**AFRICAN ECONOMIC RESEARCH CONSORTIUM**

***Collaborative PhD Programme in Economics for Sub-Saharan Africa***

***COMPREHENSIVE EXAMINATIONS IN CORE AND ELECTIVE FIELDS***

**FEBRUARY 13 – MARCH 4, 2019**

# ECONOMETRICS

**Time: 08:00 – 11:00 GMT** **Date: Friday, February 22, 2019**

## INSTRUCTIONS:

1. 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 either Question 5 or Question 6**.
2. The sections are weighted as indicated on the paper.
3. The null hypothesis and the alternative hypothesis for all the statistical tests in this examination should be indicated.

## SECTION A: (15%)

***Answer only ONE Question from this Section***

### Question 1: [15 Marks]

Consider the linear regression model (1.1).

$$y_t = \alpha_1 z_{1t} + \alpha_2 z_{2t} + v_t \qquad (1.1)$$

where $v_t \sim N(0, \sigma^2)$. The observations on $y$ and Z, where Z is the matrix of explanatory variables, are given below:

| $y_t$ | $z_{1t}$ | $z_{2t}$ |
|---|---|---|
| 4 | 1 | 0 |
| 2 | 1 | 2 |
| 3 | 1 | 2 |
| 2 | 1 | 3 |
| 1 | 1 | 4 |

(a) Find Z'Z and Z'$y$ **[3 marks]**

(b) Find $(Z'Z)^{-1}$ **[3 marks]**

(c) Find the ordinary least squares estimates for $\alpha_1$ and $\alpha_2$ **[3 marks]**

(d) Given that the estimate of $\sigma^2 = 0.1818$, estimate the variances of $\alpha_1$ and $\alpha_2$. **[3 marks]**

(e) Test the null hypothesis that $\alpha_2 = 0$ against the alternative that $\alpha_2 \neq 0$. Use 5% level of significance **[3 marks]**

## Question 2: [15 Marks]

An econometrician used a random sample of 1000 observations derived from a household survey to estimate the following consumption function

$$Y_i = \beta_0 + \beta_1 X_i + u_i \text{ , } i=1,2,....,1000 \qquad (2.1)$$

where $Y$ denotes consumption; $X$ denotes disposable income; and $u$ is an error term with the property $u_i \sim N(0, \sigma^2)$.

The regression results are presented in Table 2.1 below. Figure 2.1 presents the graph of residuals vs fitted values.

**Table 2.1 Regression results**

| | *Dependent variable:* |
|---|---|
| | Y |
| X | 0.601 |
| | (0.100) |
| Constant | 0.258 |
| | (0.058) |
| Observations | 1,000 |
| $R^2$ | 0.35 |
| Adjusted $R^2$ | 0.34 |
| Residual Std. Error | 1.830 (df = 998) |
| F-Statistic | 35.823 (df1 = 1 ; df2 = 998) |

*Note: The numbers in parentheses below the estimated intercept and slope coefficients are the standard errors*

Breusch-Pagan test: BP = 175.37, p-value = 0.00021
Breusch-Godfrey test: LM test = 0.71547, df = 2, p-value = 0.6993

