AFRICAN ECONOMIC RESEARCH CONSORTIUM

Collaborative PhD Programme in Economics for Sub-Saharan Africa

COMPREHENSIVE EXAMINATIONS IN CORE AND ELECTIVE FIELDS

FEBRUARY 11 – MARCH 2, 2015

ECONOMETRICS

Time: 08:00 – 11:00 GMT **Date: Friday, February 20, 2015**

INSTRUCTIONS:

Answer a total of FOUR questions: ONE question from Section A, ONE question from Section B, and TWO questions from Section C, one of which must be **Question 5 or 6** and the other must be **Question 7 or 8**.

The sections are weighted as indicated on the paper.

SECTION A: (15%)

Answer only ONE Question from this Section

Question 1

Consider the following polynomial earnings regression for an unspecified Sub-Saharan African country:

$$y_i = \beta_0 + \beta_1 e_i + \beta_2 e_i^2 + \beta_3 s_i + \beta_4 m_i + u_i \qquad i = 1,2,\ldots,100$$

where y_i is the average monthly earnings (in shillings) of the i-th worker; e_i is the number of years of experience of the i-th worker; e_i^2 is the square of experience; s_i is the number of completed years of schooling; m_i is a dummy variable ($m_i = 1$ if the i-th worker is male and $m_i = 0$ otherwise); and u_i denotes the usual error term which is assumed to satisfy all the classical assumptions underlying a linear regression model.

(i) Explain the reasons for including the error term u_i in the above regression equation. **(2 marks)**

(ii) State the assumptions underlying u_i, assuming that u_i satisfies all the classical assumptions underlying linear regression. **(3 marks)**

(iii) What is first order autocorrelation? Should first order autocorrelation be a concern in this model? **(3 marks)**

(iv) What is the justification for including both experience (e_i) and experience squared (e_i^2) in the regression? **(3 marks)**

(v) State the Gauss Markov Theorem. Will the Gauss Markov Theorem hold with respect to this model specification? **(4 marks)**

Question 2

(a) The output of a multiple regression from a cross-section sample obtained using E-Views is reported below

Dependent Variable: Y
Method: Least Squares
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic
Constant	-------------	32.89165	5.117003
X1	-------------	0.248017	-7.128663
X2	-------------	0.001878	-2.934275
X3	-------------	4.190533	3.070883

R-squared	0.747372
Adjusted R-squared	0.734740
S.E. of regression	39.13127
Sum squared resid	91875.38

Answer the following questions on the basis of these results:

(i) Obtain the missing coefficients of the estimated regression model and determine the variables which are significant at 1% level **(3 marks)**

(ii) Approximate and interpret the 95% confidence interval for the coefficient of X1 variable **(3 marks)**

(iii) Calculate the F-statistic and comment on the overall significance of the regression model **(3 marks)**

(iv) Determine the variance of the error term (**2 marks**)

(b) An econometrics student divided her data into two sub-samples with $n_1 = 5$ and $n_2 = 10$ observations. She regressed Y on a constant and one regressor and obtained a combined residual sum of squares (RSS) of 3.1602 for the separate sub-samples. She also obtained RSS of 6.5561 for the pooled sample. Explain how you would use this information to test for structural stability using the Chow test. (**4 marks**)

SECTION B: (25%)

Answer only ONE Question from this Section

Question 3

(a) Consider the following two-variable unrestricted VAR system:

$$y_{1t} = c_1 + \pi_{11}y_{1t-1} + \pi_{12}y_{1t-2} + \pi_{13}y_{1t-3} + \pi_{14}y_{2t-1} + \pi_{15}y_{2t-2} + \pi_{16}y_{2t-3} + \varepsilon_{1t}$$

$$y_{2t} = c_2 + \pi_{21}y_{1t-1} + \pi_{22}y_{1t-2} + \pi_{23}y_{1t-3} + \pi_{24}y_{2t-1} + \pi_{25}y_{2t-2} + \pi_{26}y_{2t-3} + \varepsilon_{2t}$$

(i) Rewrite the above unrestricted VAR in matrix form and state the advantages of the unrestricted VAR. **(5 marks)**

(ii) Define Granger causality and explain how the unrestricted VAR stated above could be used to test the null hypothesis of no unidirectional causation from y_2 to y_1. **(5 marks)**

(b) Consider the following results of performing Johansen's test for cointegration on a VAR system with 3 variables (inflation, interest rates and exchange rates). These results are obtained using real data for a certain Sub-Saharan African country.

