AFRICAN ECONOMIC RESEARCH CONSORTIUM (AERC)

Collaborative PhD Programme in Economics for Sub-Saharan Africa

COMPREHENSIVE EXAMINATIONS IN CORE AND ELECTIVE FIELDS

FEBRUARY – MARCH 2012

ECONOMETRICS

Time: 08:00 – 11:00 Hours **Date: Friday, February 24, 2012**

Instructions:

Answer a total of FOUR questions: ONE question from Section A; ONE question from Section B.
In Section C, you **MUST** answer **ONE** question from **Questions (5) and (6)**; and **ONE** question from **Questions (7) and (8)**. Statistical tables are provided.

Section (A): 15 Marks [27 minutes]

Question 1

What are the factors associated with trade tax revenue based on exports in Uganda? To answer this question, an econometrician specified the model

$$TTAXREVENUE_t = \beta_0 + \beta_1 EXRATE_t + \beta_2 OPENNESS_t + \beta_3 INFLATION_t + \beta_4 ABOLISH_t + u_t$$

t=1,2,…,66 where $TTAXREVENUE$ is the total trade tax revenue collected **(millions of Ugandan shillings);** $EXRATE$ is the exchange rate **(Ugandan shillings per US dollar)**; $OPENNESS$ **(percent of GDP)** is a variable defining the degree of openness of the Ugandan economy to international trade (i.e., $OPENNESS = \left(\frac{EXPORTS + IMPORTS}{GDP}\right) x100$);

ABOLISH is a **dummy variable** which takes the value 0 before the abolition of the tax on coffee, Uganda's main export, in 1996 and takes the value 1 from 1996 onwards; and u represents the disturbance term.

The above model is estimated using quarterly time series data for Uganda covering the period from the third quarter of 1995 to the fourth quarter of 2010. Use the reported regression results to answer the questions below:

Dependent variable: TTAXREVENUE

Independent Variable	Coefficient	Standard Error	T-ratio	p-value
CONSTANT	-415.671	27.5983	-15.06	3.20e-022
EXRATE	0.135989	0.0206435	6.587	1.18e-08
OPENNESS	13.6424	0.922725	14.78	7.81e-022
INFLATION	-0.980122	5.70591	-0.1718	0.8642
ABOLISH	37.7218	23.2892	1.620	0.1105

n=66
Mean of the dependent variable=239.9549; Standard deviation of the dependent variable=148.7702

Sum of squared residuals=119812.2; Standard error of the regression=44.31856

R-squared=0.916717; Adjusted R-squared=0.911256

Durbin Watson statistic=0.896647; rho=0.552106

Jarque-Bera statistic=0.446 (p-value=0.8002)

ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Explained/Regression	1.31881e+006 (ESS)	4	329701	167.861 (p-value=3.46e-032)
Residual/Error	119812 (RSS)	61	1964.13	
Total	1.43862e+006 (TSS)	65	22132.6	

(i) Interpret the estimated slope coefficient for *EXRATE* (i.e., $\hat{\beta}_1 = 0.135989$) in the above model. Does the positive sign of this slope coefficient agree with your intuition? **[2 marks]**

(ii) Test the null hypothesis $H_0 : \beta_1 = 0$. Use 5% level of significance. Interpret your test results **[3 marks]**

(iii) Construct a 95 percent confidence interval for β_3, the coefficient of INFLATION, and use it to test the hypothesis that inflation is a significant determinant of trade tax revenue in Uganda. **[3 marks]**

(iv) Test for significance of the overall regression in the above model. Use 5% level of significance. What is the meaning of testing for significance of the overall regression? **[3 marks]**

(v) Perform an appropriate test for normality in the above model. Use 5% level of significance. What are the consequences of non-normality with respect to the properties of the ordinary least squares (OLS) estimator and the validity of hypothesis tests? **[4 marks]**

Question 2

Good econometric practice entails having a clear understanding of the nature, causes and consequences of violations of the assumptions underlying the classical linear regression model. It is also important to know the remedies to such violations.