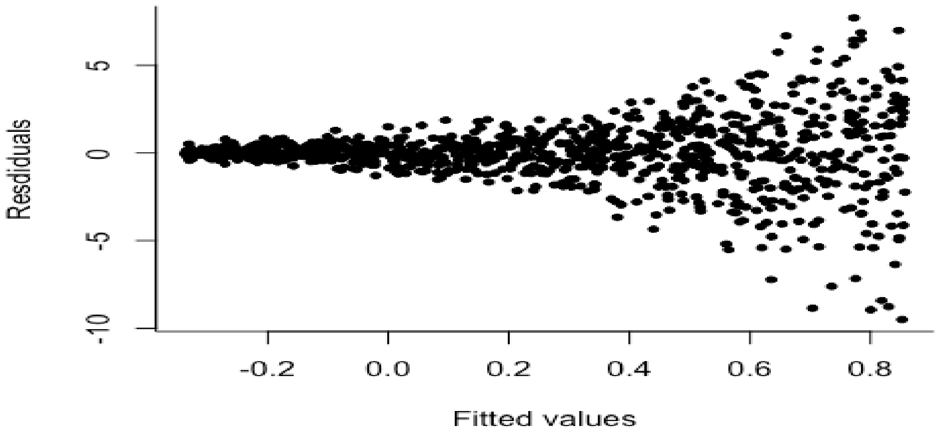


*Figure 2.1 Graph of Residuals vs fitted values*

(a) Indicate the economic meaning of the estimate for $\beta_1$ **[3 marks]**

(b) Is $\beta_1$ significant at 5% level of significance? **[3 marks]**

(c) From Figure 2.1 and the regression results reported in Table 2.1 what kind of econometric problem(s) characterize(s) the model? Briefly explain. **[3 marks]**

(d) Is serial correlation in the error term a concern in this case? Briefly explain. **[3 marks]**

(e) From the reported regression results are there indications that the estimates of the parameters of the model are likely to be biased? Briefly explain. **[3 marks]**

## SECTION B: (25%)

***Answer only ONE Question from this Section***

## Question 3: [25 Marks]

Greene (2012, pp. 694-696) presented the results of a study on whether a new method of teaching called the *Personalized System of Instruction* (PSI), significantly influenced performance in later economics courses.

The dependent variable used in the analysis is *GRADE*, (*GRADE* = 1 when the student's grade in an intermediate macroeconomics course was higher than that in the principles course; *GRADE* = 0 otherwise).

The independent variables are the student's grade point average (*GPA)*, the score on a pretest that indicates entering knowledge of the material (*TUCE)*, and the binary variable *PS*I (*PSI* = 1 *when* the student was exposed to the new teaching method; *PSI* = 0 otherwise).

The logit and probit estimation results are presented below:

| *Variable* | *Logistic* | | | | *Probit* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Coeffs* | *p-value* | *Marginal Effect* | *p-value* | *Coeffs* | *p-value* | *Marginal Effect* | *p-value* |
| constant | -13.021 | 0.008 | -- | -- | -7.452 | 0.003 | -- | -- |
| GPA | 2.826 | 0.025 | 0.534 | 0.024 | 1.626 | 0.019 | 0.533 | 0.022 |
| TUCE | 0.095 | 0.501 | 0.018 | 0.493 | 0.052 | 0.537 | 0.017 | 0.531 |
| PSI | 2.379 | 0.025 | 0.456 | 0.012 | 1.426 | 0.017 | 0.464 | 0.006 |

(a) Describe the differences and the similarities between the logit and probit results. **[4 marks]**

(b) Specify the binary probit model and its assumptions used in the estimation. **[4 marks]**

(c) Describe the difference in the assumptions underlying the binary probit and logit models**.** **[4 marks]**

(d) Using the probit model, derive the expression for the probability that *GRADE* = 1. **[5 marks]**

(e) Using the probit model, derive the likelihood function, log-likelihood function and the first order conditions with respect to the parameters of interest. **[8 marks]**

## Question 4: [25 Marks]

(a) Consider the following random walk with drift process:

$$y_t = y_{t-1} + \beta + \varepsilon_t.$$

where $\beta$ is a constant drift parameter and $\varepsilon_t$ is a white noise process.

(i) State the assumptions underlying $\varepsilon_t$ . **[2 marks]**

(ii) Give one example of a macroeconomic/financial variable which behaves like a *simple random walk*. **[1 mark]**

(iii) Show that while the random walk with drift process is non-stationary, its first difference is covariance stationary. **[6 marks]**

(b) As an economics PhD student, you have been asked to examine the relationship between real GDP, $y_t$, and money supply, $m_t$ in your country. Unit root tests you have conducted indicate that the two variables are non-stationary, and, more specifically, $y_t \sim I(1)$, and $m_t \sim I(1)$.

