Examination: Economics IV (5027) = 1560

Introduction to Econometrics (Economics IV)

Winter Semester 2003/2004

Dr. John E. Brennan

You are allowed to use a non-programmable calculator (in accordance with the instructions given by the examination office) and a translating dictionary from your native language to English (without any notes written into it). **All** of the **seven** (7) examination questions must be answered (the estimated time to spend on each question is given). This examination consists of **four** (4) pages.

Only work that is presented in a neat and orderly manner that is easy to read will be graded. All calculations should be rounded to four places following the decimal point (e.g., 15.6429)

Question 1 (25 Minutes)

In the "pure" hetroscedastic case the (n x n) variance covariance matrix of the n univariate conditional random variables, $(y_i \mid X_i)$, cannot be written as: $\sigma^2_{y|X}\,\mathbf{I}$

a. Draw a picture of the (n x n) Var-Cov $(\mathbf{y} \mid \mathbf{X})$ matrix and discuss the numerical values of Cov $[(y_i \mid X_i), (y_j \mid X_j)]$ when $i = j$ and when $i \neq j$ for $i = 1, 2, \ldots, j, \ldots, n$

b. Consider a sample drawn from a bivariate population. Order the sample according to the magnitude of the X variable. Conduct an OLS regression on this sample and plot the prediction error on the vertical axis and the X variable of the horizontal axis. Explain how an inspection of this graph could be helpful in determining if the sample came from a population where homoscedasticity is present.

c. Describe a formal test that could be used on a sample to test for the presence of hetroscedasticity.

d. If OLS is used on a sample drawn from a population where hetroscedasticity is present discuss the properties of the "best" MSE estimator, **C**.

e. Assume that a sample of size n has been drawn from a bivariate population and the n conditional variances are known, Var $(y_i \mid X_i)$. Describe how this sample could be estimated using GLS. What are the properties of the GLS estimator, $\mathbf{C}_{GLS}$.

Question 2 (10 Minutes)

The Linear Probability Model states that in the population $E(y_i \mid X_i) = X_i\,\beta = \Pr[(y_i \mid X_i) = 1]$

a. Discuss this model in terms of its theoretical shortcomings and estimation difficulties.

b. If a sample from this population is used to estimate the population parameters, β, will the estimates obtained, **c**, be unbiased and minimum variance?

c. Outline a GLS procedure for this model.

Please turn to page 2

Question 3 (30 Minutes)

Given a sample of 47 quarterly observations (y_t, x_{t2}, x_{t3}, x_{t4}). The following matrices were calculated:

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 35.2468 & -0.0225 & 0.0936 & 1.3367 \\ -0.0225 & 0.00156 & -0.00134 & 1.1462 \\ 0.0936 & -0.00134 & 0.00143 & -0.00862 \\ 1.3367 & 1.1462 & -0.00862 & 0.05403 \end{bmatrix} \qquad \mathbf{X'y} = \begin{bmatrix} 27.11 \\ 6.12 \\ 68.14 \\ 21.11 \end{bmatrix}$$

$\Sigma_t (y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4})^2 = 533108.78$

$\Sigma_s [(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4})(y_{t-1} - c_1 - c_2 x_{t-12} - c_3 x_{t-13} - c_4 x_{t-14})] = -203647.55$

$\Sigma_s[(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4})(y_{t-1} - c_1 - c_2 x_{t-12} - c_3 x_{t-13} - c_4 x_{t-14})]^2 = 1.3755\text{E}10$

$\Sigma_s[(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4}) - (y_{t-1} - c_1 - c_2 x_{t-12} - c_3 x_{t-13} - c_4 x_{t-14})]^2 = 1482042.41$

$\Sigma_s[(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4}) - (y_{t-1} - c_1 - c_2 x_{t-12} - c_3 x_{t-13} - c_4 x_{t-14})] = 1217.39$

a. Calculate the values in the (4 x 1) vector **c**.
b. Given that $\omega = 0.05$, $t_{(43)} = 2.017$, can you accept the null hypothesis that in the population $\beta_4 = 0$?
c. Calculate the sample correlation coefficient $\rho\^{} = \text{Corr}\,[(y_t | X_t), (y_{t-1} | X_{t-1})]$
d. Calculate the Durbin-Watson d statistic and test the H_o: There is no positive or negative autocorrelation present. Explain what you have discovered in terms of the OLS estimates that you calculated in part (a) of this Question. Was the hypothesis test conducted in part (b) of this Question correct? [$n = 47$, $k' = 3$: $d_U = 1.666$ and $d_L = 1.383$]
e. Define an AR (1) autoregressive process.
f. Outline a GLS estimation procedure when AR (1) is present.

