Examination: 5027
Economics III
Introduction to Econometrics
Summer Semester 2009
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 **twelve** (12) examination questions must be answered. This examination consists of **four** (4) pages and must be completed within 120 minutes.

NOTE: Answers that are NOT presented in a NEAT and ORDERLY manner, 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 (10 Minutes)

The Rectangular (uniform) univariate random variable X= x with pdf, $f(x) = 1/(b - a)$, with $a < b$ for $a \le x \le b$, with $f(x) = 0$ elsewhere. Assume: $a = -1$ and $b = 3$.

a. Pr (A) = $\{0 \le X = x \le 2.5\}$

b. Pr (B) = $\{X = x \le 0.75\}$

c. E (X) and V (X)

Question 2 (10 Minutes)

Given the discrete bivariate probability distribution for the pair of random variables X= x and Y= y, f (x, y):

Y= y \ X= x	1	3	9	f_2 (y)
2	0.020	0.080	0.100	
4	0.030	0.120	0.150	
6	0.050	0.200	0.250	
f_1 (x)				1.00

Predict a value for the random variable Y= 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= y with no knowledge of X?

b. What are your "best" linear predictions of Y= y knowing the values of X= x?

Question 3 (10 points)

Consider the discrete random variables (X, Y) with joint pmf: f (x, y) where X= x and Y= y.

a. Explain in detail how the probabilities associated with the conditional random variables (Y | X= x), g_1(y | X= x), relate to the joint probabilities, f (x, y).

b. When are the conditional probabilities (Y | X= x), g_1(y | X= x), equal to the marginal probabilities of Y= y, f_2 (y).

Please turn to page 2

Question 4 (10 Minutes)

Consider a random sample of 50 observations (x_i, y_i). The random variable Y= y is average monthly expenditure on entertainment (in EUR) and the random variable X is average monthly income (also EUR). Based on the results of a Goldfeld-Quandt Test it was concluded that hetroscedasticity exists in the sample. Assume that $v_i = x_i^2$ and that the population regression function is linear in the parameters, $E^*(Y \mid X = x) = \alpha + \beta x$.

The following matrices were calculated:

$$(X' V^{-1} X)^{-1} = \begin{bmatrix} 806.553308 & -12.68210546 \\ -12.68210546 & 0.224411245 \end{bmatrix}$$

$$X' V^{-1} y = \begin{bmatrix} 0.222466303 \\ 13.84434804 \end{bmatrix}$$

(Note: This value was calculated using the correct $\mathbf{c_{GLS}}$)

$$n\, s^2_{Y^*|X^*} = 0.3044024$$

a. Calculate the GLS estimates, $\mathbf{c_{GLS}}$, of the population parameters α and β.

b. Are these GLS estimates unbiased and minimum variance? Explain your answer.

c. Calculate the Var-Cov matrix for the vector $\mathbf{c_{GLS}}$.

Question 5 (10 Minutes)

In the pure hetroscedastic case the y_i's are uncorrelated with each other, but their conditional distributions have different variances.

a. Describe the logic behind the Goldfeld-Quandt Test

b. What is the null hypothesis in this test?

Question 6 (10 Minutes)

When time series data are used in estimating models autocorrelation is often a problem. Use the following data to calculate the Durbin-Watson d statistic and to answer the questions:

$$\Sigma_t (y_t - c_1 - c_2 x_t)^2 = 34.65905784$$

$$\Sigma_t [(y_t - c_1 - c_2 x_t) - (y_{t-1} - c_1 - c_2 x_{t-1})]^2 = 46.771496$$

a. With n = 20, k' = 1, d (upper) = 1.411, and d (lower) = 1.201, can you accept the null hypothesis that there is no positive or negative serial correlation in the regression that was performed? Explain your answer.

b. If autocorrelation is present, what effect does it have on the estimated OLS coefficients? Explain how to correct for autocorrelation in estimation.

Please turn to page 3

Question 7 (10 Minutes)

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

$$\mathbf{R}\,\mathbf{c} = \mathbf{r}$$

Where **R** is a matrix (q x k), $q \le k$ and **r** is a (q x 1) vector

if:

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

Write down the matrix **R** and the vector **r** that would impose the following constraints:

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

b. $c_2 + c_3 = 0$ and $c_4 - c_5 = 1.0$

c. $c_2 - 7\,c_3 = 6\,c_4$

Question 8 (10 Minutes)

The Runs Test (Geary Test) is a nonparametric test that can be used to check if the random variable $[y_t - E^*(Y \mid X = x_t)]$ is random over time.

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

a. Compute and explain the rationale behind the Runs (Geary) Test.

$$E(\delta) = [(2\, n_1 n_2) / (n_1 + n_2)] + 1$$

$$V(\delta) = \frac{2\, n_1 n_2\, (2\, n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2\, (n_1 + n_2 - 1)}$$

b. Can we accept the null hypothesis in this case? Explain your answer fully.