(a) Heteroscedasticity

(i) What does the term heteroscedasticity mean? **[1 mark]**

(ii) What are the consequences of completely ignoring heteroscedasticity and performing an OLS? **[2 marks]**

(iii) Mention one method of detecting heteroscedasticity **[1 mark]**

(iv) Briefly explain one remedy that you can adopt if you detect heteroscedasticity and you know the error variances **[1.5 marks]**

(v) Briefly explain one remedy that you can adopt if you detect heteroscedasticity and you do not know the error variances **[2 marks]**

(b) Autocorrelation

(i) What does the term autocorrelation mean? **[1 mark]**

(ii) What are the consequences of completely ignoring autocorrelation and performing an OLS? **[2 marks]**

(iii) Mention one method of detecting first order autocorrelation **[1 mark]**

(iv) Briefly explain one remedy that you can adopt if you detect first order autocorrelation and you know the value of the autocorrelation coefficient ρ . **[1.5 marks]**

(v) Briefly explain one remedy that you can adopt if you detect first order autocorrelation and you do not know the value of the autocorrelation coefficient ρ in which case ρ becomes an additional parameter to be estimated. **2 marks]**

Section (B): 25% [45 minutes]

Question 3

(a) Testing for Granger causality is routine in many contexts in economics.

(i) Briefly describe the meaning and significance of "Granger Causality". **[3 marks]**

(ii) In order to investigate the relationship between GNP (denoted by y_t) and government spending (denoted by g_t) a bivariate VAR model was estimated by the ordinary least squares method using 30 annual observations. The results were as follows:

$$y_t = 0.25 + 0.60 y_{t-1} + 0.80 g_{t-1}, \; R^2 = 0.73$$
$$\quad (1.0) \quad (5.3) \qquad (1.3)$$
$$g_t = 0.87 + 0.10 y_{t-1} + 0.40 g_{t-1}, \; R^2 = 0.90$$
$$\quad (5.0) \quad (3.0) \qquad (3.2)$$

where the t statistics are given in brackets below the parameter estimates.

Using the above results, can you decide in favor of the Wagnerian hypothesis that GNP is a Granger cause for government expenditure, or in favor of the Keynesian hypothesis that government expenditures are a Granger cause for GNP, or for neither hypothesis? Explain. **[3 marks]**

(iii) Write the VAR model estimated in part (ii) in matrix form and check for stability of the estimated VAR. **[4 marks]**

(b) What is long-run causality? How would you go about testing for long-run causality between GNP and government spending? **[3 marks]**

(c) Consider the following partial adjustment model of money demand:

$$\ln M_t{}^* = \beta_1 + \beta_2 \ln Y_t + \beta_3 \ln R_t + u_t$$

$$(\ln M_t - \ln M_{t-1}) = \lambda(\ln M_t{}^* - \ln M_{t-1})$$

where $M_t{}^*$ and M_t represents the desired and actual level of money demand, respectively; Y is a measure of real income (or output) and R is a nominal interest rate measure; λ represents the speed of adjustment $(0 \le \lambda \le 1)$. Assume that there is no autocorrelation in the original/untransformed model.

(i) Discuss the a priori expected signs of all the parameters based on economic theory. **[2 marks]**

(ii) Show that the transformed model incorporating the partial adjustment mechanism is given by the following autoregressive model:
$\ln M_t = \lambda\beta_1 + \lambda\beta_2 \ln Y_t + \lambda\beta_3 \ln R_t + (1-\lambda)\ln M_{t-1} + \lambda u_t$ **[2 marks]**

(iii) Determine the short-run and long-run elasticities of money demand with respect to income based on this transformed model. **[4 marks]**

(iv) In what ways do partial adjustment models differ from error correction models associated with cointegrated systems? **[4 marks]**

