BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER 2012-13

ECON C342/FIN C332: ECONOMETRICS **ECON F241: ECONOMETRIC METHODS**
MAX MARKS: 45 **DATE: 08/MAY/2013**
TIME: hr.3.00

COMPREHENSIVE EXAMINATION

Instructions: This paper consist of two parts: PART-A (Closed Book – 20 Marks) and PART-B (Open Book – 25 Marks). After completing PART-A submit and attempt PART-B (Open Book). Answer to the point.

ID. No: **NAME:**

PART – A (CLOSED BOOK) **MARKS: 20**

Note: Attempt all questions. Write Legibly. You must show all your work to receive full credit. Any assumptions you make should be clearly indicated. Do not simply write down a final answer to the problems without an explanation when required. No partial marking for answers. Questions 1-10 carries1.0 marks each and Q 11-15 - 2.0 marks each.

1. Each factory which makes Lays Potato Chips makes thousands of chips every day. Suppose you sample 100 chips to ensure the quality control machines are running correctly (like all things, some mistakes are expected even when they is running well). What technique (Binomial distribution/ Poisson/t-test/Z-test) should you use to determine if the mistakes are numerous enough to determine if the machine is broken? Why?

2. Suppose cargo captain Kasi Reddy wants to invest in a new starship. To determine what she can afford, she needs to figure out what she can expect to transport (in cubic meters). Suppose she estimates the cubic meters of goods she can expect to transport using the latest orders from her busiest run: from Earth to Bajor. Based on previous experience, she knows that the standard deviation in cubic meters is 3,000. (A war just broke out and it will decrease trade—so she can't use her previous data to determine the new mean—but it's reasonable to assume the standard deviation is the same.) With 95% confidence, she wants to limit her error to no more than 600 cubic meters on either side of her mean. How many orders will she have to sample? (HINT: at 95% confidence, z=1.96.)

3. Consider Rachel, who works for an auto insurance company. She wants to know if men are worse drivers than women. In a survey of 100 men, 12 have caused a major accident in the past five years. In a survey of 110 women, 6 have caused a major accident in the past five years. At 95% confidence (z=1.96) do men have a higher accident rate than women?

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

5. Suppose you run a regression with an earner's income predicting his/her children's level of education. If you add the earner's level of education as an explanatory variable, what econometric problem, if any, will you get?

6. Suppose you ran a regression of AGE predicting INCOME (thousands of rupees). If your estimated line was INCOME = -1.5 + 1.2*AGE + u, how much more income do you expect to make for every two years you age?

7. Francis runs a regression with a sample of 33 and with 17 explanatory variables (including the intercept, which is included in k). His R^2 is 0.70. What is his adjusted R^2?

8. A random sample of twelve automobiles showed the following figures for miles achieved on a gallon of gas. Assume the population distribution is normal. From the data:

$$n = 12, \sum_{i=1}^{12} X_i = 232.9, \sum_{i=1}^{12} X_i^2 = 4533.49, and \sum_{i=1}^{12} (X_i - \overline{X})^2 = 13.2867.$$

From t-table for α = .05 $t_{\frac{\alpha}{2},11} = 2.201$. Find an approximate 95% confidence interval for the population mean.

9. In 1995, the government bond yield in the United States was 10.62 percent. A random sample of government bond yields in nine foreign countries was: 11.04, 6.34, 10.94, 13.00, 7.34, 13.09, 4.78, 10.62, 6.87.
The mean foreign bond yield was 9.34 with variance 9.31. Assume that government bond yields are normally distributed.

At the 5 percent level of significance, test whether the government bond yields in the rest of the world during 1985 were lower than in the United States (state the null and alternative hypothesis, evaluate the test statistic, draw the decision criteria or evaluate the P-value, and state your conclusion).

10. For 528 random people sampled in the Current Population Survey, a researcher constructed a binary variable (Y=1 if "low wage earner") equal to 1 if person's wage rate was less than Rs.7.00 per hour and 0 otherwise. He ran a linear regression of Y on X1=1, X2 (education), and X3 (=1 if female, 0 otherwise). The results are tabulated below

Variable	**O.L.S. Coefficient Estimate**	**Mean**
$\mathbf{X_1}$	**1.05**	**1.0**
$\mathbf{X_2}$	**-0.05**	**13.09**
$\mathbf{X_3}$	**0.16**	**0.46**

What proportion of the sample were women? What proportion of the sample were "low wage earners"?

