Examination: 5027
Economics III
<u>Introduction to Econometrics</u>
Winter Semester 2006 / 2007
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). **<u>All</u>** of the **<u>twelve</u>** (12) examination questions must be answered (the estimated time to spend on each question is given). This examination consists of **<u>four</u>** (4) pages and must be completed within 120 minutes.

<u>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)</u>

Question 1 (10 Minutes)

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

a. What is the total "unexplained" variation of Y given $X=x_i$?

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

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

Question 2 (10 Minutes)

You are a researcher in the Sociology Department at the Martin-Luther-Universität Halle-Wittenberg located in Halle (Saale). You are interested in the relationship between private consumption and family income in Sachsen-Anhalt. Annual data has been obtained from the responsible statistical agency from 1975 to 2005 (31 observations on consumption and income). Some of your older colleagues, however, have indicated to you that your study should take into account the German political reunification that took place in 1990.

a. Explain in detail a BLP estimation procedure that would allow the estimated level of consumption to be different before and after the structural change. Describe the **y** vector and the **X** matrix that would be used and highlight any estimation problems that might arise as a result of your specification.

b. Explain in detail a BLP estimation procedure that would allow the estimated marginal propensity to consume (mpc) to be different before and after the structural change. Describe the **y** vector and the **X** matrix that would be used and highlight any estimation problems that might arise as a result of your specification.

Question 3 (10 Minutes)

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

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

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

Please turn to Page 2

Question 4 (10 Minutes)

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

Y \ X	8.2	13.4	18.4	23.4	$f_2(y)$
1.7	0.035	0.058	0.072	0.125	
3.7	0.081	0.091	0.058	0.085	
5.7	0.092	0.069	0.068	0.166	
$f_1(x)$					

Given: $E(X) = 17.0684$, $E(X^2) = 326.0474$, $E(XY) = 66.15428$

a. What is your "best" MSE prediction for the value of Y when x = 13.4?

b. What is your BLP prediction of Y knowing that x = 13.4?

c. What is your BPP prediction of Y knowing that x = 13.4?

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

Question 5 (10 Minutes)

Consider the continuous random variable $X \sim$ Rectangular with parameters $a < b$, $f(x) = 1 / (b - a)$ for $a \le x \le b$, with $f(x) = 0$ elsewhere. Calculate the probability of each of the following events occurring for $a = -2$ and $b = 6$.

a. $A = \{-1 \le x \le 4\}$

b. $B = \{(-1 \le x \le 4) \mid x \le 4\}$

c. $C = \{x = 4\}$

Question 6 (10 Minutes)

A time series sample of 32 observations concerning the random variables $X = x_t$ and $Y = y_t$ is available, where t = 1960, 1961, …, 1991. The observed values are from a population where the random vector $(X, Y) \sim f(x, y)$ has some unknown joint population probability distribution. An OLS estimation of the BLP was conducted and a Durbin-Watson statistic calculated, $d = 0.1381$ with $d_L = 1.373$ and $d_U = 1.502$.

$$\mathbf{X^{*\prime} X^*} = \begin{bmatrix} 0.28094836 & 25.5056702 \\ 25.5056702 & 2516.79842 \end{bmatrix} \qquad (\mathbf{X^{*\prime} X^*})^{-1} = \begin{bmatrix} 44.504794 & -0.451019 \\ -0.451019 & 0.00496804 \end{bmatrix}$$

$$\mathbf{X^{*\prime} y^*} = \begin{bmatrix} 25.3744888 \\ 2469.16389 \end{bmatrix}$$

$$\Sigma_t (y_t^* - c_1 x_{t1}^* - c_2 x_{t2}^*) = 1.28504512$$
$$\Sigma_t (y_t^* - c_1 x_{t1}^* - c_2 x_{t2}^*)^2 = 27.8588238$$
$$[\Sigma_t (y_t^* - c_1 x_{t1}^* - c_2 x_{t2}^*)]^2 = 1.65134095$$

a. Calculate the GLS estimates for c_1 and c_2.

b. Calculate the Var (c_1), Var (c_2), and the Cov (c_1, c_2).

c. Explain why the Durbin-Watson statistic ranges between zero and four and what it means at each of its values.

Please turn to Page 3

Question 7 (10 Minutes)

The Runs Test (Geary Test) is a nonparametric test that can be used to check whether the univariate 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.

Question 8 (10 Minutes)

$$\underset{(n \times n)}{\text{Var-Cov}(Y \mid \mathbf{X})} = \begin{array}{c} \\ 1 \\ 2 \\ \\ t \\ \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \cdots\ t & \cdots & n \end{array} \\ \left[\begin{array}{lllll} \mathbf{V(Y|X_{1j}=x_{1j})} & C(Y|X_{1j}=x_{1j}, Y|X_{2j}=x_{2j}) & C(Y|X_{1j}=x_{1j}, Y|X_{tj}=x_{tj}) & C(Y|X_{1j}=x_{1j}, Y|X_{nj}=x_{nj}) \\ C(Y|X_{2j}=x_{2j}, Y|X_{1j}=x_{1j}) & \mathbf{V(Y|X_{2j}=x_{2j})} & C(Y|X_{2j}=x_{2j}, Y|X_{tj}=x_{tj}) & C(Y|X_{2j}=x_{2j}, Y|X_{nj}=x_{nj}) \\ & & & \\ C(Y|X_{tj}=x_{tj}, Y|X_{1j}=x_{1j}) & C(Y|X_{tj}=x_{tj}, Y|X_{2j}=x_{2j}) & \mathbf{V(Y|X_{tj}=x_{tj})} & C(Y|X_{tj}=x_{tj}, Y|X_{nj}=x_{nj}) \\ & & & \\ C(Y|X_{nj}=x_{nj}, Y|X_{1j}=x_{1j}) & C(Y|X_{nj}=x_{nj}, Y|X_{2j}=x_{2j} & C(Y|X_{nj}=x_{nj}, Y|X_{tj}=x_{tj}) & \mathbf{V(Y|X_{nj}=x_{nj})} \end{array}\right] \end{array}$$

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 $[(Y \mid X_{tj} = x_{tj}), (Y \mid X_{t-1j} = x_{t-1j})]$?

Question 9 (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 autocorrelation likely to be present?

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 Logit Model.

Please turn to Page 4

Question 10 (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:

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

$$\text{Estimated BLP} = c_1 + c_2\,x_2 + c_3\,x_3 + c_4\,x_4 + c_5 x_5$$

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 11 (10 Minutes)

A group of business management students here at the Otto-von-Guericke Universität Magdeburg can be categorized in the following manner. Twenty-eight percent (28%) of them are students from foreign countries. Concerning only those students from countries other than Germany, seventy-four percent (74%) of them live in apartments located on the campus. Sixty-eight percent (68%) of the students from Germany live in rented apartments off the campus. What is the probability that a business management student selected at random is:

a. A student from a foreign country who is living off campus.

b. A student from a foreign country who is living on the campus.

c. A German student who is living on the campus or a student from a foreign country living off campus.

d. A business management student who is living on the campus.

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

Question 12 (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**?

This is the end of the examination.

<u>GOOD LUCK !!</u>