Examination: 5027
Economics III
Introduction to Econometrics
Winter Semester 2009 / 2010
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 (the estimated time to spend on each question is given). 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)

Consider the continuous univariate random variable X= x ~ Rectangular with $a < b$,

$$f(x) = 1 / (b - a) \text{ and } F(x) = (x - a) / (b - a) \text{ for } a \leq x \leq b.$$

Calculate the following using $a = -2$ and $b = 6$.

a. $\Pr\{0.5 \leq X = x \leq 1.0\}$

b. $\Pr\{0.5 \leq X = x \leq 1.0 \mid X = x \leq 2\}$

Question 2 (10 Minutes)

Given the discrete bivariate probability distribution, $(X, Y) \sim f(x, y)$.

Y \ X	1	3	9	$f_2(y)$
2	0.018	0.080	0.100	
4	0.030	0.122	0.150	
6	0.050	0.200	0.250	
$f_1(x)$				1.00

a. What is your "best" prediction for the value of the random variable Y= y with no knowledge of the random variable X= x?

b. What is your "best" prediction for the value of the random variable $(Y \mid X = 1)$?

c. Is Y stochastically independent of X? Explain your answer in complete detail.

Please turn to page 2

Question 3 (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= 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 and that the population regression function is linear in the parameters, $E^*(Y \mid X= x) = \alpha + \beta x$.

The following matrices were calculated:

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

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

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

b. If OLS estimates, $\mathbf{c}_{OLS}$, had been calculated, $(\mathbf{X'} \mathbf{X})^{-1} \mathbf{X'} \mathbf{y}$, what problems would exist in these estimates?

Question 4 (10 Minutes)

Consider the random variables Y= y and X= x with joint pdf: $(X, Y) \sim f(x, y)$.

a. Explain in detail how the univariate random variable Y= y differs from the univariate conditional random variable $(Y \mid X= x)$.

b. Under what conditions will these two random variables be exactly the same?

Question 5 (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:

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

$$\Sigma_t [(y_t - c_1 - c_2 x_t)(y_{t-1} - c_1 - c_2 x_{t-1})]^2 = 153.524278$$

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

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 6 (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), $q \leq k$ 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 7 (10 Minutes)

The Runs Test (Geary Test) is a nonparametric test that can be used to test for the presence of autocorrelation in a sample.

(− − − −)(+ +)(− − − −)(+ + + + + + + + + + + + + +)(−)(+)(− − − − − − − − − − − − − − −)

a. 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 8 (10 Minutes)

A sample measure of "goodness of fit" is the coefficient of determination, R^2.

a. Why is R^2 thought of as a measure of the linear association between y_i and the sample estimate of the conditional expectation, $E^*(Y \mid X = x_i)$?

b. Explain in detail why R^2 should not be used when sample estimation of the BPP has been conducted.

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 and Y are stochastically independent, explain the implication for the CEF and the BLP.

Please turn to page 4

Question 10 (10 Minutes)

Consider a random sample of 15 annual observations (y_t, x_{t2}, x_{t3}). The following matrices were calculated:

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

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

$$\mathbf{X'y} = \begin{bmatrix} 29135 \\ 62905821 \\ 247934 \end{bmatrix} \qquad \Sigma_t\,(y_t - c_1 - c_2\,x_{t1} - c_3\,x_{t2})^2 = 1976.85539$$

(based on the correct values of the parameter estimates)

a. Calculate the vector **c**.
b. What are the consequences associated with multicollinearity in OLS regression. Explain your answer in detail discussing the difference between "perfect" and "near" multicollinearity.

Question 11 (10 Minutes)

Var-Cov($Y \mid \mathbf{X}$) = (n x n)

	1	2	...	t	...	n
1	$\mathbf{V(Y \mid X_{1j}=x_{1j})}$	$C(Y \mid X_{1j}=x_{1j}, Y \mid X_{2j}=x_{2j})$		$C(Y \mid X_{1j}=x_{1j}, Y \mid X_{tj}=x_{tj})$		$C(Y \mid X_{1j}=x_{1j}, Y \mid X_{nj}=x_{nj})$
2	$C(Y \mid X_{2j}=x_{2j}, Y \mid X_{1j}=x_{1j})$	$\mathbf{V(Y \mid X_{2j}=x_{2j})}$		$C(Y \mid X_{2j}=x_{2j}, Y \mid X_{tj}=x_{tj})$		$C(Y \mid X_{2j}=x_{2j}, Y \mid X_{nj}=x_{nj})$
t	$C(Y \mid X_{tj}=x_{tj}, Y \mid X_{1j}=x_{1j})$	$C(Y \mid X_{tj}=x_{tj}, Y \mid X_{2j}=x_{2j})$		$\mathbf{V(Y \mid X_{tj}=x_{tj})}$		$C(Y \mid X_{tj}=x_{tj}, Y \mid X_{nj}=x_{nj})$
n	$C(Y \mid X_{nj}=x_{nj}, Y \mid X_{1j}=x_{1j})$	$C(Y \mid X_{nj}=x_{nj}, Y \mid X_{2j}=x_{2j}$		$C(Y \mid X_{nj}=x_{nj}, Y \mid X_{tj}=x_{tj})$		$\mathbf{V(Y \mid X_{nj}=x_{nj})}$

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

b. Completely define the First-Order Autoregressive Process, AR(1). What is the value of Corr $[(Y \mid X_{tj}= x_{tj}), (Y \mid X_{t-5j}= x_{t-5j})]$?

Question 12 (10 Minutes)

In the Linear Probability Model: $E(Y \mid Z= z_i, W= w_i) = \beta_1 + \beta_2 z_i + \beta_3 w_i = p_i$ and $V(Y \mid Z= z_i, W= w_i) = p_i (1 - p_i)$ where $p_i = \Pr(Y = 1 \mid Z= z_i, W=w_i)$.

a. Critically discuss any estimation problems that might be involved in the model.

b. Is autocorrelation likely to be present?

This is the end of the examination.

GOOD LUCK !!