KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2015/2016

THIRD SEMESTER EXAMINATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

EES 905: COMPREHENSIVE ECONOMETRIC THEORY AND PRACTICE

DATE: Friday 5th August 2016 **TIME: 9.00a.m-12.00p.m**

INSTRUCTIONS:

1. You are required to answer **Four (4)** questions in total. **ONE (1)** from Section I, **ONE (1)** from Section II and **TWO (2)** from Section III.
2. Relevant formulae and Tables are provided when necessary.

SECTION I

Instructions: section I is composed of Questions (1) and (2). **Choose ONE (1)** question from this section.

Question (1) [15 marks]

Consider the linear regression model

$$Y_t = \alpha + \beta X_t + e_t \qquad t = 1, 2, \ldots, T \qquad (1.0)$$

Answer the following questions.

(a) State the assumptions of the classical linear regression model in order to estimate the unknown parameters in (1.0).

(b) Find the least squares estimators for α and β.

(c) Derive the variance for least squares estimators for α and β.

Suppose a researcher applied model (1.0) to estimate the relationship between food expenditure (Y) and income (X), using the observations from 100 households in a particular country (both quantities in USD). The estimated results are as follows:

$$\hat{Y}_t = 83.42 + 0.1021\, X_t \qquad R^2 = 0.385$$

$$(se) \quad (43.41) \quad (0.0209)$$

(d) Construct a 95% confidence interval for β.

(e) Test the hypotheses that $\alpha = 0$ and $\beta = 0$.

(f) Evaluate and interpret the coefficients of the model.

(g) Interpret the coefficient of determination.

Question (2) [15 marks]

Consider the simple regression

$$Y_t = \beta_1 + \beta_2 X_t + e_t \qquad t = 1, 2, 3, 4, 5$$

where the e_t are independent errors with $E(e_t) = 0$ and $var(e_t) = \sigma^2 X_t^2$. Suppose you have the following observations:

$$Y = (4, 3, 1, 0, 2) \qquad X = (1, 2, 1, 3, 4)$$

(a) Find the generalized least squares estimates of β_1 and β_2.

(b) What are the consequences of heteroskedasticity for the least squares estimator?

(c) How would one detect the problem of heteroskedasticity?

(d) How would one deal with the problem of heteroscedasticity in order to find efficient estimates? Explain.

SECTION II

INSTRUCTIONS: Section II is composed of Questions (3) and (4). Choose ONE question from this section.

Question (3) [25 marks]

Suppose you are working in a reputable research institution in Kenya and you are given an assignment to estimate a production function that will describe the contribution of labor and capital to the country's output (measured in gross domestic product, GDP). Economic theory suggests that output (Q) is a function of capital (K) and labor (L) as given below:

$$Q = f(K, L) \tag{3.0}$$

Since equation (3.0) does not specify the form of the production technology, you decided to estimate the three (3) competing models:

(i) Production function in linear form: $Q_i = \alpha_0 + \alpha_1 K_i + \alpha_2 L_i + e_i$ (3.1)

i = 1979, 1980, 1981, …, 2001, 2002

(ii) Production function in Cobb-Douglas form: (3.2)

$$Q_i = AK_i^{\beta_1} L_i^{\beta_2}$$

or $\ln Q_i = \beta_0 + \beta_1 \ln K_i + \beta_2 \ln L_i + u_i$ where $\beta_0 = \ln(A)$

(iii) Production function in translog form: (3.3)

$$\ln Q_i = \log A + \delta_k \log K_i + \delta_L \log L_i + \delta_{KK}[\log K_i]^2 + \delta_{LL}[\log L_i]^2 + \delta_{LK} \log K_i \log L_i + v_i$$

If restrictions: $\delta_{KK}=\delta_{LL}=-1/2\delta_{LK}$ is binding, then the production technology leads to the Constant Elasticity of Substitution (CES) function.

The estimation results using STATA program are given on *Appendix A*.