(i) Explain how you would use the Engle-Granger procedure to test for cointegration. **[3 marks]**

(ii) Outline the major weaknesses of the Engle – Granger test for cointegration. **[2 marks]**

Explain the three major steps involved in Pesaran's ARDL bounds tests for cointegration between four variables: $y_t, x_{1t}, x_{2t}$, and $x_{3t}$ **[4 marks]**

(c) 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:

$$y_t = 0.25 + 0.60 y_{t-1} + 0.80 g_{t-1}, \; R^2 = 0.73$$

(1.0) (5.3) (1.3)

$$g_t = 0.87 + 0.10 y_{t-1} + 0.40 g_{t-1}, \; R^2 = 0.90$$

(5.0) (3.0) (3.2)

where $R^2$ is the coefficient of determination and 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. **[7 marks]**

## SECTION C: (60%)

***Answer TWO Questions from this Section, ONE of which MUST be Question 5 OR 6***

### Question 5: [30 Marks]

(a) A common assumption in count data models is that, for given $x_i$, the count variable $y_i$ has a Poisson distribution with expectation $E[y_i|x_i] = \lambda_i = \exp\{x_i'\beta\}$ given by

$$f(y_i|x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad y_i = 0, 1, 2, \dots$$

(i) State some characteristics of the dependent variable in a count data model. **[3 marks]**

(ii) What are the major limitations of the Poisson model? **[3 marks]**

(iii) How would you avoid the limitations of the Poisson model identified in (ii) above? Briefly explain **[2 marks]**

(b) Hurdle models allow for a systematic difference in the statistical process governing individuals. The probability function of the hurdle-at-zero is given by:

$f(y = 0) = g_1(0)$, probability of zero outcome

$f(y = k) = (1 - g_1(0))g_2(k)$, probability function for positive integers, k = 1, 2, …

Let $f_1$ and $f_2$ be the pdf for non-negative integers such that:

$$g_1(0) = f_1(0) \qquad \text{and} \qquad g_2(k) = \frac{f_2(k)}{1-f_2(0)}$$

(c) Derive the likelihood function for an independent sample of *N* observations assuming that $f_1$ and $f_2$ are Poisson distributions with parameters $\lambda_1 = \exp(x_i'\beta_1)$ and $\lambda_2 = \exp(x_i'\beta_2)$, respectively. **[10 marks]**

(d) Suppose that absolute convergence holds for a group of countries, $i = 1, 2, \dots N$, $t = 1, 2, \dots, T$. The conventional growth model, which uses panel data, is given as:

$$\boldsymbol{y_{it} = \alpha_i + \delta\, y_{i,t-1} + v_{it}} \qquad (5.1)$$

where $y_{it}$ is the log of income of the $i$th country in period $t$; $\alpha_i$ is the individual-specific effects; $\delta = (1 - b)$, $0 < b < 1$, implies absolute income convergence; and $v_{it} \sim iid\,(0, \sigma_v^2)$ is the disturbance term.

(i) Differentiate between static and dynamic panel data models. **[3 marks]**

(ii) Differentiate between the Anderson-Hsiao estimators and the Arellano & Bond estimators. **[3 marks]**

(iii) If you are to use GMM to estimate Model (5.1) using the full sample of *T* observations, what are the valid instruments that may be used in the estimation. Enumerate all moment conditions and show the matrix of instruments, $Z_i$. **[6 marks]**