11.One of the central questions in labor economics is to estimate the rate of returns to education. A student has considered the following two models based upon cross-sectional data (standard errors are in parentheses):

MODEL 1:

$$\begin{aligned} \widehat{learn} &= \underset{(0.14)}{0.21} + \underset{(0.016)}{0.093} educ + \underset{(0.003)}{0.01} exp - \underset{(0.05)}{0.30} female + \underset{(0.06)}{0.06} married \\ R^2 &= 0.250, \ \overline{R}^2 = 0.241, \ n = 253 \end{aligned}$$

MODEL 2:

$$\begin{aligned} \widehat{learn} &= \underset{(0.16)}{0.14} + \underset{(0.011)}{0.095} educ + \underset{(0.0036)}{0.015} exp - \underset{(0.075)}{0.15} female + \underset{(0.08)}{0.17} married - \underset{(0.0001)}{0.0002} exp^2 - \underset{(0.08)}{0.20} f\_marr \\ R^2 &= 0.285, \ \overline{R}^2 = 0.273, n = 253 \end{aligned}$$

where *learn* = log(earnings), earnings are the average hourly earnings, *educ* is years of education, *exp* is the years of working experience, *female* is a dummy variable which is 1 for female and 0 otherwise, and *married* is another dummy variable that takes value 1 for married worker and 0 otherwise, and f_marr = female · married.

Based on Model 1, what is the percentage difference in earnings between a single male and a married female? Controlling education, which group of people makes more money than the other one on average?

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

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

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

Use the relevant test and test the hypothesis errors are homoscedastic at 5% level.

13. In a study of the determinants of the demand for computers, taking the following form:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 p_t + \alpha_3 m_t + u_t$$

$y_t - computer\_demand$

$x_t - national\_income$

$p_t - computer\_price$

$m_t - marketing\_\exp enditure$

A firm then wanted to determine if the price of a computer and the amount spent on marketing were jointly significant determinants. They then estimated the above unrestricted model to produce a RSS of 0.96 and a restricted version of the model (without price or marketing variables) and produced a RSS of 0.98. If there are 120 observations for the tests, are the price of a computer and marketing expenditure jointly significant? The table value of test statistic is 3.07 at 5% level of significance.

14. 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.

15. Given the following set of results based on a LPM, using 60 observations:

$$\hat{p}_i = 0.7 - 0.5d_t + 0.9f_t + 0.7y_t$$
$$(0.6) \quad (0.1) \quad (0.3) \quad (0.2)$$
$$\mathrm{R}^2 = 0.2, DW = 1.89$$

Where the dependent variable is whether a country defaults on its bank loans (1) or not (0). The explanatory variables are *d* (democracy), *f* (fixed exchange rate), *y* (income) (all in logs).

Interpret the coefficients on the above model. Are the individual variables significant?

End of PART –A

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 the following demand equation is estimated: $\ln q_t = \beta_0 + \beta_1 \ln p_t + \varepsilon_t$ (1)

where q_t is the per capital quantity consumed in period t (measured in grams) and p_t is the price in period t (measured in dollars/gram).

Suppose a sample of data for 42 time periods is used to obtain the following estimated demand equation (with standard errors in parentheses): $\ln \hat{q}_t = -5.0 - 0.4 \ln p_t$ (2)

(1.6) (0.2375) $\hat{\sigma}^2 = 0.18, R^2 = 0.49$

Answer the following questions.

a) Interpret β_1 in (1) using q_t and p_t. Interpret $R^2 = 0.49$ in (2).

b) Test the following hypothesis using the significance test approach at the 1% significance level.

$H_0 : \beta_1 \geq 0$

$H_1 : \beta_1 < 0$

What happens if one uses the interval approach at the 10% significance level?

c) Find the p-value of the test statistic in (b).

d) If q_t is now measured in kilograms, how do the change affect the two estimates in Model (2). Give your reasoning.

(4.0)

B2) Suppose that the Gurgavo Golf Associate (GGA) wants to compare the mean distances traveled by four brands of golf balls when struck by a driver. A completely randomized design is employed with Tigger Wedge, the GGA's robot golfer, using a driver to hit a random sample of 40 golf balls (10 balls for each brand) in a random sequence. The distance is recorded for each hit, and the resulting means are organized by brand in the following table:

	Mean
Brand A	251.28
Brand B	261.06
Brand C	269.95
Brand D	249.42

Judy Bates was hired by the GGA to determine whether the mean distances traveled by the four brands of golf balls were the same. To do this she estimated the following linear regression model:

$$\textbf{Distance}_i = \beta_0 + \beta_1 \cdot \textbf{BrandA}_i + \beta_2 \cdot \textbf{BrandB}_i + \beta_3 \cdot \textbf{BrandC}_i + \varepsilon_i$$

Where: Distance_i = Distance travelled for the ith golf ball

BrandA_i = Dummy variable equal to 1 for the distance travelled by BrandA golf balls; 0 otherwise.