Question 4 (15 Minutes)

The Runs Test (Geary Test) is a nonparametric test that can be used to check if the conditional random variable $(y_t | X_t)$ is correlated with $(y_{t-1} | X_{t-1})$. An OLS regression was run on the sample and the following prediction errors, e_t, were calculated.

T	e_t	t	e_t	t	e_t
1965	0.1875	1970	0.2134	1975	0.2187
1966	– 0.3482	1971	– 0.0421	1976	0.0463
1967	– 0.0173	1972	– 0.6234	1977	– 0.2419
1968	– 0.4713	1973	0.1267	1978	– 0.1234
1969	– 0.0964	1974	0.1167	1979	– 0.0853

a. Compute and explain the rationale behind the Runs (Geary) Test.
   $E(\delta) = [(2\, n_1\, n_2) / (n_1 + n_2)] + 1$
   $V(\delta) = \{2\, n_1\, n_2\, [(2\, n_1\, n_2) - n_1 - n_2]\} / \{(n_1 + n_2)^2\, (n_1 + n_2 - 1)\}$
b. Can the null hypothesis be accepted in this case? Explain your answer fully.

Please turn to page 3

Question 5 (15 Minutes)

Given the discrete bivariate probability distribution for the pair of random variables X and Y.

Y \ X	0.8	2.6	4.8	6.2	f_2 (y)
0.73	0.0914	0.0642	0.0542	0.1239	
0.86	0.0655	0.1144	0.0719	0.0644	
0.94	0.1252	0.0655	0.0732	0.0862	
f_1 (x)					

Predict a value for the random variable Y using the "best" minimum mean squared error predictor (calculate the specific values in each case).

a. What is your "best" prediction for the value of Y when X = 6.2?

b. What are your "best" linear predictions of Y knowing that X = 6.2?

c. What is the "best" prediction when the value of X is unknown?

d. Is Y mean independent of X? Explain your answer in complete detail.

Question 6 (15 Minutes)

An often used measure of "goodness of fit" in a sample is called the coefficient of determination, R^2.

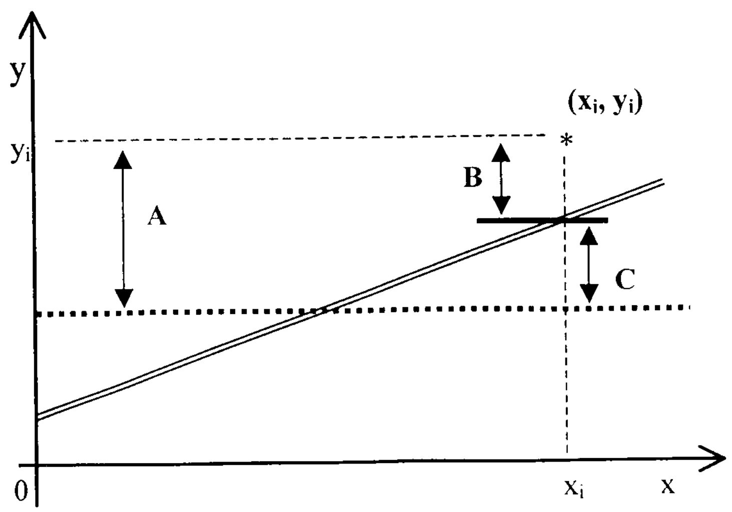
$R^2 = \{\Sigma_i [E(y_i | X_i) - \bar{y}]^2\} / \{\Sigma_i (y_i - \bar{y})^2\}$

a. Is this equation valid when $E(y_i | X_i) = E^{**}(y_i | X_i) = BPP$? Explain your answer in detail.

b. Explain the situation when $R^2 = 1.0$.

c. Explain the situation when $R^2 = 0.0$.

d. Referring to the picture below, explain how R^2 measures "goodness of fit" in a sample.



Please turn to page 4

Question 7 (10 Minutes)

Three major problems for OLS regression are (1) multicollinearity, (2) hetroscedasticity, and (3) autocorrelation.

a. What are the consequences of "perfect" and "near" multicollinearity for the OLS estimate vector **c**?

b. What are the consequences of hetroscedasticy for the OLS estimate vector **c**?

c. What are the consequences of autocorrelation for the OLS estimate vector **c**?

Define the problem in each case and describe the estimation problems in detail. Can you prescribe a solution for each of these estimation problems?

This is the end of the examination.

<u>GOOD LUCK !!</u>