BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2013-14

Course No.: ECON F241/ECON C342/FIN C332/MGTS C443 **Course Title: ECONOMETRIC METHODS**
Date: 13 MAY 2014 **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**

For the following multiple choice questions, choose the correct best alternative and put a tick (√) against that letter. Corrections/Overwriting/illegible answers are invalid. Each multiple choice question carries equal mark of 0.50. (Total: 8 Marks)

1. Suppose you know for sure that a variable does not belong in a regression as an explanatory variable. If someone includes this variable in their regression, in general this will create

A. bias and increase variance
B. bias and decrease variance
C. no bias and increase variance
D. no bias and decrease variance
E. None of the above

2. Suppose we estimate the equation: $y_i = \beta_1 + \beta_2 x_{i2} + \beta 3\ x_{i3}$ .If x_{i2} and x_{i3} are closely but not perfectly correlated, then the least-squares estimators of their coefficients

A. will have large standard errors.
B. will be zero.
C. cannot be computed.
D. will be biased.
E. None of the above

3. A researcher runs $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + u_i$ and then $\sigma^2_{Ui} = \gamma_0 + \gamma_1 X_{1,i} + \gamma_2 X_{2,i} + \gamma_3 X_{1,i} X_{2,i} + \gamma_4 X_{1,i} + \gamma_5 X^2_{2,i} + \varepsilon_i$ where $\sigma^2_{Ui} = \mathrm{Var}(u_i)$. If $p\ (F) = 0.74$ in the second regression, what would be your conclusion?

A. there is no multicollinearity problem in the first regression
B. there is multicollinearity problem in the first regression
C. there is no heteroscedasticity problem in the first regression
D. there is negative serial correlation in residuals
E. None of the above

4. A researcher runs $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 D_i + \beta_3\ D_i\ X_{1,i} + u_i$ where is the salary of worker i, $X_{1,i}$ is the years of experience, and Di is a dummy that takes the value of 1 if the worker is female and 0 otherwise. Which of the following would indicate that female workers' salary increases at a slower rate with years of experience.

A. $\beta_1 < 0$
B. $\beta_2 < 0$
C. $\beta_2 > 0$
D. $\beta_3 < 0$
E. None of these

5. Of the definitions below, which best describes the property "Unbiasedness of the OLS estimators"?

A. The average of the OLS estimates across a big number of samples equals the true parameter value
B. The distribution of the OLS estimators has small mass for values far away from the true parameter value
C. The average of the OLS estimates across an infinite number of samples equals the true parameter value
D. The distribution of the OLS estimators has a mean very close to the true parameter value
E. None of the above descriptions describes properly "Unbiasedness of the OLS estimators"

6. **Suppose you estimate the following model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$. We want to test the following null hypothesis, $H_0: \beta_3=0$ against $H_1: \beta_3 \neq 0$. The p value for β_3 is 0.04.**

I. You can reject the null hypothesis at 5 % significance level
II. You can reject the null hypothesis at 4 % significance level
III. You can reject the null hypothesis at 3 % significance level

Which combination of statements is true?

A. (I) only
B. (I) and (II) only
C. (II) and (III) only
D. (III) only
E. (I),(II) and (III)

7. **Suppose you estimate the following model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$. Adding more and more number of regressors will lead to which of the below likely consequences?**

I. Bias II. Multicollinearity III. Heteroskedasticity

A. I only.
B. II only.
C. III only.
D. I and II only.
E. I, II, and III.

8. **Dropping a variable can be a solution to a multicollinearity problem because it**

A. avoids bias
B. increases t values
C. eliminates the collinearity
D. could decrease mean square error
E. All the above

9. **Suppose your specification is $y = \beta x + \gamma Male + \theta Female + \delta Weekday + \lambda Weekend + \varepsilon$**

A. there is no problem with this specification because the intercept has been omitted
B. there is high collinearity but not perfect collinearity
C. there is perfect collinearity
D. there is inconsistency
E. None of the above

10. **If the error term in a regression is heteroskedastic, the least-squares estimators for the coefficients will be**

A. Biased and inconsistent.
B. best linear unbiased estimators.
C. best unbiased estimators.
D. Both A and B
E. None of these.