(a) Evaluate the results for model (3.1) and interpret the coefficients. Are the estimated coefficients statistically significant?

(b) Evaluate the results for model (3.2) in the standard way and interpret the coefficients. Are the estimated coefficients statistically significant?

(c) Test the hypothesis of constant returns to scale, that is H0: $\beta_1 + \beta_2 = 1$ vs. H0: $\beta_1 + \beta_2 \neq 1$.

(d) Stata output shows the re-estimate model of (3.2) imposing restriction of constant returns to scale. Evaluate the results and interpret the estimated coefficients. Are they significant?

(e) Test the restrictions that $\delta_{KK}=\delta_{LL}=-1/2\delta_{LK}$.

(f) Based on the estimated coefficients and different tests, which model will you choose? Why? According to your model of choice, how would you describe the production processes in Kenya?

Question (4) [25 marks]

Suppose you are researcher and constructed the following rather simplistic model of the Kenyan economy:

$$C_t = \beta_1 Y_t + \gamma_{11} + \gamma_{12} C_{t-1} + u_{1t} \qquad (4.1)$$

$$I_t = \beta_2 Y_t + \gamma_{21} + \gamma_{22} r_t + u_{2t} \qquad (4.2)$$

$$Y_t = C_t + I_t + G_t \qquad (4.3)$$

where C_t is private consumption expenditure in year t

I_t is private investment expenditure in year t

Y_t is gross national expenditure in year t

G_t is government expenditure in year t

R_t is a weighted average of interest rates in year t.

The variables C_t, I_t and Y_t are considered endogenous variables while G_t and r_t are considered exogenous variables.

a) Briefly explain what signs you would expect *a priori* for each of the regression coefficients and the theoretical basis for your expectations.

b) Derive algebraically the reduced form coefficients of Y_t, where the reduced form equation is given as:

$$Y_t = \pi_{10} + \pi_{11} C_{t-1} + \pi_{12} r_t + \pi_{13} G_t + v_{1t} \qquad (4.4)$$

c) Using first, the order condition and then, the rank condition, examine equation (4.1) and (4.2) to determine their identification status.

d) Based on the identification status found in (c), what is the appropriate estimation method may be used to consistently estimates (4.1) and (4.2). Explain.

e) The results of estimation for regression model in equation (4.2) using OLS and 2SLS as well as the estimated reduced form coefficients are presented in *Appendix* 2. Based on your overall analysis, which estimation results are preferred. Justify and interpret the set of coefficients you have chosen.

SECTION III

INSTRUCTIONS: Section III is composed of Questions (5) and (7). Choose any TWO (2) questions from this section.

Question (5) [30 marks]

Consider the following three dimensional VAR(p) model:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t \qquad (5.1)$$

Where $y_t = (x_t, z_t, w_t)'$

A_i is a (3×3) matrix of coefficients, $i = 1, 2, \ldots, p$

(a) Derive the error correction representation of (5.1).

(b) Explain the concept of error correction representation in econometric modelling.

(c) Denote the matrix of the long-run coefficients obtained from the above model as π, what are the implications of the following in estimating the VAR system?

(i) If the rank of π is zero, that is $r(\pi) = 0$

(ii) If the rank of π is 3, that is $r(\pi) = 3$

(iii) If the rank of π is 2, that is $r(\pi) = 2$

(d) What is cointegration?

(e) What are the advantages of using Johansen cointegration test over the Engle-Granger 2-step cointegration test?

Question (6) [30 marks]

A common assumption in count data models is that, for given x_i, the count variable y_i has a Poisson distribution with expectation $\lambda_i \equiv \exp\{x_i'\beta\}$. Therefore, it follows that the density of the Poisson regression model for a single observation is given by

$$f(y_i/x_i) = \frac{e^{-\exp(x_i'\beta)}\exp(x_i'\beta)^{y_i}}{y_i!} \qquad y = 0, 1, 2, \ldots$$

(a) Give some characteristics of count data.

