Original

Examination: 5027
Economics III
Introduction to Econometrics
Summer Semester 2006
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 **ten** (10) examination questions must be answered (the estimated time to spend on each question is given). This examination consists of **four** (4) pages and must be completed within 120 minutes.

Answers that are NOT presented in a NEAT and ORDERLY manner that is easy to read will NOT be graded. All calculations should be rounded to four places following the decimal point (e.g., 15.6429)

Question 1 (20 Minutes)

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

Y \ X	8.2	12.1	16.4	21.8	f_2 (y)
1.7	0.0350	0.0580	0.0440	0.1247	
3.6	0.0810	0.0910	0.0583	0.0840	
8.2	0.1430	0.0500	0.0650	0.1660	
f_1 (x)					

a. What is the "best" MSE prediction for the value of Y when X = 16.4?
b. Using the BLP, the "best" MSE prediction of Y given that X = 16.4 is?
c. Using the BPP, the "best" MSE prediction of Y given that X = 16.4 is?
d. Do both the BLP and the BPP produce a prediction, c, where c = E(Y), when the value of the random variable X = x, is equal to: x = E(X)? Explain your answer in detail.
e. Which of the two prediction methods, the BLP or the BPP, produces the "best" prediction? Explain your answer.

Question 2 (10 Minutes)

The Runs Test (Geary Test) is a nonparametric test that can be used to check if the conditional random variable $(Y \mid X = x_t)$ is correlated with $(Y \mid X = x_{t-1})$.

(– – – –)(+ +)(– – – –)(+ + + + + + + + + + + + +)(– – –)(+)(– – – – – – – – – – –)(+ +)

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 2

Question 3 (10 Minutes)

Consider the continuous Rectangular (uniform) univariate probability distribution with parameters $a < b$, $f(x) = 1 / (b - a)$ for $a \leq x \leq b$, with $f(x) = 0$ elsewhere. Calculate the probability of each of the following events occurring over the interval [0, 3].

a. $A = \{0 \leq X \leq ½\}$

b. $B = \{X \leq ½\}$

c. $C = \{X = ½\}$

d. $D = \{¼ \leq X \leq ¾\}$

e. $E = \{1 < X \leq 2\}$

Question 4 (10 Minutes)

Consider an annual time-series sample consisting of the 34 observed values of the random variables Y, X, and W: (y_t, x_t, w_t) for t = 1972, 1973, …, 2005. An estimate of the correlation coefficient, $\hat{c}$, between the univariate conditional random variables $(Y \mid X_t = x_t, W_t = w_t)$ and $(Y \mid X_{t-1} = x_{t-1}, W_{t-1} = w_{t-1})$ can be obtained from the Durbin-Watson d statistic:

$$d = 2\,(1 - \hat{c})\,.$$

f. Explain in detail all the steps necessary to calculate the Durbin-Watson statistic, d, using this sample data.

g. After completing your work in part (a) of this question, you obtained the Durbin-Watson statistic, $d = 2.7135$, with $n = 34$, $k' = 2$, $d_L = 1.333$ and $d_U = 1.580$. Perform the Durbin-Watson test and report your result.

h. Explain WHY the Durbin-Watson d statistic is a number near 2.0 when null hypothesis is not rejected. Why does $0 \leq d \leq 4.0$?

Question 5 (10 Minutes)

The Linear Probability Model states that in the population:
$E(Y \mid Z = z_i, W = w_i) = \beta_1 + \beta_2 z_i + \beta_3 w_i = p_i$ and $V(Y \mid Z = z_i, W = w_i) = p_i (1 - p_i)$ where $p_i = Pr(Y = 1 \mid Z = z_i, W = w_i)$. The dichotomous random variable Y signifies automobile ownership by individuals, the continuous random variable Z measures disposable income, and the dummy variable W is equal to one if the individual has an advanced degree from an institution of higher learning.

a. Critically discuss this model and any estimation problems that might be involved.

b. Is it likely that autocorrelation is present due to the inclusion of two qualitative variables in this model?

c. What is multicollinearity and is it a problem with probability models in general? What about this case?

d. Explain how to estimate this model specification as a Probit Model.