11. **With heteroskedasticity we should use weighted least squares where**

A. by doing so we maximize R-square
B. use bigger weights on those observations with error terms that have bigger variances
C. we use bigger weights on those observations with error terms that have smaller variances
D. the weights are bigger whenever the coefficient estimates are more reliable
E. All of the above

12. **Suppose a production function is estimated of the form $y_i = \beta_1 + \beta_2 x_i$, where y_i denotes output in gallons and x_i denotes labor input. Now suppose the output data are converted to liters (there are about 3.8 liters in a gallon) and the equation is re-estimated. Which of the following are true?**

A. $\hat{\beta}_2$ will increase by a factor of 3.8 .
B. The sum of squared residuals will increase by a factor of $(3.8)^2$.
C. The r^2 value will be unaffected.
D. All of the above.
E. Both A and B

13. **Suppose you have run the following regression: $y = \alpha + \beta x + \gamma Urban + \theta Immigrant + \delta Urban*Immigrant + \varepsilon$ where Urban is a dummy indicating that an individual lives in a city rather than in a rural area, and Immigrant is a dummy indicating that an individual is an immigrant rather than a native. The coefficient γ is interpreted as the ceteris paribus difference in y between**

A. an urban native and a rural native
B. an urban immigrant and a rural immigrant
C. an urban native and a urban immigrant
D. rural native and a rural immigrant
E. None of the above

14. **Assume that the model is $y_t = \beta_0 + \beta_1 x_t + u_t$. Consider the alternative estimator β_1 bar$_=(y_2-y_1)/(x_2-x_1)$.**

A. β_1 bar is an unbiased estimator of β_1
B. β_1 bar is a consistent estimator of β_1
C. β_1 bar is an efficient estimator of β_1
D. All of the above
E. Both B and C

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 (towards zero).
B. to be biased down (away from zero).
C. to be unbiased.
D. To be consistent
E. Cannot be determined from information given.

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

A. minus one.
B. zero.
C. one.
D. two.
E. four.

For the following short answer questions write the correct best answer in the space provided.

17. **An econometrician estimates the following model:**

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 female + u$$

Suppose the econometrician wants to know if the effect of experience varied by education level. How can she modify the model .To answer, just write down the new population model.

(1 M)

Ans:

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

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

Note: (standard errors in brackets). **Given: d_L = 1.48 and d_U is 1.57**

A. Calculate the value of correlation coefficient r. **(1 M)**

Ans:

B. Is there any evidence of 1st order autocorrelation.

(1 M)

Ans:

19. Consider the aggregate production model:

$$\ln X_t = \beta_0 + \beta_1 \ln L_t + \beta_2 \ln K_t + u_t$$

where X_t is an index of India's GNP in constant rupees for year 2004-05.L_t is a labor input index, K_t is a capital input index, and u_t is a disturbance term.

A regression of this equation from annual data for 1929-1948 yields:

$$\ln X_t = -4.0576 + 1.6167 \ln Lt + 0.2197 \ln Kt \qquad \text{---(1)}$$
$$(0.209) \qquad (2.289)$$

with R^2 = .9759, s(standard deviation) = .04579.