Rank	Eigenvalue	Trace test (p-value)	Maximum eigenvalue test (p-value)
0	0.23449	35.168 (0.0101)	27.256 (0.0046)
1	0.066593	7.9116 (0.4821)	7.0292(0.4942)
2	0.0086135	0.88238 (0.3476)	0.88238 (0.3476)

(i) Interpret the results indicating the number of cointegrating vectors. **(5 marks)**

(ii) Use these reported results to comment on the stability of the VAR system. **(5 marks)**

(iii) Do the rank test and maximum eigenvalue test yield conflicting conclusions at 5% level of significance? **(5 marks)**

Question 4

(a) Consider the model $y = X\beta + \varepsilon$ where y is an *nx1* vector of observations on the dependent variable; X is an *nxk* matrix of observations on the independent variable; β is a *kx1* vector of parameters to be estimated and ε is an *nx1* vector of errors such that $E(\varepsilon\varepsilon') = \sigma^2 I$. Assume all the other assumptions of the classical linear regression model are met.

(i) Setup the log-likelihood function and derive the maximum likelihood estimators of β and s^2. **(5 marks)**

(ii) Comment on the properties of the maximum likelihood estimators of β and s^2. **(5 marks)**

(b) Compare and contrast the theoretical concepts behind the estimation of linear regression models with normally distributed disturbances through ordinary least squares (OLS) and maximum likelihood (ML). What are the differences in the estimation results from the two approaches assuming that the disturbances are normally distributed? **(5 marks)**

(c) Consider the following OLS results from fitting a Cobb-Douglas production function relating GDP (Y) (measured in billion dollars) to capital input (K) and labour input (L) using data for the manufacturing sector in 20 U.S. states:

$$\ln(\hat{Y}_i) = -1.6524 + 0.846 \ln K_i + 0.3397 \ln L_i$$

$t-ratio$ (-2.7259) (9.0625) (1.8295)

Total sum of squares = 2.76525462

Explained sum of squares = 2.75165006

(i) Interpret the coefficient estimates of the model and determine their statistical significance. **(5 marks)**

(ii) What do these results infer with respect to economies of scale? What should be the appropriate method for testing for constant returns to scale in this model? **(5 marks)**

SECTION C (60%)

Answer TWO Questions from this Section,

ONE of which must be Question 5 or 6

And THE OTHER must be Question 7 or 8

Question 5

(a) One of the most popular ARCH models for modelling returns on financial assets is the ARCH in Mean (ARCH-M) model. Suppose that the objective is to model the returns of a particular financial asset (RFA) as a function of GDP using the ARCH(p)-M model

(i) Fully specify an ARCH(p)-M model relating RFA and GDP **(2 marks)**

(ii) Why is it necessary to impose certain parameter restrictions with respect to the above model? **(2 marks)**

(iii) Explain why the ARCH-M model is the preferred specification for modelling the returns on financial assets **(4 marks)**

(b) Asymmetric response is an issue that commonly receives special consideration in ARCH/GARCH modelling:

(i) Explain the meaning of asymmetric response in the context of ARCH/GARCH models. **(5 marks)**

(ii) If the objective was to capture asymmetric responses in the relationships briefly explain one ARCH/GARCH specification you would consider to be appropriate. **(5 marks)**

(c) Consider the following three alternative ARCH/GARCH specifications relating two variables y and x:

Model 1: $y_t = x_t'\beta + \varepsilon_t$; $\varepsilon_t = u_t\sigma_t$; and $s_t^2 = a_0 + a_1 e_{t-1}^2 + a_2 e_{t-2}^2 + \ldots + a_p e_{t-p}^2$ where u_t is $i.i.d.(0,1)$

