KENYATTA UNIVERSITY

EXAMINATION FOR THE DEGREE OF MASTER OF ECONOMICS

ECO 801: COMPUTER SIMULATION AND ECONOMIC DATA ANALYSIS

3RD TRIMESTER 2024/2025 **TIME: 3 HOURS**

INSTRUCTIONS:

Answer all questions SECTION I and one question in SECTION II

NB: STATA OUTPUT, Word File and Do-File Documents Will Be Examined

SECTION I

QUESTION 1

A researcher investigated the factors determining income velocity in Kenya using quarterly data from the first quarter of 2002 to the fourth quarter of 2022. Table 1 (see appendix) below explains the description of variables used in the model. Use Table 2 and 3 (see appendix) to

i) Justify the use of Vector Error Correction Model (4 Marks)

ii) Explain [clearly but concisely] any diagnostic tests that would conduct before interpreting the model presented in Table 3 (6 Marks)

iii) Interpret the error correction term and the long run equation of the model whose dependent variable is log of income velocity and clearly cite the implications of the findings (6 Marks)

iv) Interpret the results in figure 1 and make possible policy inferences based on the information available (4 Marks)

QUESTION TWO

Using the data provided in the appendix load the data in STATA software and:

a) Create and label dummy variables to represent marriage and occupation (3 Marks)

b) Generate a new variable called experience calculated as the square root of the product of age and tenure (3 Marks)

c) Create a new variable called department by combining the subordinate staff and business support category NB: bank teller and managers category should be retained as they are (3 Marks)

d) Regress Experience against Education, department, and Marriage and conduct relevant diagnostic tests (6 Marks)

e) Comment on the validity of your regression model and interpret the results (5 Marks)

SECTION II

QUESTION THREE

You have been commissioned to investigate factors that determine salary in the banking industry. To answer the question, you obtain data from 136 bankers and adopt the following regression model (Use STATA the data provided in the appendix)

$$lsal_i = \beta_0 + \beta_1 tenure_i + \beta_2 Grade_i + \beta_3 Age_i + \beta_4 Dept_i + \beta_1 MaritalS_i + \varepsilon_i$$

Where $lsal_i$ represent the log of monthly salary for individual i

$tenure_i$ represent no. of years in banking service for individual i

$Grade_i$ represents no. of schooling years for individual i

Age_i represent age in years for individual i

$Dept1_i$ rep the department an individual is employed such that $Dept_i = \begin{cases} 1, & Manager \\ 2, & Surbodinate \\ 3, & Business\ Support \\ 4, & Bank\ teller \end{cases}$

$MaritalS_i$ represent marital status such that $MS_i = \begin{cases} 0, & Single \\ 1, & Married \end{cases}$

i) Explain the nature of the distribution of salary and plot the graph (4 Marks)

ii) Construct and comment on the results obtained from the kernel density estimate and normality test and the extent to which the findings affect reliability of the regression model results

(4 Marks)

iii) Test for heteroskedasticity and multicollinearity and comment on the findings of the heteroskedasticity test presented in Table 4. What are the consequences of presence of heteroskedasticity (6 Marks)

iv) Interpret the coefficients and comment on the strength of the equation (6 Marks)

b)	Dep Variable: Natural log of Salary
Tenure	0.0252***
	(3.77)
Education	0.0858***
	(5.79)
Age	-0.0140
	(-1.13)
Business Support	0.119
	(1.39)
Manager	0.757***
	(6.44)
Marital status	0.0910
	(1.20)
_cons	10.20***
	(18.78)
N	136

t statistics in parentheses
Bank teller and NOT married are used as reference category
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

QUESTION FOUR

a) The following regression estimates examined the trade-off between time spend Leisure and working. Work was measured in hours per week. Education is measured using no of schooling hours. Income is measured in Kenya Shillings while Health was measured using an index from 0 to 100 such that zero (0) indicates poor health status. NB: Standard errors estimates in parenthesis.

$$Health = \underset{}{18} + \underset{(2.5)}{0.5Educ} - \underset{(0.1)}{5.625Work} + \underset{(2.01)}{0.21Income} - \underset{(0.045)}{0.05Inc^2} \qquad \underset{(0.005)}{Adj.\,R\,Sq. = 0.55}$$

Required

i) Test the significance of the variables: assume critical value of $t_{0.025} \approx 2$ (5 Marks)

ii) Interpret all the coefficients and make possible policy inferences (5 Marks)

b) The following regression estimates the determinants of income. Income was measured in hours per week. Experience, Education and age are measured in years and other control variables (X)

$$Income = \alpha_0 + \alpha_1 Exp + \alpha_2 Educ + \alpha_3 Age + \alpha_4 X + \varepsilon$$

Assume that $E(X, \varepsilon) = 0$

i) You suspect that Experience is endogenous, demonstrate how would you test for endogeneity (5 Marks)

ii) Assume that endogeneity is established, how would you estimate the model (5 marks)

APPENDIX

Table 1: Description and Measurement of Variables

Abbreviation	Variable	Type	Description
LNV2	Log of Income Velocity	$\frac{GDP}{MS}$	Derived from the quantity theory of money. GDP was used as a proxy of income. Money supply was measured using broad money (M2)
LNGDP	Log of GDP	Income	Represent the real output or real GDP of the economy. Deflated using GDP deflator
LNINF	Log of interest	Financial variable	Real interest rate at time t. Deflated using GDP deflator. Nominal interest rates was represented by T-bill rates
LNBRAN	PBO	Institutional variable	Number of bank branches in Kenya
LNMON	Bank assets-GDP ratio	Institutional variable	This represents the degree monetization and by extension the level of financial system development
LNREER	Log of real exchange rate	Financial variable	Real exchange rate at time t.