A regression of the equation for the sample 1949-1967 yields:

$$\ln Xt = -1.9564 + 0.8336 \ln Lt + 0.6631 \ln Kt \qquad \text{---(2)}$$
$$(0.2488) \qquad (0.2289)$$

with R^2 = .9904, s = .02185

A. Calculate the degrees of freedom for equation 1 and 2.

(1 M)

Ans:

B. Test the hypothesis that the variances σ^2 are the same in both samples. Critical value of test statistic is 3.52

(1 M)

Ans:

20. With reference to the following simultaneous equation model,

$$\mathbf{Y_1 - 3Y_2 + 2X_1 - X_2 - U_1 = 0}$$
$$\mathbf{Y_2 - Y_3 - X_{3\,-}\, U_2 = 0}$$
$$\mathbf{Y_3 - Y_1 + Y_2 + 2X_3 - U_3 = 0}$$

A. Use the order condition to check for over-identification, under-identification or exact identification for each of the above three equations. **(1 M)**

Ans:

B. Use the rank condition procedure to calculate the rank for each of the above three equations

(1 M)

Ans:

Space for Rough Work:

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2013-14

Course No.: ECON F241/ECON C342/FIN C332/MGTS C443 **Course Title: ECONOMETRIC METHODS**
Date: 13 MAY 2014

PART-B (OPEN BOOK) **(Max. Marks: 25)**

NOTE: Attempt all questions. Start each question from a new page. Answer to the point.

1. The following equation was estimated for BITS PILANI –Pilani campus, where standard errors are in parenthesis:

BITSAT (hat) = 1,028.1 + 19.30 hsize –45.09 male – 169.81 ACB + 62.31 male*ACB
(6.29) (3.83) (4.29) (12.71) (18.15)

Where hsize is the number of students in BITS Pilani campus, male and ACB are dummy variables indicating that the student is male or under effect of ACB, respectively, and male*ACB is an interaction of the two.

A. What is the estimated difference in BITSAT score between ACB females and non-ACB females?
B. What is the estimated difference in BITSAT score between ACB males and non-ACB males? Can we construct a t-statistic for this particular case. If yes, then calculate the value of t-statistic. If no, then give reasons that why we can't construct ?
C. State the null hypothesis that the effect of being male does not vary by ACB.
D. What is the marginal effect of being male on BITSAT (hat)? Briefly interpret this result. **(4 M)**

2. Consider the simple regression model

$$\log(\text{scrap rate}) = \beta_0 + \beta_1 \text{grant} + u$$

where scrap rate is given for Honda Motors Plant, Gurgaon and grant is a dummy variable indicating whether the Honda Motors plant received a job training grant or not.
(Note : Scrap rate is defined as the Percentage of failed items or material that cannot be repaired or restored, and is therefore condemned and discarded)

A. Suppose the ability of the Honda Motor employees, which is unobserved, is an important omitted variable in this model. Now, consider adding as an explanatory variable log ($scrap_{lastyear}$), the log of last period's scrap rate. Is this inclusion a sensible decision?
B. How would the inclusion of log ($scrap_{lastyear}$) variable affect your estimates of β_1?
C. Finally, suppose that the reporting of scrap rates is voluntary, so that the sample of plants is not a random sample of all plants in the population. With this type of sample selection, will the estimates of β_1 be unbiased? (You can assume that there are no reasons why β_1 would be biased other than sample selection.) **(3 M)**

3. Suppose for the following regression equation:

$$\text{lprice} = \beta_0 + \beta_1 \text{ lproptax} + \mu$$

where the logarithm of the price of an apartment, **lprice**, is regressed on the logarithm of the property tax of the district where the apartment is located, **lproptax.** The following is regression output for 33 districts of Rajasthan.

	Coefficient	Std. Error	t-ratio	p-value
Const	13.34193	.2273745	58.68	0.000
lproptax	-.5733679	.0382489	-14.99	0.000

The descriptive statistics of the variables are given below:

Variable	Obs.	Mean	Std. Dev.
lprice	33	9.941057	.409255
lproptax	33	5.931405	.3963666
price	33	22511.51	9208.856
proptax	33	40.82372	16.85371

A. Comment the effect of lproptax on lprice.
B. Derive the formula of calculating the marginal effect of proptax on price from results reported?
C. Using this formula, calculate the marginal effect referring to the average point. (Here Average point assumes that for a particular variable let us say price, we can approximate price= average of price). **(3 M)**

