AFRICAN ECONOMIC RESEARCH CONSORTIUM

Collaborative Masters Programme in Economics for Anglophone Africa (Except Nigeria)

JOINT FACILITY FOR ELECTIVES (JFE) 2013

JUNE – SEPTEMBER

ECONOMETRICS THEORY AND PRACTICE I

First Semester: Final Examination

Duration: 3 Hours **Date: Wednesday, July 31, 2013**

INSTRUCTIONS:

1. Answer **TWO QUESTIONS** in **Section A** and **TWO QUESTIONS** in **Section B**.
2. You are required to answer a total of **FOUR (4)** questions.
3. All questions carry equal marks.
4. Present your work in a clear and orderly manner.

Section A:

Answer TWO Questions from this Section. Note that Question 2 is Compulsory

Question 1

Consider the Classical Linear Regression Model (CLRM)

$y = X\beta + \varepsilon$ **(1)**

where y is an $(N \times 1)$ vector, X is an $(N \times k)$ matrix, β is $(k \times 1)$ vector of population parameters to be estimated, and ε is an error term with the property $\varepsilon \sim N(0, \sigma^2 I)$.

(i) Explain the implication of the exogeneity assumption, $E(\varepsilon | X) = 0$, in estimating the model using the ordinary least squares (OLS) estimator. **(5 points)**

(ii) Prove the Gauss-Markov Theorem which says: "If it is assumed that $E(\varepsilon | X) = 0$ and $Var(\varepsilon | X) = \sigma^2 I$ in the CLRM, then the OLS estimator, $\hat{\beta}$, is more efficient than any other *linear*, *unbiased* estimator". **(5 points)**

(iii) Suppose, one of the regressors in X, is correlated with z (which can be considered to be an instrument), but the endogeneity test suggest that all the regressors in X are exogenous. Would you recommend using the instrumental variable estimator to obtain estimates of β? Why? **(5 points)**

Question 2 (Compulsory)

Mr. John Kariuki is carrying out a research which involves among other things estimation of demand for cloth function. The demand function is specified as follows:

$$q_t = b_1 + b_2 x_{2t} + b_2 x_{3t} + g p_t + e_t \qquad \textbf{(2)}$$

where q_t is the log of quantity, p_t is the log of price, x_{2t} is the log of income, and x_{3t} is a dummy variable representing regular demand shifts. p_t is suspected to be endogenous, and, x_{4t} and x_{5t} have been identified as instruments for p_t. Mr. Kariuki decides to estimate equation (1) using both OLS and two-stage-least-squares (TSLS), using both x_{4t} and x_{5t} as instruments (Results are presented in Table 1). Endogeneity test, and heteroskedasticity tests are carried out, and the results are also reported below. Furthermore, Mr. Kariuki estimates the reduced form for p_t, and the results are reported in Table 2. **(Note**: *All continuous variables are stationary*)

Table 1 Regression results

	OLS		TSLS	
p_t	-0.863***	(0.067)	-1.451***	(0.114)
x_{2t}	0.951***	(0.067)	1.308***	(0.097)
x_{3t}	-22.38***	(4.270)	-27.38***	(5.455)
Intercept	83.16***	(6.774)	103.4***	(8.979)
N	120		120	
R^2	0.676		0.460	
Standard errors in parentheses				
"* p<0.05 ** p<0.01 *** p<0.001"				

Test of endogeneity of p - H0: Regressor is exogenous

Durbin-Wu-Hausman Chi-square test	99.515
p-value	0.0000

Pagan-Hall general test statistic for heteroskedasticity: 18.508 Chi-sq(4) P-value = 0.0432
(Ho: Disturbance is homoscedastic)

Table 2 Results for reduced form equation for p

	Reduced form regression	
x_{2t}	0.565***	(0.050)
x_{3t}	-12.43**	(3.956)
x_{4t}	-0.436***	(0.062)
x_{5t}	-0.213***	(0.063)
Intercept	81.14***	(7.217)
N	120	
R^2	0.723	
Standard errors in parentheses		
"* p<0.05		
H0: $x_{4t} = x_{5t} = 0$		
F-test = 71.81, p-value = 0.0000		

(i) Given the two methods of estimation:- OLS and TSLS, which method would you recommend? Why? **(3 points)**

(ii) Would you recommend using only one instrument: either x_{4t} or x_{5t}, and thus use a simple instrumental variable estimator rather than TSLS? Why? **(4 points)**

(iii) Are there indications that Kariuki should worry about weak instruments? Why should one worry about weak instruments? **(4 points)**

(iv) Would you recommend using GMM? Why? **(4 points)**

Question 3:

Consider the linear regression model

$$y = X\beta + \epsilon$$

where each component of X is viewed as being an *exogenous regressor* if it is uncorrelated with the error term, ϵ, or *an endogenous regressor* if it is correlated with the error term. Suppose, there are components of X which are endogenous, and that there is a vector of instruments Z that satisfies $E[\epsilon_i | z_i] = 0$; and also suppose that the number of instruments, l, exceeds the number of regressor, q. That is the model is over-identified, $(l > p)$.

