Examination: 5027
Economics III
<u>Introduction to Econometrics</u>
Winter Semester 2005 / 06
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 (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:

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

a. Explain in detail all the steps necessary to calculate the Durbin-Watson statistic, d, using this sample data.
b. 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.
c. Explain WHY the Durbin-Watson d statistic is a number near 2.0 when null hypothesis is not rejected. Why does $0 \le d \le 4.0$?

Question 2 (10 Minutes)

Given a certain bivariate population with continuous random variables X and Y, where:

E(Y) = 3.409, V(Y) = 0.8314, E(X) = 2.076, V(X) = 0.5532, and C(X, Y) = – 0.5216.

When the choice of predictors, c, is confined to linear functions of X=x, c = h (x), then either the BLP or the BPP could be used to predict the value of Y.

a. Using the BLP, what is the best MSE prediction of Y given that X = 4.12?
b. Using the BPP, what is the best MSE prediction of Y given that X = 4.12?
c. 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.
d. Which of these two predictions, the BLP or the BPP, do you think produces a value that is the closest to the CEF? Explain your answer.

Please turn to Page 2

Question 3 (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 4 (10 Minutes)

In order to complete their work for a baccalaureate degree in the English language program of the Faculty of Economics and Management, 92% of the students complete the course *Principles of Economics I* and 38% of the students take the course *Introduction to Econometrics* while here in Magdeburg (the others either get credit for taking equivalent courses elsewhere or do not take the courses). A recent survey found that 13% of the students who take the econometrics course here in Magdeburg had not taken the course *Principles of Economics I* here in our faculty's study program.

a. What is the percentage of students who take both courses here in Magdeburg?

b. Are the decisions to take these two courses made independently by students?

c. Considering the group of students who do not take the econometrics course here in Magdeburg, what is your "best" prediction regarding the probability that they take the course *Principles of Economics I* here in Magdeburg?

Question 5 (10 Minutes)

The Linear Probability Model states that in the population:
$E(Y \mid X = x_i, W = w_i) = \beta_1 + \beta_2 x_i + \beta_3 w_i = p_i$ and $V(Y \mid X = x_i, W = w_i) = p_i(1 - p_i)$ where $p_i = Pr(Y = 1 \mid X = x_i, W = w_i)$. The dichotomous random variable Y signifies home ownership by married couples, the continuous random variable X measures total household income, and the dummy variable W is equal to one if either (or both) the man or the wife has a university degree.

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?

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

Please turn to Page 3

Question 6 (20 Minutes)

Given a sample of 47 quarterly observations $(y_t, x_{t2}, x_{t3}, x_{t4})$ for $t = 1, 2, \ldots, 47$. The following matrices were calculated [assume for all t, $V(Y \mid X_{t2}=x_{t2}, X_{t3}=x_{t3}, X_{t4}=x_{t4}) = \sigma^2_{Y|X}$]:

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

$$\Sigma_{s=2}^{47} [(y_s - c_1 - c_2 x_{s2} - c_3 x_{s3} - c_4 x_{s4})(y_{s-1} - c_1 - c_2 x_{s-12} - c_3 x_{s-13} - c_4 x_{s-14})] = -203647.55$$

$$\Sigma_{t=1}^{47} [(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4})(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4})] = 533108.78$$

$$\Sigma_{s=2}^{47} [(y_s - c_1 - c_2 x_{s2} - c_3 x_{s3} - c_4 x_{s4})(y_{s-1} - c_1 - c_2 x_{s-12} - c_3 x_{s-13} - c_4 x_{s-14})]^2 = 1.3755\text{E}10$$

$$\Sigma_{s=2}^{47} [(y_s - c_1 - c_2 x_{s2} - c_3 x_{s3} - c_4 x_{s4}) - (y_{s-1} - c_1 - c_2 x_{s-12} - c_3 x_{s-13} - c_4 x_{s-14})]^2 = 1482042.41$$

$$\Sigma_{s=2}^{47} [(y_t - c_1 - c_2 x_{t2} - c_3 x_{t3} - c_4 x_{t4}) - (y_{t-1} - c_1 - c_2 x_{t-12} - c_3 x_{t-13} - c_4 x_{t-14})] = 1217.39$$

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 $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$?

c. Calculate an estimate of the correlation coefficient by two different methods, $\hat{\rho} = \text{Corr}\,[(Y \mid X_{t2}=x_{t2}, X_{t3}=x_{t3}, X_{t4}=x_{t4}), (Y \mid X_{t-12}=x_{t-12}, X_{t-13}=x_{t-13}, X_{t-14}=x_{t-14})]$

Question 7 (10 Minutes)

Consider a cross-sectional sample of 40 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 5% critical value of 2.22 for an F with 18 df, conduct the Goldfeld-Quandt Test on the sample of 40 observations given in this question. Report your results and explain what you have learned from this test.

Please turn to Page 4

Question 8 (15 Minutes)

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

d. Draw a picture of the (n x n) Var-Cov $(Y \mid \mathbf{X})$ matrix and discuss why the Cov $[(Y \mid X = x_i), (Y \mid X = x_j)]$ when $i \neq j$ cannot be replaced with the correlation coefficients.

e. The GLS estimator can be written as: $\mathbf{c}_{GLS} = (\mathbf{X}^{*\prime}\,\mathbf{X}^{*})^{-1}\,\mathbf{X}^{*\prime}\,\mathbf{y}^{*}$. The estimator can also be written as $(\mathbf{X}'\,\mathbf{P}'\,\mathbf{P}\,\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{P}'\,\mathbf{P}\,\mathbf{y} = (\mathbf{X}'\,\mathbf{V}^{-1}\,\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{V}^{-1}\,\mathbf{y}$. Explain why these expressions are equivalent.

f. Given a sample of size n from a bivariate population where the values of the n conditional variances are known, Cov $[(Y \mid X = x_i), (Y \mid X = x_i)]$. Outline the GLS estimation procedure to be used on this sample. What are the properties of the GLS estimator, $\mathbf{C}_{GLS}$.

Question 9 (10 Minutes)

Given the 1027 observations from the University of Michigan study regarding household savings behavior. Assume that the random variable Y is used to measure the savings rate and X the household income: (X, Y) ~ Bivariate Normal. It is known that in Michigan, as a matter of good banking practice, it is impossible to borrow more than 1/3 of your current income. It is also deemed highly unlikely that a family could save more that 2/3 of their current income. Estimate the sample BLP using this data set.

a. Explain in detail the properties of the estimators obtained.

b. Explain the difference in this case between the use of OLS and GLS.

Question 10 (10 Minutes)

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

a. What is the difference between VARIATION and VARIANCE?

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

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

d. Explain why R^2 should not be used when the sample estimate of the BPP has been calculated.

This is the end of the examination.

<u>GOOD LUCK !!</u>