) Which of the statements you have just made are statistically significant at the 5% level? Which are significant at the 1% level?

ID.NO:	NAME:

B). We then did a White test (using X_2 as the independent variable to explain the heteroskedasticity) and found a test statistic value of 5.02 (with a p-value of 0.025). We re-ran the previous regression using a heteroskedasticity consistent estimator (HCE) and obtained:

	960	**- 109 x X_1**	**+120 x X_2**	**-10.449 x D**	**-203 x Z**
Y =	(4 x 10^{-5})	(0.023)	(0.067)	(0.04)	(0.005)

where the numbers in parentheses are P-values for testing the hypothesis that the coefficient equals zero.

When presenting final results in a project, would you use our OLS results of part A) or HCE results of part B)? Why?

(4.00)

B5) A) Write a two-equation system in "supply and demand form," that is, with the same variable $y1$ (typically, "quantity") appearing on the left-hand side:

$$\boldsymbol{y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1}$$

$$\boldsymbol{y1 = \alpha_2 y_2 + \beta_2 z_2 + u_2.}$$

(i) If $\alpha_1 = 0$ or $\alpha_2 = 0$, explain why a reduced form exists for y_1. (Remember, a reduced form expresses y_1 as a linear function of the exogenous variables and the structural errors.) If $\alpha_1 \neq 0$ and $\alpha_2 = 0$, find the reduced form for y_2.

(ii) If $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, and $\alpha_1 \neq \alpha_2$, find the reduced form for y_1. Does y_2 have a reduced form in this case?

(iii) Is the condition $\alpha_1 \neq \alpha_2$ likely to be met in supply and demand examples? Explain.

B) Given the following Simultaneous Equation Model, use the order and rank conditions of identification to determine whether the model is identified or not. (The y's represent the endogenous variables and the z's the exogenous variables).

$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}z_1 + u_1$ -------------------------------------- (1)

$y_2 = \alpha_{21}y_1 + \beta_{21}z_1 + \beta_{22}z_2 + \beta_{23}z_3 + u_2$ ------------------------------ (2)

$y_3 = \alpha_{32}y_2 + \beta_{31}z_1 + \beta_{32}z_2 + \beta_{33}z_3 + \beta_{34}z_4 + u_3$---------------------(3)

(4.00)

ID.NO:	**NAME:**