(b) Give an example of researchable economic problem where count data models may be applied.

(c) Derive a log-likelihood function, which the Poisson ML estimator maximizes.

(d) Derive the expression for the marginal effects, that is, the impact of a change in x_{ik} on the expected value of y_i. [where x_{ik} is a continuous variable.]

(e) What are the major deficiencies of the Poisson model?

Question (7) [30 marks]

Panel data refers to the pooling of cross-section observations over several periods. Suppose there are observations on N individuals, i = 1, 2, ..., N, taken over T periods of time, t = 1, 2, ..., T. Consider the following panel data model where the dependent variable, y_{it} may be described as a linear function of the individual's previous realizations, and of explanatory variables, x_{it} based on some fundamental economic theory.

If the relationship is static, the econometric expression maybe written as:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it} \qquad (7.1)$$

where: $u_{it} \sim IID(0,\sigma_{it}^2)$, and x_{it} is uncorrelated with u_{it}.

(a) What are the conventional assumptions on the behavior of the individual-specific effects, α_i to be considered before estimating the model in equation (7.1)?

(b) Depending on the assumptions on the behavior of α_i, what estimators would you suggest to consistently and efficiently estimate the model in equation (7.1)?

(c) One of the many advantages of using panel data is that it is able to address the problem of omitted variable bias. Show how a fixed-effect model is able to rectify this problem.

(d) Suppose y_{it} is the production output in 35 African countries from 2001 to 2004, and the corresponding $x_{it} = (x_{1it}, x_{2it})'$ are production inputs, respectively, capital and labor. All variables are expressed in logs. The coefficients from the estimated model with constant returns to scale, and some test results are presented below.

Table 7.1: Panel data estimates of the growth equation

Dependent variable: log of production	*FE* Coefficients (t-values)	*RE* Coefficients (t-values)
Constant	2.123*** (9.66)	1.820*** (5.73)
Elasticity of output with respect to capital	0.371*** (8.25)	0.447*** (10.67)
Elasticity of output with respect to labor	0.629*** (13.98)	0.553*** (6.29)
Adjusted R-squared	0.99	0.73
No. of cross-sections	35	35
Total no. of observations	140	140
Hausman test: $\chi^2(2) = 18.70$***		
White test heteroskedasticity: $\chi^2(5) = 4.87$		

Figures in parenthesis are t-values. *** Significance at 1%, ** Significance at 5 %

(i) What conclusion can you draw from the computed value of the Hausman test? State the null and alternative hypotheses.

(ii) What conclusion can you draw from the heteroskedasticity test?

(iii) Interpret the coefficients of the model of your choice based on the tests (i) and (ii) above.