## Question 6: [30 marks]

(a) A Tobit model is used to estimate the relationship between credit card expenditures and some observed characteristics of an individual such as income, age and number of derogatory reports. Observations on 100 surveyed individuals were used to estimate the model based on the Excel data file "credit card expenditures". The data summary statistics and estimation results obtained using STATA econometric software package are shown below.

```
. import excel "C:\Users\Documents\Credit Card Expenditure.xls", sheet("Sheet1") firstrow
case(lower)
. sum avgexp income age mdr
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      avgexp |        100    189.0231    294.2446          0    1898.03
      income |        100      3.3693    1.629013        1.5         10
         age |        100       32.08    7.828567         20         55
         mdr |        100         .36     1.01025          0          7

. tobit avgexp income age mdr, ll
Tobit regression                                  Number of obs   =        100
                                                  LR chi2(3)      =      27.85
                                                  Prob > chi2     =     0.0000
Log likelihood = -537.20541                       Pseudo R2       =     0.0253
------------------------------------------------------------------------------
      avgexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   91.63029   22.10062     4.15   0.000     47.76667    135.4939
         age |  -4.310459   4.777255    -0.90   0.369    -13.79199    5.171069
         mdr |  -218.9261   71.06661    -3.08   0.003    -359.9737   -77.87858
       _cons |   4.012551   150.8896     0.03   0.979    -295.4616    303.4867
-------------+----------------------------------------------------------------
      /sigma |   324.8963   28.00537                      269.3134    380.4792
------------------------------------------------------------------------------
  Obs. summary:         28  left-censored observations at avgexp<=0
                        72     uncensored observations
                         0 right-censored observations

. regress avgexp income age mdr
      Source |       SS       df       MS              Number of obs =     100

-------------+------------------------------           F(  3,    96) =    7.23

       Model |  1579949.55     3  526649.848           Prob > F      =  0.0002

    Residual |  6991458.92    96   72827.697           R-squared     =  0.1843
-------------+------------------------------           Adj R-squared =  0.1588
       Total |  8571408.46    99  86579.8834           Root MSE      =  269.87
------------------------------------------------------------------------------
      avgexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   72.63788   17.28555     4.20   0.000     38.32634    106.9494
         age |  -1.523183   3.606736    -0.42   0.674    -8.682497    5.636131
         mdr |  -53.12481   26.93669    -1.97   0.051    -106.5937    .3440971
       _cons |   12.27294   117.5757     0.10   0.917     -221.113    245.6589
------------------------------------------------------------------------------
```

(i) Explain why a censored regression model may be used to estimate the above relationship. **[2 marks]**

(ii) Specify the censored regression model and state its assumptions. **[2 marks]**

(iii) What are the major weaknesses of the censored regression model? **[2 marks]**

(iv) Derive, based on the censored regression model you specified in (ii), the probability that the outcome $y$ is zero given $x$, and the conditional expectation of expenditure variable given a positive outcome, $E\{y_i|y_i > 0\}$. **[5 marks]**

(v) Show, based on parts (ii) and (iv), that $E\{y_i|x_i\} = \Phi(\cdot)\, x_i'\beta + \sigma\, \phi(\cdot)$ where $\Phi(\cdot) = \Phi(x_i'\beta/\sigma)$ and $\phi(\cdot) = \phi(x_i'\beta/\sigma)$, are respectively the CDF and pdf of the standard normal distribution. **[5 marks]**

(vi) Compare the reported censored regression results with the OLS results **[4 marks]**

(b) Answer the following questions pertaining to treatment evaluation in microeconometrics.

(i) Explain the concept of the treatment evaluation in microeconometrics. **[4 marks]**

(ii) Define counterfactual, average treatment effects on the treated (ATET), and average selection bias in relation to treatment evaluation. **[6 marks]**

## Question 7: [30 Marks]

Suppose you were asked to estimate the following consumption function for which ***consumption*** is a function of ***permanent income***.

$$C_t = \alpha + \beta_1 I_t + \beta_2 Y_t^* + u_t \quad \text{.....................................(7.1)}$$

where $C_t$ is private consumption; $I_t$ is the inflation rate; $Y_t^*$ is permanent income; $\alpha$, $\beta_1$ and $\beta_2$ are parameters to be estimated; and $u_t$ is an error term with the property $N(0,\sigma^2)$.

For estimation purposes, there is data for the consumption variable, $C_t$. However, there is no data for permanent income, $Y_t^*$, but there is data for ***actual income***, $Y_t$. As a result, the model you will end estimating is

$$C_t = \alpha + \beta_1 I_t + \beta_2 Y_t + v_t \qquad (7.2)$$

where, $Y_t$ is actual income, and $v_t$ is an error term that follows a normal distribution with zero mean and constant variance. You are told that the rate of inflation, $I_t$, in the model is exogenous.