Model 2: $y_t = x_t'\beta + \delta\sigma_t^2 + \varepsilon_t$; $\varepsilon_t = u_t\sigma_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2$ where u_t is $i.i.d.(0,1)$

Model 3: $y_t = x_t'\beta + \varepsilon_t$; $\varepsilon_t = u_t\sigma_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \lambda_1 d_{t-1}\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$ where u_t is $i.i.d.(0,1)$ and d_{t-1} is a dummy variable that takes the value 1 if $\varepsilon_{t-1} < 0$ and is equal to 0 if $\varepsilon_{t-1} \geq 0$.

(i) Explain the empirical justification for specifying Model 2 (**2 marks**)

(ii) Explain the empirical justification for specifying Model 3 (**2 marks**)

(iii) State the parameter restrictions with respect to Model 1 and indicate the reasons for these parameter restrictions (**2 marks**)

(iv) Explain the meaning of a likelihood function and setup the log-likelihood function for maximum likelihood estimation of the parameters of Model 1. (**4 marks)**

(v) What are the problems arising from maximum likelihood estimation of the parameters of Model 1? (**2 marks**)

Question 6

Moving average models are often considered in the analysis of stationary time series. Answer the questions below pertaining to the MA(3) model: $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$.

(i) Explain the meaning of stationarity and invertibility and state the necessary stationarity or invertibility condition, whichever is appropriate for the MA(3) model. **(5 marks)**

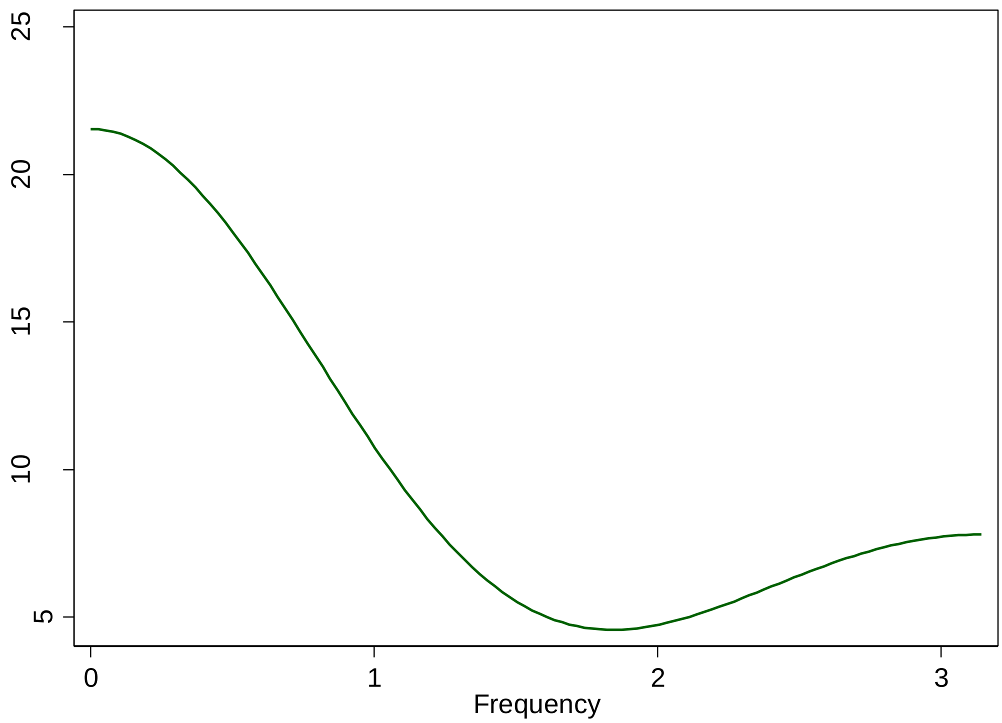
(ii) Explain the meaning of a state space representation and setup a state-space presentation for MA(3). **(5 marks)**

(iii) One strategy for estimating the MA(3) model is by conditional maximum likelihood. Explain how you would setup the conditional maximum likelihood of the parameters of MA(3). **(5 marks)**