* * * * * * * * *

APPENDIX A. Stata Output: Estimation of Production Functions Applied to Kenyan Data (1979-2002)

```
. insheet using "C:\Users\Portege\Documents\Production Data set Kenya example.csv"
(4 vars, 24 obs)
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |        24      1990.5    7.071068       1979       2002
     capital |        24    482.0658    80.96568     352.08     615.87
         gdp |        24      187.31    38.35524     122.59     238.74
      labour |        24    11.63036    2.742587        7.5     16.296

. regress gdp capital labour
      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  2,    21) =  272.10
       Model |  32578.684     2  16289.342             Prob > F      =  0.0000
    Residual | 1257.17593    21  59.8655205            R-squared     =  0.9628
-------------+------------------------------           Adj R-squared =  0.9593
       Total | 33835.8599    23  1471.12434            Root MSE      =  7.7373

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     capital |   .3714857   .3578549     1.04   0.311    -.3727142    1.115686
      labour |   2.759341   10.56447     0.26   0.796    -19.21067    24.72935
       _cons |  -23.86271   50.32395    -0.47   0.640    -128.5171    80.79168
------------------------------------------------------------------------------

. generate lgdp = log(gdp)
. generate llabour=log(labour)
. generate lcapital=log(capital)
```

```
. regress lgdp lcapital llabour

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  2,    21) =  328.65
       Model |  1.03419948     2  .517099741           Prob > F      =  0.0000
    Residual |  .033041791    21  .001573419           R-squared     =  0.9690
-------------+------------------------------           Adj R-squared =  0.9661
       Total |  1.06724127    23  .046401795           Root MSE      =  .03967

------------------------------------------------------------------------------
        lgdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcapital |  -.3687685   .7067532    -0.52   0.607    -1.838542    1.101005
     llabour |   1.141347   .4964515     2.30   0.032     .1089202    2.173775
       _cons |   4.715227   3.156225     1.49   0.150    -1.848503    11.27896
------------------------------------------------------------------------------

. test lcapital = 1 - llabour

 ( 1)  lcapital + llabour = 1

       F(  1,    21) =    1.13
            Prob > F =    0.3006

. constraint 1 lcapital = 1 - llabour

. cnsreg lgdp lcapital llabour, constraint(1)

Constrained linear regression                      Number of obs   =        24
                                                   Root MSE        =    0.0398

 ( 1)  lcapital + llabour = 1
------------------------------------------------------------------------------
        lgdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcapital |   .3715558   .1139353     3.26   0.004     .1352685    .6078431
     llabour |   .6284442   .1139353     5.52   0.000     .3921569    .8647315
       _cons |   1.395991   .4259823     3.28   0.003     .5125576    2.279424
------------------------------------------------------------------------------
```

```
. generate lk_sq = lcapital^2

. generate ll_sq = llabour^2

. generate lk_ll = lcapital*llabour


. regress lgdp lcapital llabour lk_sq ll_sq lk_ll

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  5,    18) =  312.91
       Model |  1.05510223     5  .211020447           Prob > F      =  0.0000
    Residual |  .012139041    18  .000674391           R-squared     =  0.9886
-------------+------------------------------           Adj R-squared =  0.9855
       Total |  1.06724127    23  .046401795           Root MSE      =  .02597

------------------------------------------------------------------------------
        lgdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcapital |   324.6393   416.3109     0.78   0.446    -549.9974    1199.276
     llabour |  -224.5402    284.208    -0.79   0.440    -821.6391    372.5587
       lk_sq |  -36.30283   46.16611    -0.79   0.442    -133.2942    60.68855
       ll_sq |  -18.69105   21.45867    -0.87   0.395    -63.77405    26.39195
       lk_ll |   51.14569   62.99436     0.81   0.427    -81.20055    183.4919
       _cons |   -726.587   938.2975    -0.77   0.449    -2697.877    1244.703
------------------------------------------------------------------------------


. test (lk_sq = ll_sq) (lk_sq = (-1/2)*lk_ll)

 ( 1)  lk_sq - ll_sq = 0
 ( 2)  lk_sq + .5*lk_ll = 0

       F(  2,    18) =    1.46
            Prob > F =    0.2585
```

APPENDIX 2. Estimated Regression Results for Equation (4.2)

A. Using OLS method

Dependent Var (i)	Coef.	Std. Err.	t	P>|t|
y	.3833252	.0274894	13.94	0.000
r	-.1821539	.2229979	-0.82	0.418
_cons	-5.643832	4.351007	-1.30	0.201

B. Estimated Reduced Form Method

Dependent Var (i)	Coef.	Std. Err.	t	P>|t|
l_cons	.461371	.0685352	6.73	0.000
r	-.281381	.2469569	-1.14	0.261
g	1.618368	.4421462	3.66	0.001
_cons	-9.734629	5.1762	-1.88	0.067

C. Using 2SLS method

Dependent Var (i)	Coef.	Std. Err.	t	P>|t|
yhat	.4100774	.0329115	12.46	0.000
r	-.1458295	.2471141	-0.59	0.558
_cons	-8.494901	4.961258	-1.71	0.094