Course: HUL 215 (Econometric Methods)
Instructor: Sujoy Chakravarty
Test: Major Exam (2 Hours)
Total Points: 54
Date: 3rd May, 2007

NAME- _____
GROUP -_____
ENTRY NUMBER - _____

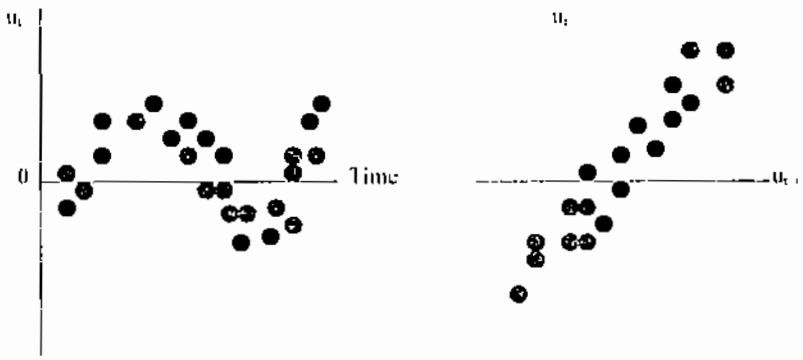
There are two (2) sections in this test. The first contains 12 multiple choice questions worth 1 point each. The second section contains 8 questions worth a total of 42 points , which involve calculation and interpretation. Please attempt all questions.

Section 1 - Multiple Choice Questions (12 points)

1. Heteroscedasticity in your data is a problem because:

a. ordinary OLS assumes that the data are homoscedastic and calculates the point estimates of regression parameters accordingly
b. ordinary OLS assumes that the data are homoscedastic and calculates the standard error estimates of the parameters accordingly
c. it is contagious
d. it biases the parameter point estimates.
e. It makes your estimates biased.

2. By inspection of the figure below you understand that



a. your data set suffers from perfect autocorrelation.
b. there is evidence of negative serial correlation.

c. it is an obvious case of heteroskedasticity because for large values of X the spread of the residuals is smaller than that of small values of X.
d. there is evidence of positive serial correlation.
e. none of the above.

3. You probably want to avoid log-log specifications if:

a. it is possible for Y or X to take on zero or negative values
b. the elasticity of Y with respect to X is one
c. you have values of Y which are large and values of X which are small
d. you cannot remember the definition of a logarithm
e. There is significant non-normality in the data.

4. When using dummy variables in dealing with a categorical variable, one needs to fit,

a. k dummies if a variable has k categorical values it can take.
b. only one dummy that increases monotonically.
c. (k-1) binary dummies
d. (k+1) binary dummies
e. none of the above.

5. Micronumerosity may occur when

a. We have only numerical data
b. We are trying to fit too few variables with a large number of observations.
c. The standard errors are very small.
d. The dependent variable is numeric and small.
e. We are trying to fit too many variables with too little data.

6. If we perform OLS estimation disregarding heteroscedasticity,

a. the OLS estimators are linear and biased
b. the OLS estimators are non-linear, unbiased but minimum variance
c. the OLS estimators are linear, unbiased and minimum variance.
d. the OLS estimators are linear, unbiased and not minimum variance.
e. the OLS estimators are linear and biased and maximum variance.

7. If a regression of quantity demanded (in million units) of candy bars, on different Geographical areas gives us the following output,

$\widehat{Q} = 42 - 12N + 23E + 42S$
(se) (12) (10) (22) (6)
n=54

Standard errors (se) given in parentheses. N, E and S stand respectively for the North, East and the Southern Regions. We can say,

a. The quantity demanded in the West is 42 million units.
b. The quantity demanded in the South is 80 million units.
c. The quantity demanded in the North is 56 million units.
d. The quantity demanded in the South is 84 million units.
e. Both a. and d.

8. Based on vaues and significance (at the 1% level) of parameters from the regression given above in question 7,

a. The quantity demanded in the North is 30 million units
b. The quantity demanded in the East is 65 million units.
c. The quantity demanded in the East is 42 million units.
d. The quantity demanded in the North is 40 million units.
e. None of the above.

9. A researcher desires to perform the Goldfield Quandt test on a dataset with 2 variables running the regression.
$Y_i = \beta_1 + \beta_2 X_i + u_i$ on 100 observations

He also assumes that the error structure is $\sigma_i^2 = \sigma^2 X_i^2$. The RSS obtained from partitioning the dataset into two groups of 40 and deleting 20 central observations are, $RSS_1 = 566.3$ and $RSS_2 = 768.2$. Assuming that
$\lambda = \frac{RSS_2/[(n-c)/2-k)]}{RSS_1/[(n-c)/2-k)]}$ follows the F distribution with appropriate degrees of freedom,

a. There is significant heteroscedasticity of the assumed form at the 5% level of significance
b. There is significant heteroscedasticity of the assumed form at the 1% level of

significance.
c. There is no heteroscedasticity of the assumed form in the dataset.
d. There is heteroscedasticity in the dataset but not of the form given above.
e. Both a. and b.

10. The linear probability model,

a. Assumes that the errors are logistically distributed.
b. Runs an OLS regression with dummy independent variables.
c. Assumes that probabilities of occurence of an event are piecewise linear.
d. is obtained by running an OLS regression with a binary dependent variable.
e. None of the above.

11. With respect to Logit and Probit models, one can say that,

a. They are Linear regression models with Normally distributed errors
b. They are Linear regression models with Poisson distributed errors.
c. They are Non-linear regression models that use the logistic and Normal cumulative distributions respectively to model choice probability
d. They have homoscedastic errors.
e. R^2 is a meaningful statistic that defines model fit.