(a) What are the econometric implications of estimating by OLS Equation (7.2) instead of Equation (7.1)? *Use algebraic illustrations to explain the implications*. **[4 marks]**

(b) Suppose that the data set also includes data for other three variables: $Z_{1t}$, $Z_{2t}$, and $Z_{3t}$, which are correlated with the actual income, $Y_t$ but are not correlated with the error term, $v_t$. Furthermore, unit root test results indicate that all variables in the data set are covariance stationary. You are asked to estimate Equation (7.2) using the two stage least squares (2SLS) estimator.

(i) What is the econometric meaning of the inflation rate, $I_t$, in the model being exogenous? **[1 mark]**

(ii) What would be the main reason(s) for using the 2SLS estimator? Explain. **[2 marks]**

(iii) Explain how you will go about carrying out estimation in the 1st stage of 2SLS. **[1 mark]**

(iv) Will the OLS estimator in the 1st stage be unbiased and consistent? Explain. **[3 marks]**

(v) What is the econometric rationale for the 1st stage in 2SLS estimation? **[3 marks]**

(vi) Use the intuition underlying 2SLS estimation to derive the 2SLS estimator for the model:

$y=X\beta+u$ with Z denoting a matrix of instruments, and assuming that the number of moment conditions exceeds the number of parameters to be estimated. **[8 marks]**

(vii) Prove that the 2SLS estimator will be *consistent* if $Z_{1t}$, $Z_{2t}$, and $Z_{3t}$ are exogenous. **[6 marks]**

(viii) Given the fact that all the variables are covariance stationary, will it be necessary to carry out a *cointegration test*? Explain. **[2 marks]**

## Question 8: [30 Marks]

(a) Consider the two data generating processes (DGPs) indicated below:

$x_t = \delta + x_{t-1} + \upsilon_t$-----------------------(8.1)

$y_t = \phi x_{t-1} + \varepsilon_t$--------------------------(8.2)

where $\varepsilon_t$ and $v_t$ are *independent white noise processes*.

(i) What kind of stochastic process is represented by Equation (8.1)? **[3 marks]**

(ii) What kind of stochastic process is represented by Equation (8.2)? **[3 marks]**

(iii) Suppose that you obtain a sample of T observations from the two aforementioned DGPs, and you estimate the following model:

$$y_t = \beta x_{t-1} + u_t \text{----------------------(8.3)}$$

where $u_t$ is a zero mean error term with a constant variance, $\sigma^2$, and follows a normal distribution.

(1) Write down the OLS estimator for $\beta$ in Equation (8.3), taking into account the moving average representation for $x_t$ in Equation (8.1) **[5 marks]**

(2) Determine whether the OLS estimator will be consistent. **[5 marks]**

(3) What is the meaning of a super-consistent estimator? **[4 marks]**

(b) A PhD candidate collected a sample of 100 observations on two variables, $x_t$, and $y_t$ from an unknown data generation process (the two time series variables are plotted in Figure 8.1 below), and estimated the following simple regression model by OLS

$$Y_t = \alpha + \beta X_t + u_t \text{………………………..(8.4)}$$

where $u_t$ is a zero mean error term with a constant variance, $\sigma^2$, and follows a normal distribution. Tables 8.1 and 8.2 present the regression results