Question 9 (10 Minutes)

Stochastic independence and mean independence both imply that $C(X, Y) = 0$.

a. Explain the difference between stochastic and mean independence.

b. If X= x and Y= y are stochastically independent, explain the implication for the CEF and the BLP.

Question 10 (10 points)

At the faculty of Economics and Management, 75% of the baccalaureate students in the English language program take the course Economics I, 65% of them take the course Economics II, and 85% of them take at least one of the two courses.

a. What proportion of students takes both courses?

b. What is the probability that a student will take Economics II given that he or she has taken Economics I?

c. Are the decisions made by students to take these two courses independent?

Please turn to page 4

Question 11 (10 Minutes)

Consider a random sample of 15 annual observations (y_t, $x_{t\,2}$, $x_{t\,3}$):

det $\mathbf{X'\,X}$ = 306223760 Calculate estimates of the population BLP = $E^*(\mathbf{Y} \mid \mathbf{X}) = \beta\,\mathbf{X}$

$$\mathbf{X'\,X} = \begin{bmatrix} 15 & 31895 & 120 \\ 31895 & 68922513 & 272144 \\ 120 & 272144 & 1240 \end{bmatrix} \quad (\mathbf{X'\,X})^{-1} = \begin{bmatrix} 37.232772 & -0.0225081 & 1.3367066 \\ -0.0225081 & 0.0000137 & -0.0008319 \\ 1.3367066 & -0.0008319 & 0.054035 \end{bmatrix}$$

$$\mathbf{X'\,y} = \begin{bmatrix} 29135 \\ 62905821 \\ 247934 \end{bmatrix} \qquad \Sigma_t\,(y_t - c_1 - c_2\,x_{t\,1} - c_3\,x_{t\,2})^2 = 1976.85539$$

(based on the correct values of the parameter estimates)

a. Calculate the vector **c**.
b. Calculate the matrix Var-Cov (**c**).
c. Given that $\omega = 0.05$, $t_{(12)} = 2.179$, can you accept the null hypothesis that in the population the vector β is equal to the null vector **0**?
d. What are the consequences associated with multicollinearity in OLS regression. Explain your answer in detail discussing the difference between "perfect" and "near" multicollinearity.

Question 12 (10 points)

There are n different sets, i = 1, 2, …, n, of the (k-1) explanatory variables, j = 2, 3, …, k. The i^{th} set would be written as: ($X_2 = x_{i2}$, $X_3 = x_{i3}$, …., $X_k = x_{ik}$). Consequently, the notation, $X_j = x_{ij}$, means that the (k-1) explanatory variables are observed with their i^{th} value. Consider the (n x n) covariance matrix of the n univariate conditional random variables, $(Y|X_j = x_{ij})$:

Cov (Y | **X**) =
(n x n)

	1	2	…	i	…	n
1	**$V(Y\|X_j = x_{1j})$**	$C(Y\|X_j = x_{1j}, Y\|X_j = x_{2j})$		$C(Y\|X_j = x_{1j}, Y\|X_j = x_{ij})$		$C(Y\|X_j = x_{1j}, Y\|X_j = x_{nj})$
2	$C(Y\|X_j = x_{2j}, Y\|X_j = x_{1j})$	**$V(Y\|X_j = x_{2j})$**		$C(Y\|X_j = x_{2j}, Y\|X_j = x_{ij})$		$C(Y\|X_j = x_{2j}, Y\|X_j = x_{nj})$
i	$C(Y\|X_j = x_{ij}, Y\|X_j = x_{1j})$	$C(Y\|X_j = x_{ij}, Y\|X_j = x_{2j})$		**$V(Y\|X_j = x_{ij})$**		$C(Y\|X_j = x_{ij}, Y\|X_j = x_{nj})$
n	$C(Y\|X_j = x_{nj}, Y\|X_j = x_{1j})$	$C(Y\|X_j = x_{nj}, Y\|X_j = x_{2j}$		$C(Y\|X_j = x_{nj}, Y\|X_j = x_{ij})$		**$V(Y\|X_j = x_{nj})$**

a. Completely explain what we mean by the term "Hetroscedasticity" and what it tells us in a model where Y = Savings Rate and X = Family Income.

b. Explain the conditions that are necessary for the covariances in this matrix to be exchanged for correlation coefficients.

c. Completely define the First-Order Autoregressive Process, AR(1). With AR(1), What is the value of Corr $[(Y \mid X_j = x_{3j}), (Y \mid X_j = x_{8j})]$?

This is the end of the examination.
<u>GOOD LUCK !!</u>