12. Multicollinearity between different regressors in a model,

a. can be detected by taking simple pairwise correlations.
b. is usually signalled by a high R^2 coupled with very few significant variables.
c. necessarily makes the (X'X) matrix singular.
d. is always due to mistakes in data collection.
e. deflates the variance of regressors.

Section 2 - Short Answer Questions - 6 questions (48 POINTS)
PLEASE SHOW ALL OF YOUR WORKING

1. Given that RSS_t and RSS_{t-1} in a 4 variable time series regression with 38 observation are 286 and 192.4 respectively, is there significant serial correlation in this sample? (6 points) (SHOW YOUR WORKING CLEARLY)

2. Given results from 3 variable regression of 80 observations as

$$(X'X)^{-1} = \begin{bmatrix} .0072 & .012 & .025 \\ .012 & .0015 & .06 \\ .015 & .06 & .012 \end{bmatrix} \text{ and } X'y = \begin{bmatrix} 250 \\ 320 \\ 40 \end{bmatrix}$$

a. Find the coefficient estimates for $\widehat{\beta_1}$ $\widehat{\beta_2}$ and $\widehat{\beta_3}$ (1 point)

b. Assuming that the error variance is 2.1, test the hypothesis $H_0 : \beta_2 = 2\beta_3$ against the alternative hypothesis $H_0 : \beta_2 > 2\beta_3$ (3 points)

c. Construct a 99% confidence interval for β_2 (2 points)

3. A four variable regression yields the following results

$$\widehat{Y}_i = 0.67 + 2.056X_{2i} + 3.52X_{3i} + 4.66X_{4i}$$

se (0.36) (0.98) (1.8) (0.45)

The $R^2 = 0.48$ on the above regression. When we regress X_2 on X_3 and X_4 we get the following results,

$$\widehat{X}_{2i} = 3.678X_{3i} + 5.95X_{4i}$$

se (1.45) (3.23) R^2=0.68

a. What is the probable phenomenon at work in the regressions above? What in the above regressions is/are indicative of this phenomenon (3 points)

b. Name 2 ways we can attempt to deal with this phenomenon. Can some of these methods cause undesirable effects? Explain briefly (3 points)

4. A researcher is studying the effect of X on Y using a time series of 80 observations.

Assuming an Autoregressive relationship of the first order (AR -1 scheme) he models $u_t = \rho u_{t-1} + e_t$ where e_t is a white noise error term.

On regressing u_t on u_{t-1}, he obtains $\hat{u}_t = .42u_{t-1}$ and $\hat{\sigma}_e^2 = 2.56$

Using the information above construct the Var-Cov matrix of errors assuming that ONLY 1st order serial correlation exists with higher order autocovariances equal to zero. Furthermore assume that the error variance on the original Y on X regression is homoscedastic with $\widehat{\sigma_u^2} = 0.85$. (3 points)

5. An ANOVA table for a 3 variable regression with n=35 yields the following information:

SOURCE	SS	df
ESS due to X_2	114.89	1
ESS due to X_2 and X_3	34.84	2
RSS (of 3 var. model)	494.50	38

Is the contribution from adding X_3 significant to the model at the 1 % level of significance? (3 points)

6. A researcher wants to determine whether there is a structural change in his data in the period 1985-2007 as compared to 1960-1984. He is investigating the effect of farm support in crores of Rupees on the poverty rate measured in % of people in the population consuming below 2400 calories per day. He has time series (yearly) data on the poverty rate and farm support from 1960-2007.

doing so the researcher runs the following regression

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

where, Y = poverty rate, X = Farm Support Level, t= year.

D= 1 if observation is in the period 1985 -2007
= 0 if the observation is in the period 1960 - 1984

He obtains the regression line (with standard errors in parentheses),

$\widehat{Y}_i = .45 + .54D_i + .26X_i + .10X_iD_i$
se (.19) (.27) (.23) (.02)

$R^2 = .82$

Are the two regression relationships (1960 -1984) and (1985-2007) parallel (at 5% level of significance)? SHOW YOUR WORKING (6 points)

7.From a logit estimation of binary responses on whether an individual bought a car or not, Maruti-Suzuki wants to use a basket of demographic individual characteristics to predict the probability that an individuai will purchase a car. Using cross sectional data on 560 individuals in the Delhi area, and four characteristics, age (in years), gender (M=1 /F=0), household income (in lakhs) and whether an individual already owns a car or not (1/0 binary variable) the research department estimates the following logit. The significance (p) values on the z-statistic on the variables is given onder the regression line

$\widehat{Y}_i = 1.45AGE_i + .68GENDER_i + 0.76INCOME_i + 1.50OWNSCAR_i$
sig. (.025) (.013) (0.0001) (.08)

a. Using only variables that are significant at the 5% level calculate the probability of car purchase of a Male 35 year old with an income of 3 who owns a car (4 points)

b. What is the change in probability for a unit change in income for a person with an income 3 ? (4 points)

8. 3. In a 3 variable regression the correlation coefficient between X_2 and X_3 is given as 0.42. Another sample taken from the same population gives the corresponding correlation coefficient to be 0.50. Given that the error variance is equal (to 0.60) for both, find the Variance of the coefficient $\widehat{\beta_2}$ for both samples, given that $\sum x_i^2 = 235.8$ for the first sample and $\sum x_i^2 = 278.3$ for the second. (4 points) (SHOW YOUR WORKING CLEARLY !!!)