Question 4

(a) Distinguish between the standard logit model and the ordered logit model. **[3 marks]**

(b) Dummy dependent variables are commonly used in econometrics to model various decision-making situations. In labor economics, for example, dummy variables are used to model the decisions of individuals to work full-time. Are older workers more likely to work full-time? Are single people less likely to work full-time? To answer these questions, an econometrician specified the following linear probability model: $D_i = \beta_1 + \beta_2 AGE_i + \beta_3 SINGLE_i + \varepsilon_i$ i=1,2,…,1000 where D=1 if a particular individual worked full-time during the reference period and D=0 otherwise; AGE is the individual's age (in years); and SINGLE is a dummy variable that takes the value 1 if the individual is single and takes the value 0 otherwise.

(i) Show that the above linear probability model can be written equivalently as $p_i = \beta_1 + \beta_2 AGE_i + \beta_3 SINGLE_i + \varepsilon_i$ where $p_i = p(D_i = 1/\{AGE_i, SINGLE_i\})$ and discuss the limitations of a linear probability model. **[3 marks]**

(ii) With respect to the decision to work full-time, suppose that the econometrician opted to specify the logit model $p_i = \dfrac{1}{1+e^{-(\beta_1+\beta_2 AGE_i+\beta_3 SINGLE_i+\varepsilon_i)}}$, rewrite the model in terms of the logarithm of the odds-ratio and explain the meaning of the odds-ratio in the present context. **[3 marks]**

(iii) Show that for the transformed model in part (ii), $\dfrac{dp_i}{dAGE} = \beta_2 p_i(1-p_i)$ and, hence, explain how the sign of β_2 influences the probability of working full-time for a particular individual of a given AGE. **[3 marks]**

(iv) Describe any problems that might arise in the estimation of the parameters of the logit model specified in part (ii) **[3 marks]**

(v) When the logit model specified in (ii) was estimated by maximum likelihood procedure, the following results were obtained.

$$\ln\left(\frac{\hat{p}_i}{1-p_i}\right) = -1.734 + 0.012 AGE_i - 0.046 SINGLE_i$$

Standard error (0.454) (0.004) (0.077)

Number of observations=30

Based on these results, do older workers have greater probability of working full-time? Does a single worker have a lower probability of working full-time**?**

[3 marks]

(vi) Use the results reported in part (v) to determine the probability that a randomly selected 40-year old married individual will work full-time. **[3 marks]**

(vii) How would you setup the probit counterpart to the logit model in part (ii)?

[4 marks]

Section (C): 60% [108 minutes]

<u>Choose and answer ONE from questions (5) and (6)</u>

Question 5

(a) Two probability concepts that feature in the time series econometrics literature are subjective probability and transition probability. Explain the contexts in which these concepts are used in time series econometrics. **[6 marks]**

(b) It has been noted that the frequency domain and the time domain are opposite sides of the same coin in the analysis of time series:

(i) Distinguish between the frequency domain and time domain approaches to the analysis of time series. **[3 marks]**

(ii) In what sense are the frequency domain and time domain opposite sides of the same coin in the analysis of time series? **[3 marks]**

(iii) Of particular significance in the frequency domain approach to the analysis of time series is the concept of the spectral density function. What is a spectral density function? What role does it play in spectral analysis of time series? **[3 marks]**

(iv) What is an autocovariance function? Derive the autocovariance function for the MA(3) process $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$. **[3 marks]**

(v) Consider the spectral density function given by $f(w) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j e^{-iwj}$ where $\gamma_j = \gamma_{-j} = E(y_t - \mu)(y_{t-j} - \mu)$ is the autocovariance at lag j. Show that the spectral density function can be rewritten as $f(w) = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \cos(wj)\right\}$.