(iv) Explain the meaning of a conditional forecast and derive the 2-step ahead conditional forecast for MA(3) as well as the variance of the forecast error. **(5 marks)**

(v) Derive the spectral density function for the above MA(3) model. What information does the spectral density function provide? **(5 marks)**

(vi) Interpret the following estimated spectral density from fitting an MA(3) model to the first difference of the price, in Kenyan Shillings, of Kenya's Bamburi Cement stocks using real monthly data covering the period from January 2001 to December 2010. Identify the full ARMA specification that takes into account the differencing and explain the reason for fitting the MA(3) model to the first difference of the stock prices. **(5 marks)**



Question 7

(a) What inferential problems are encountered in the interpretation of the results of a pooled panel data model estimated by OLS? What precautionary measures should be taken to safeguard against it? **(5 marks)**

(b) Explain the assumptions and shortcomings behind the formulation of fixed effects models **(5 marks)**

(c) Consider the following panel data model

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it},\ i = 1,2,...,10; t = 1,2,...,20$$

(i) Given that the estimated R^2 of the pooled model is 0.812408 and that of the fixed effects model is 0.944073. Obtain the F-statistic for testing the significance of the ten individual unit effects. Comment on your results. **(4 marks)**

(ii) Test for separate regressions for each of the ten units (Chow test) given that the residual sums of squares for the pooled model is 1755850 and the combined individual residual sum of squares for all the units is 324728.6. **(4 marks)**

(d) Answer the following questions on random effects models:

(i) What are the assumptions underlying the panel data random effects model? What approaches are used to estimate such a model? **(4 marks)**

(ii) Explain the estimation of random effects models when the number of cross-section units is less than the number of regressors (constant included) in the pooled model. **(2 marks)**

(iii) Which test and distribution is commonly used in testing for random effects? **(2 marks)**

(iv) A graduate student estimated both fixed effects and random effects models using a balanced panel in order to determine the preferred model. He estimated eleven coefficients in both models the resulting value of Hausman's test statistic was **8.52**. Was the student right in choosing the fixed effects model? **(4 marks)**

Question 8

(a) What problems make the linear limited dependent variable model (LPM) to be unattractive for analyzing limited dependent variables? Can these problems be overcome**?** **(5 marks)**

(b) Answer the following questions:

(i) Specify a logit model with k regressors expressed in terms of the logarithm of the the odds ratio and use it to derive the corresponding cdf. **(2 marks)**

(ii) What methods and distributions are employed in estimating the logit, probit and linear probability models? **(4 marks)**

(iii) Write the formulas for determining the effect of a unit change in the $k-th$ regressor on the probability that the dependent variable Y=1 for the logit, probit and linear probability models **(3 marks)**

(c) Using mortgage application data from a particular bank in 2013, a researcher wanted to find out whether gender was a significant factor in denying an applicant a bank loan. The dependent variable was Y=1 if the mortgage application was denied and zero otherwise. The explanatory variables included the loan-salary ratio, X, and a dummy variable representing gender (Female=1; and 0 otherwise). Fitting a probit model to 2,380 observations, the researcher obtained the following results (standard errors in brackets):

$$\Pr(Y = 1/X, Female) = -2.26 + 2.74X + 0.71 Female$$
$$(0.16) \quad (0.44) \quad (0.083)$$

(i) Comment on the sign and significance of the coefficient on the female dummy variable **(2 marks)**

(ii) What is the approximate difference in denial probabilities between a female and a male when both have the same value of X at 0.3? **(3 marks)**

(iii) What do your results in (i) and (ii) suggest about gender discrimination in loan acquisition for this particular bank? **(4 marks)**

(d) Answer the following questions:

(i) What are the implications of using OLS for estimating truncated and censored regression (Tobit) models? **(2 marks)**

(ii) Formulate a Tobit model with censoring from below at zero where the latent variable y^* is linear in regressors with homoscedastic normally distributed errors. What is the distribution of the latent variable y^* implied in this model? **(5 marks)**