Please turn to Page 3

Question 6 (10 Minutes)

Sample estimates, **c**, of the (k x 1) vector β of population parameters can be made subject to q linear constraints imposed by:

R c = r, where **R** is a matrix (q x k), $q \leq k$ and **r** is a (q x 1) vector.

$$BLP = c_1 + c_2\, x_2 + c_3\, x_3 + c_4\, x_4 + c_5 x_5$$

Write down the matrix **R** and the vector **r** for the following constraints:

a. $c_2 - c_3 = 0$ and $c_4 + c_5 = 1$

b. $c_2 = c_3 = c_4 = 2$

c. $c_2 - 3\, c_3 = 5\, c_4$

d. $c_2 = c_3$ and $c_4 = c_5$

Question 7 (15 Minutes)

Two estimation problems, that are often prevalent in economic data, can cause severe problems when interpreting the results of the BLP estimated coefficients. These problems are hetroscedasticity and autocorrelation.

a. Explain what hetroscedasticity is and exactly the kind of problems it causes with the OLS estimated (kx1) vector **c**? In what kind of data is it likely to appear?

b. Give a description of a GLS procedure to correct this problem.

c. What are the consequences of autocorrelation for the OLS estimate vector **c**? What kind of data is likely to have this problem and why?

d. Define and explain why the First-Order Autoregressive Process, AR(1), allows the implementation of GLS as a solution to this problem.

Question 8 (10 Minutes)

A group of engineering students can be categorized in the following manner. Forty-five percent of them are students from African countries. Concerning the students from African countries, fourteen percent of them live on the campus. Sixteen percent of the students from the rest of the world live off the campus. What is the probability that an engineering student selected at random is:

a. A student from an African country who is living off campus.

b. A student from an African country who is living on the campus.

c. A student not from an African country living on the campus or a student from an African country living off campus.

d. The engineering student who is living on the campus.

e. A student from an African country given that we know that he/she lives on the campus.

Please turn to Page 4

Question 9 (10 Minutes)

Consider a cross-sectional sample of 20 observations where the random variable Y is average weekly expenditure on food and the random variable X is average household weekly income. The sample was divided in the middle according to the value of the explanatory variable:

$$\det(\mathbf{X_A}'\,\mathbf{X_A}) = 50353.5731$$

$$(\mathbf{X_A}'\,\mathbf{X_A})^{-1} = \begin{bmatrix} 61908.51/\det & -1089.87/\det \\ -1089.87/\det & 20/\det \end{bmatrix} \qquad \mathbf{X_A}'\,\mathbf{y_A} = \begin{bmatrix} 406.68 \\ 22992.44 \end{bmatrix}$$

$$\Sigma_i\,(y_{Ai} - c_{A1} - c_{A2}\,x_{Ai})^2 = 402.785$$

$$\det(\mathbf{X_B}'\,\mathbf{X_B}) = 68707.8771$$

$$(\mathbf{X_B}'\,\mathbf{X_B})^{-1} = \begin{bmatrix} 148297.72/\det & -1702.13/\det \\ -1702.13/\det & 20/\det \end{bmatrix} \qquad \mathbf{X_B}'\,\mathbf{y_B} = \begin{bmatrix} 537.1 \\ 46442.60 \end{bmatrix}$$

$$\Sigma_i\,(y_{Bi} - c_{B1} - c_{B2}\,x_{Bi})^2 = 1348.80$$

a. The Goldfeld-Quandt Test is based on an F-test with $[(n / 2) - k]$ df in both the numerator and denominator. Explain the null hypothesis for this test.
b. Given a 0.05 critical value of 3.44 for an F-test with the appropriate number of df, conduct the Goldfeld-Quandt Test on the sample of 20 observations given in this question. Report your results and explain what you have learned from this test.

Question 10 (15 Minutes)

Assume that the bivariate random vector (Z, W) has a joint pmf or pdf f (z, w). In the "pure" hetroscedastic case the (n x n) variance covariance matrix of the n univariate conditional random variables, $(W \mid Z = z_i)$, cannot be written as: $\sigma^2_{W|Z}\,\mathbf{I}$

i. Draw a picture of the (n x n) Var-Cov $(W \mid Z)$ matrix and discuss why the Cov $[(W \mid Z = z_i), (W \mid Z = z_j)]$ when $i \neq j$ cannot be replaced with the correlation coefficients.

j. The GLS estimator can be written as: $\mathbf{c}_{GLS} = (\mathbf{Z}^{*\prime}\,\mathbf{Z}^{*})^{-1}\,\mathbf{Z}^{*\prime}\,\mathbf{w}^{*}$. It can also be written as $(\mathbf{Z}'\,\mathbf{P}'\,\mathbf{P}\,\mathbf{Z})^{-1}\,\mathbf{Z}'\,\mathbf{P}'\,\mathbf{P}\,\mathbf{w} = (\mathbf{Z}'\,\mathbf{V}^{-1}\,\mathbf{Z})^{-1}\,\mathbf{Z}'\,\mathbf{V}^{-1}\,\mathbf{w}$. Explain why these expressions are equivalent.
k. Given a sample of size n from a bivariate population where the values of the n conditional variances are known, Cov $[(W \mid Z = z_i), (W \mid Z = z_i)]$. Outline the GLS estimation procedure to be used on this sample. What are the properties of the GLS estimator, $\mathbf{C}_{GLS}$.

This is the end of the examination.

GOOD LUCK !!