4. Consider the following model:

$$GNPt = \beta_1 + \beta_2 M_t + \beta_3 M_{t-1} + \beta_4 (M_t - M_{t-1}) + u_t$$

Where GNPt = GNP at time t, Mt = money supply at time t, M_{t-1} = money supply at time (t − 1), and ($M_t - M_{t-1}$) = change in the money supply between time t and time (t − 1). This model thus postulates that the level of GNP at time t is a function of the money supply at time t and time (t- 1) as well as the change in the money supply between these time periods.

Assuming you have the data to estimate the preceding model, would you succeed in estimating all the coefficients of this model? Why or why not?

A. If not, what coefficients can be estimated?
B. Suppose that the $\beta_3 M_{t-1}$ term were absent from the model. Would your answer to (a) be the same? **(3 M)**

5. You are given the following data

A. Calculate the Durbin–Watson statistic?

B. Is there positive serial correlation in the disturbances?

Y, expenditure (in rupees in crores)	X, time	Y cap, (estimated Y)
281.4	1(=1999)	261.4208
288.1	2	276.6026
290.0	3	291.7844
307.3	4	306.9661
316.1	5	322.1479
322.5	6	337.3297
338.4	7	52.5115
353.3	8	367.6933
373.7	9	382.8751
397.7	10	398.0569
418.1	11	413.2386
430.1	12	428.4206
452.7	13	443.6022
469.1	14	458.7840
476.9	15(=2013)	473.9658

*Obtained from the regression $Y_t = \beta_0 + \beta_1 X_t + u_t$. Significance level can be assumed as 5 %. **(3 M)**

6. For a particular economic problem, we regress the test score against the average income and its square term as follows:

Dependent Variable: TESTSCR

Included observations: 420

TESTSCR=607.3017+ 3.850995(AVGINC) - 0.042308(AVGINC^2) ; SSR=67510.32 --- (1)

(2.901754) (0.268094) (0.004780)

TESTSCR=625.3836+ 1.878550(AVGINC) ; SSR=74905.20 --- (2)

(1.532405) (0.090504)

R^2=0.507558, Adjusted R^2=0.506380

A. Compute the missing R-squared and adjusted R-squared in equation 1 assuming both regression equations have same total sum of squares.

Assume the model in equation 1 is the correct one, but you did not realize it and instead run the model in equation 2. You also happen to do some covariance analysis between income and its square term and get the following values:

VAR (INCOME)	52.08917
VAR (INCOME SQUARE)	123050.9
COVARIANCE	2428.427

B. After you see the result in equation 1, do you still believe that your model give you a correct (that is, Unbiased) estimate of slope coefficient? Explain it? **(1+2=3 M)**

7. Suppose we have data on the quantity Q of a product traded at various locations when the market is in equilibrium, the market price P_Q in each location, average consumer income I in each location, and the price of materials P_M in each location. For this, consider the following simultaneous equation macroeconomic model :

Demand: $QD = \beta_0 + \beta_1 \cdot P_Q + \beta_2 \cdot I + u_D$, where "$u_D$" is an error term in the Demand equation,

Supply: $QS = \beta_3 + \beta_4 \cdot P_Q + \beta_5 \cdot P_M + u_S$, where "$u_S$" is an error term in the Supply equation

A. Explain which variables are assumed to be exogenous and endogenous in the above simultaneous equation model. Which variables are called the pre-determined variables?

B. Derive the reduced form equations for the given demand – supply simultaneous equation model

C. Estimate the reduced form equations when β_0=1.234 , β_1=5.345, β_2 =0.288, $\beta_3 = \beta_1 \cdot 2.791$, β_4 = (2 x1.421 x 1.149543), $\beta_5 = 3.824+\beta_2$, u_D=0.666 and u_S =0.0259 (x sign implies multiplication)

(2+2+2=6 M)

****