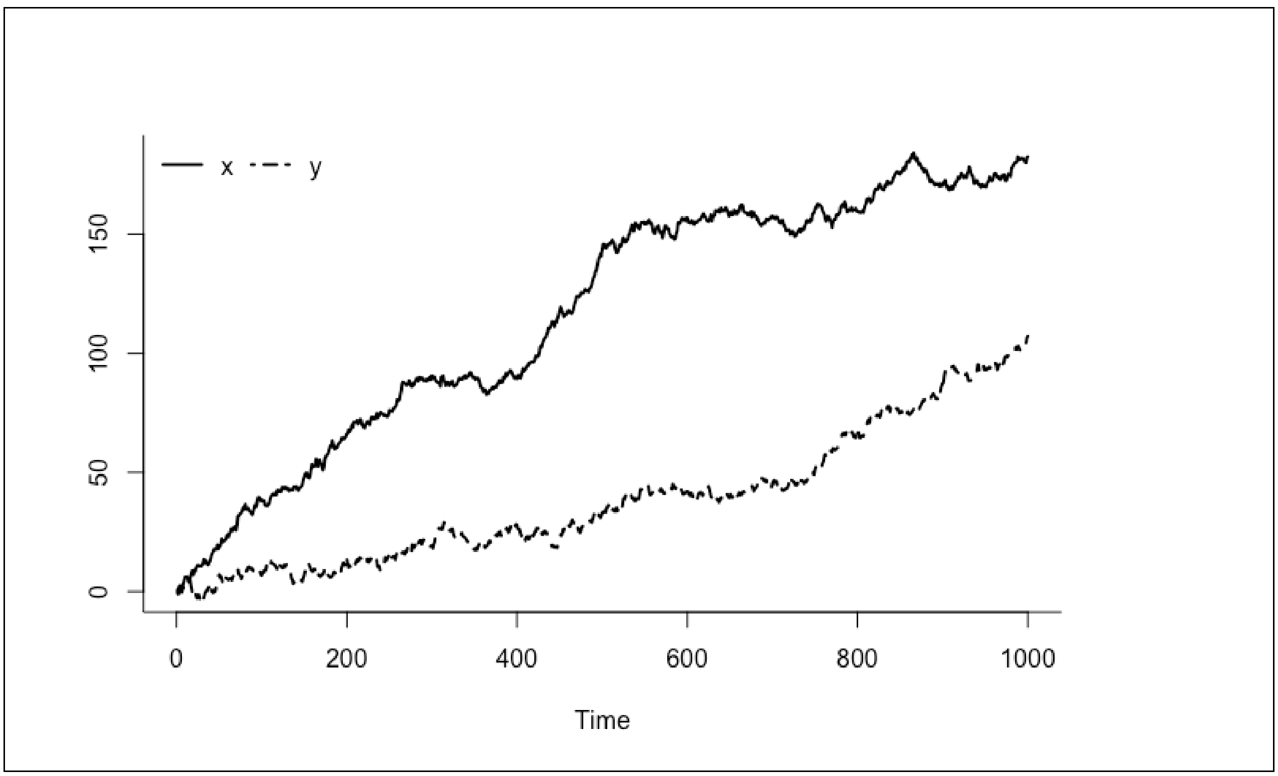


*Figure 8.1 Plot of the time series variables*

Table 8.1 OLS regression results

| | *Dependent variable:* |
|---|---|
| | $Y_t$ |
| $X_t$ | $0.472^{***}$ |
| | (0.008) |
| Constant | $-16.488^{***}$ |
| | (1.070) |
| Observations | 1,000 |
| $R^2$ | 0.760 |
| Adjusted $R^2$ | 0.760 |
| Residual Std. Error | 13.940 (df = 998) |
| F Statistic | $3,161.821^{***}$ (df = 1; 998) |
| *Note:* | $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$ |

DW = 0.0063

The numbers reported in parentheses below the estimated intercept and slope parameters are the standard errors

Table 8.2 OLS regression results after including a lagged dependent variable as another explanatory variable

| | *Dependent variable:* |
|---|---|
| | $Y_t$ |
| $Y_{t-1}$ | $0.995^{***}$ |
| | (0.002) |
| $X_t$ | 0.002 |
| | (0.001) |
| Constant | -0.093 |
| | (0.085) |
| Observations | 999 |
| $R^2$ | 0.999 |
| Adjusted $R^2$ | 0.999 |
| Residual Std. Error | 0.992 (df = 996) |
| F Statistic | $407,823.000^{***}$ (df = 2; 996) |
| *Note:* | $^{*}p <0 .1$, $^{**}p < 0.05$, $^{***}p < 0.01$ |

The numbers reported in parentheses below the estimated intercept and slope parameters are the standard errors

(i) Look at Table 8.1 and Figure 8.1. What kind of the data generation processes do you suspect generated the data for the two variables? Explain. **[4 marks]**

(ii) Are the OLS estimates reported in Table 8.1 unbiased/consistent? Explain. **[2 marks]**

(iii) What conclusion(s) can you draw from the regression results in Table 8.2? Explain. **[4 marks]**