(i) Explain why the generalized method of moments (GMM) estimator would be the appropriate estimator in this case. **(6 points)**

(ii) If we let W_n with the property $plim\ W_N = W$, to be an $l \times l$ full-rank symmetric weighting matrix, the GMM estimator in the linear instrumental variable model is given by $\hat{\beta}_{GMM} = [X'ZW_NZ'X]^{-1}X'ZW_NZ'y$. Show that $plim\hat{\beta}_{GMM} = \beta$. **(6 points)**

(iii) To what extent is the claim that *"the method of moments, ordinary least squares, and instrumental variable estimators are all special cases of the GMM estimator"* true? **(3 points)**

Section B:

Answer TWO Questions from this Section. Note that Question 5 is Compulsory

Question 4

Given the following AR(1) stochastic process

$$y_t = c + by_{t-1} + e_t \qquad \textbf{(1)}$$

where e_t is a white noise process, $e_t \sim N(0, s^2)$.

(i) Derive the moving average representation of the process (assuming initial value of y_t, y_{-1} is equal to zero). **(5 points)**

(ii) Derive the mean, variance and autocovariances of the process if $|b|<1$ Comment on your results? **(4 points)**

(iii) Suppose $c=0$ **and** $b=1$ in equation 1, what specific kind of a process is equation 1? **(2 points)**

(iv) Given the assumption in (iii) derive the mean, variance and autocovariances of the process. Comment on your results. **(4 points)**

Question 5 (Compulsory)

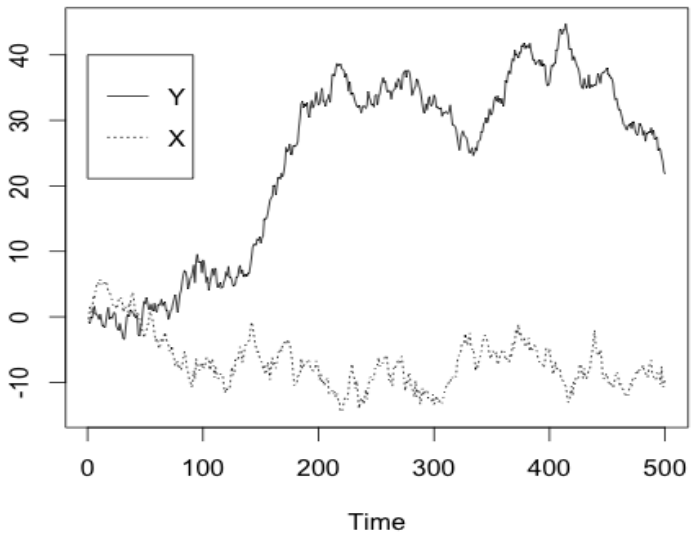
Table B.1 below presents regression results for the following model. The two time series variables are plotted in Figure B.1

$$Y_t = a + bX_t + e_t$$

Table B.1: Regression of Y_t on X_t: OLS results

	Estimate	Std. error	T-statistic
Intercept	10.19	1.01	10.06
X_t	-2.01	0.13	-15.81
T=500	$\bar{R}^2 = 0.65$	DW=0.21	

Figure B.1: X and Y



(i) Are there indications that the two variables are cointegrated? Explain. **(5 points)**

(ii) What is the meaning of *cointegration*? **(2.5 points)**

(iii) Explain what do we mean by '*spurious regressions'* **(2.5 points)**

(iv) Explain the course of action that, if you were given the data, you would take to examine the relationship between the two variables? **(5 points)**

Question 6

(a) Briefly, explain the Sims' (1980) modeling philosophy. **(3 points)**

(b) Given a simple vector autoregressive model (where x_t and y_t are stationary):

$$y_t = \alpha_{10} + \beta_{11} y_{t-1} + \text{---} + \beta_{1p} y_{t-p} + \gamma_{11} x_{t-1} + \text{---} + \gamma_{1p} x_{t-p} + u_{1t}$$

$$x_t = \alpha_{20} + \beta_{21} y_{t-1} + \text{---} + \beta_{2p} y_{t-p} + \gamma_{21} x_{t-1} + \text{---} + \gamma_{2p} x_{t-p} + u_{2r}$$

(i) Which factors would you take into account in determining the ('optimal') number of lags to include in the model? Which criteria would you use to determine that number? **(3 points)**

(ii) Which method would you use to estimate the model? Why? **(2 points)**

(iii) Explain how you would use the Granger causality test to examine the relationship between the two variables. **(4 points)**

(iv) How useful/helpful is the variance decomposition in uncovering interrelationships among variables in the system? What is (are) the main limitation(s) of the variance decomposition? **(3 points)**