[3 marks]

(vi) In light of the results in parts (iv) and (v), derive the spectral density function for the MA(3) process $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$. **[3 marks]**

(c) *A*ccording to Stock and Watson, cointegration could arise in situations for which two variables y_t and x_t share common trends. Explain how this situation could arise. **[6 marks]**

Question 6

(a) The simplest and by far the most popular GARCH model is GARCH (1,1) for which the conditional variance is given by $\sigma^2{}_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}{}^2 + \beta_1 \sigma_{t-1}{}^2$.

(i) In what situations would a GARCH (1,1) model be appropriate? **[2 marks]**

(ii) Show that the unconditional variance is given by $\sigma^2 = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$ and describe any necessary restrictions on the parameters for the results to be meaningful. **[3 marks]**

(iii) For the model $y_t = \beta_0 + \beta_1 x_t + u_t$ for which u_t possesses the aforementioned GARCH (1,1) structure, setup log-likelihood function and discuss the problems with maximum likelihood estimation of the parameters of the model. **[3 marks]**

(iv) How would you go about testing for GARCH (1,1) in practice? **[3 marks]**

(v) Consider the following results of estimating an ARCH (2) model. Comment on the possible presence of ARCH effects. Use 5% level of significance. **[3 marks]**

Model 2: WLS (ARCH) estimates using the 112 observations 2000:03-2009:06
Dependent variable: Inflation
Variable used as weight: 1/sigma

	coefficient	std. error	t-ratio	p-value	
const	0.688815	0.750566	0.9177	0.3608	
Exchange_Rate	0.00286290	0.000948315	3.019	0.0032	***
Discount_Rate	0.157767	0.0320601	4.921	3.07e-06	***

alpha(0)	0.399446	0.251331	1.589	0.1149	
alpha(1)	1.17616	0.0926312	12.70	3.18e-023	***
alpha(2)	-0.267079	0.0934300	-2.859	0.0051	***

Statistics based on the weighted data:

Sum squared resid 100.2912 S.E. of regression 0.959220
R-squared 0.529357 Adjusted R-squared 0.520722

F(2, 109)	61.29906	P-value(F)	1.45e-18
Log-likelihood	-152.7375	Akaike criterion	311.4750
Schwarz criterion	319.6305	Hannan-Quinn	314.7840
rho	0.963974	Durbin-Watson	0.112569

Statistics based on the original data:

Mean dependent var	6.264189	S.D. dependent var	2.348855
Sum squared resid	486.9793	S.E. of regression	2.113694

(b) Recently, there has been a proliferation of empirical papers employing Autoregressive Distributed Lag (ARDL) models. One plausible reason for the popularity of ARDL models is the link between this model and the ideas of cointegration and error correction.

(i) Consider the following ARDL(1,1) model: $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \delta y_{t-1} + u_t$. Rewrite the above model using the lag (or backshift) operator and show that for this model, the long-run multiplier is equal to $\frac{\beta_1 + \beta_2}{(1-\delta)}$. Explain the meaning of the long-run multiplier in the context of the above ARDL model. **[3 marks]**

(ii) The ARDL model in part (i) can be rewritten in terms of the first differences of the variables as $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + (\delta - 1)\left[y_{t-1} - \frac{(\beta_1 + \beta_2)}{(1-\delta)} x_{t-1} \right] + u_t$. Explain the significance of the expressions $\left[y_{t-1} - \frac{(\beta_1 + \beta_2)}{(1-\delta)} x_{t-1} \right]$ and $(\delta - 1)$ in the transformed model. **[3 marks]**

(iii) Suppose that y_t and x_t are both I(1). Explain the meaning of I(1) and discuss its implications with respect to the status of Δy_t and Δx_t. **[3 marks]**

(iv) Suppose that x_t and y_t are both I (1) and cointegrated. Explain the meaning of cointegration and discuss the implications of cointegration with respect to the status of $\left[y_{t-1} - \frac{(\beta_1 + \beta_2)}{(1-\delta)} x_{t-1} \right]$. **[3 marks]**