BrandB_i = Dummy variable equal to 1 for the distance travelled by BrandB golf balls; 0 otherwise.

BrandC_i = Dummy variable equal to 1 for the distance travelled by BrandC golf balls; 0 otherwise.

Judy produced the following results for her USGA employers:

$$\text{Distance}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{BrandA}_i + \hat{\beta}_2 \cdot \text{BrandB}_i + \hat{\beta}_3 \cdot \text{BrandC}_i$$

(157.06) (0.83) (5.18) (9.14) *(t statistics in parentheses)*

$\bar{R}^2 = 0.73$ ESS = 2709.2 RSS = 907.86 TSS = 3617.06

a) Test, using Judy's results and a 5% level of significance, whether the brand of golf ball matters in explaining the mean distance travelled.

b) Judy forgot to include the values of her parameter estimates. Help her out by calculating the numerical values for each of the estimated coefficients above.

(6.0)

B3) Consider the following regression output using data from the Indian Engineering Student Lifestyle Survey for the year 2010 from IITs.

Dependent Variable: GPA
Method: Least Squares
Sample: 1 816
Included observations: 483

Variable		Coefficient	Std. Error
Intercept	$\hat{\beta}_0$	1.544003	0.222638
SAT	$\hat{\beta}_1$	0.001273	0.000166
SKIP	$\hat{\beta}_2$	-0.026975	0.006148
FEMALE	$\hat{\beta}_3$	0.055791	0.030676
CASTE	$\hat{\beta}_4$	0.077675	0.063062

R-squared	0.146135	Mean dependent var	3.261801
Adjusted R-squared	0.138989	S.D. dependent var	0.355621
S.E. of regression	0.329983	F-statistic	20.45178
Sum squared resid	52.04885	Prob(F-statistic)	0.000000

Where: GPA is the student's self-reported Grade Point Average.
SAT is the student's self-reported Exam Test score
SKIP is the number of classes skipped each semester
FEMALE is a dummy variable equal to 1 if the student is female; 0 otherwise
CASTE is a dummy variable equal to 0 if the student self-identifies as a person of SC/ST;1 otherwise.

a) Develop and complete hypothesis tests for each of the coefficients of this model, stating the null and alternative hypotheses. Include a sentence that outlines the intuition of your alternative hypothesis. Test your hypotheses at the 5% level of significance.

b) Use the regression output and correlation matrix below, and apply the specification criteria we studied to determine whether or not CASTE is an irrelevant variable in this equation.

Dependent Variable: GPA……Method: Least Squares …..Sample:1 816

Included observations: 492

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Intercept	1.637299	0.218032	7.509452	0.0000
SAT	0.001257	0.000163	7.730587	0.0000
SKIP	-0.026104	0.006021	-4.335478	0.0000
FEMALE	0.055360	0.030366	1.823066	0.0689

R-squared	0.135735	Mean dependent var	3.261951
Adjusted R-squared	0.130422	S.D. dependent var	0.354232
S.E. of regression	0.330325	F-statistic	25.54719
Sum squared resid	53.24800	Prob(F-statistic)	0.000000

	GPA	SAT	SKIP	FEMALE	CASTE
GPA	1.000000				
SAT	0.311218	1.000000			
SKIP	-0.173419	0.083665	1.000000		
FEMALE	0.062354	-0.121624	-0.112665	1.000000	
CASTE	0.111654	0.115445	-0.084311	0.062931	1.000000

(6.0)

B4) Consider the following model: $WAGE_t = \beta\ EDU_t + u_t$ where t = 1, 2, 3…T. Suppose we estimate the model by ordinary least squares(OLS) and get $\hat{\beta}_{OLS} = 2$ and $R^2 = 1$. Now suppose there is heteroscedasticity of the form $Var(u_t) = \sigma^2 Z_t^2$ where Z_t is any variable. If we use generaliased least squares GLS to estimate the model, can we say that $\hat{\beta}_{GLS} = 2$ and $R^2 = 2$ in the transformed model also equals one? If yes, prove it. If not, give a counter example.

(4.0)

B5)

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

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

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. Use the reported information to implement a parametric test for autocorrelated errors in this regression model. Use a significance level of 0.05 and state clearly the null and alternatives under test. Draw the inference.

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

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

Is there any evidence of 1st order autocorrelation?

(5.0)

******