(v) Explain the significance of the "delta method" in asymptotic theory. How would you use the delta method to compute the (approximate) variance of the long-run multiplier $\frac{\beta_1 + \beta_2}{(1-\delta)}$ as noted in part (i) above. **[4 marks]**

Choose and answer ONE from questions (7) and (8)

Question 7

Consider a balanced panel data model. Take the original formulated model as,

$$y_{it} = x_{it}^{'}\beta + \varepsilon_{it}$$

(i) Re-write the above model in terms of deviations from group means and in terms of group means; **[3 Marks]**

(ii) For each of the three models write down the matrix of sum of squares and cross products; **[5 Marks]**

(iii) Show that the matrix total sum of squares is the sum of 'within'and 'between'; **[8 Marks]**

(iv) Repeat question (iii) for cross-products; **[5 Marks]**

(v) Show that the estimate of β in the original model is the weighted average of the corresponding β from the "within" and "between" means. **[9 Marks]**

Question 8

(a) We would like to analyze the relationship between wealth (e.g. the number of cattle or other assets) and child labour, using household data collected in several villages in rural Africa. Suppose we have data on Y_{iv} and X_{iv} where Y_{iv} is the number of hours worked by child *i* in village *v* and X_{iv} is family wealth.

(i) Explain how you could estimate the relationship between wealth and child labour using pooled OLS, Random Effect (RE) and Fixed Effect (FE) estimators. Write down the estimation equations and the conditions needed to guarantee consistency. Compare the FE and the OLS estimators: Is either of these two estimators based on weaker assumptions than the other? **[6 Marks]**

(ii) Describe one example how conventional exogeneity (often called "contemporaneous" exogeneity) could be violated. **[3 Marks]**

(iii) Are there any variables that you would like to include as control variables? Why? Why would you not want to include the variable "Number of hours child attended school"? **[3 Marks]**

(iv) An alternative estimator would be OLS with village-dummy variables. Can you explain under which conditions on the sampling process (i.e. data collection process) the estimates of these village-dummies would be consistent? **[5 Marks]**

(v) Now suppose that you were able to re-interview the same households two years later again, such that you have data on Y_{ivt} and X_{ivt} where the subscript t refers to time. How could this additional data help you to deal with any remaining concerns about endogeneity? **[3 Marks]**

(b) Beegle, Dehejia and Gatti analyze the effects of transitory income shocks on the extent of child labour, using household panel data in rural western Tanzania collected for the period from 1991 to 1994. They estimate this relationship using the following equation:

$$Y_{ijt} = \alpha_j + \delta_t + \beta_1 X_{ijt} + \beta_2 shock_{jt} + \beta_3 (shock_{jt} \cdot assets_{jt}) + \beta_4 assets_{jt} + U_{ijt}$$

where the subscript t represents time, j represents household and i represents child in the household. Y_{ijt} is child labour hours, α_j is a household specific constant, δ_t is a time specific constant, X_{ijt} contains a set of control variables including individual, household and community characteristics, $shock_{jt}$ is a binary indicator of crop loss due to pests (insects, rodents) or calamities such as fire or theft, $assets_{jt}$ represent non-land household wealth and U_{ijt} is an error term.

The following table represents key results of the estimation:

Dependent variable:	Hours worked
Method:	FE
Shock: crop loss	** 6.08 (2.87)
Log per capita assets	* 0.47 (0.27)
Shock · Log per capita assets	* - 0.47 (0.28)
Observations	5591
R-squared	0.38

where ***, **, * indicates significance at the 1%, 5% and 10% level, respectively. Notice that the indicator shock is binary, that the mean number of hours worked is 18.13 and that the mean value of log per capita assets is 10.3.

(i) Interpret this regression output and compare the effects for a very poor (log assets of zero) and an average household. **[5 Marks]**

(ii) In the above regression, the authors have controlled for several characteristics, including mother's and father's years of schooling. Is the choice of these control variables reasonable? **[5 Marks]**