Table 2: Johansen Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.370808	41.81170	29.79707	0.0013
At most 1	0.050085	4.282861	15.49471	0.8794
At most 2	0.001491	0.120847	3.841466	0.7281

Table 3: Vector Error Correction Estimates
Sample (adjusted): 2002Q4 2022Q4

Included observations: 77 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
LNV2(-1)	1.000000
LNBRAN(-1)	0.013764 (0.06646) [0.20711]
LNMON(-1)	0.893191 (0.06214) [14.3740]
LNREER(-1)	-0.236899 (0.02541) [-9.32305]
LNINFL(-1)	-0.027536 (0.00809) [-3.40158]
C	0.728192

Error Correction:	D(LNV2)	D(LNBRAN)	D(LNMON)	D(LNREER)	D(LNINFL)
CointEq1	-0.412137 (0.15849) [-2.60036]	-0.290519 (0.04986) [-5.82722]	-0.371259 (0.16799) [-2.21003]	0.556624 (0.11815) [4.71121]	1.638130 (1.00365) [1.63217]
D(LNV2(-1))	-0.650223 (0.19471) [-3.33946]	0.115058 (0.06125) [1.87857]	0.896259 (0.20637) [4.34288]	-0.451506 (0.14515) [-3.11069]	1.596020 (1.23299) [1.29443]
D(LNV2(-2))	-0.466101 (0.15877) [-2.93572]	-0.067205 (0.04994) [-1.34565]	0.590028 (0.16828) [3.50620]	-0.424531 (0.11836) [-3.58693]	-0.375462 (1.00540) [-0.37344]
D(LNBRAN(-1))	0.395552 (0.21580) [1.83296]	0.122680 (0.06788) [1.80726]	-0.076030 (0.22873) [-0.33241]	-0.204681 (0.16087) [-1.27235]	-1.808298 (1.36655) [-1.32326]
D(LNBRAN(-2))	-0.759545 (0.20427) [-3.71825]	0.412687 (0.06426) [6.42248]	1.454398 (0.21651) [6.71736]	-0.416983 (0.15228) [-2.73831]	-0.732573 (1.29357) [-0.56632]

D(LNMON(-1))	-0.242651	0.230747	0.220451	-0.103312	1.377374
	(0.18036)	(0.05673)	(0.19117)	(0.13445)	(1.14214)
	[-1.34535]	[4.06711]	[1.15318]	[-0.76839]	[1.20596]
D(LNMON(-2))	-0.124360	-0.030705	0.147860	-0.455457	0.138777
	(0.12968)	(0.04079)	(0.13745)	(0.09667)	(0.82120)
	[-0.95897]	[-0.75272]	[1.07574]	[-4.71142]	[0.16899]
D(LNREER(-1))	0.065609	-0.009686	-0.034227	0.720736	2.780373
	(0.16062)	(0.05052)	(0.17024)	(0.11973)	(1.01711)
	[0.40848]	[-0.19171]	[-0.20105]	[6.01952]	[2.73360]
D(LNREER(-2))	-0.027003	-0.130264	0.019389	-0.037794	-0.835519
	(0.16760)	(0.05272)	(0.17764)	(0.12494)	(1.06133)
	[-0.16112]	[-2.47084]	[0.10915]	[-0.30250]	[-0.78724]
D(LNINFL(-1))	-0.018945	-0.000181	0.013368	0.001430	0.076367
	(0.02104)	(0.00662)	(0.02230)	(0.01568)	(0.13322)
	[-0.90053]	[-0.02742]	[0.59949]	[0.09121]	[0.57323]
D(LNINFL(-2))	0.010182	0.001243	-0.018229	0.011309	0.057711
	(0.02025)	(0.00637)	(0.02146)	(0.01510)	(0.12824)
	[0.50280]	[0.19509]	[-0.84926]	[0.74908]	[0.45002]
C	-0.017597	0.006043	0.011422	0.018074	-0.083632
	(0.01156)	(0.00363)	(0.01225)	(0.00861)	(0.07317)
	[-1.52286]	[1.66240]	[0.93255]	[2.09821]	[-1.14291]
R-squared	0.661948	0.524334	0.724646	0.647996	0.171130
Adj. R-squared	0.604739	0.443836	0.678048	0.588426	0.030860
Sum sq. resids	0.410333	0.040602	0.460974	0.228023	16.45454
S.E. equation	0.079453	0.024993	0.084213	0.059229	0.503137
F-statistic	11.57073	6.513672	15.55091	10.87792	1.220002
Log likelihood	92.27347	181.3302	87.79317	114.8931	-49.84491
Akaike AIC	-2.085025	-4.398186	-1.968654	-2.672549	1.606361
Schwarz SC	-1.719757	-4.032918	-1.603385	-2.307280	1.971630
Mean dependent	-0.014997	0.011420	0.020899	0.033481	-0.047029
S.D. dependent	0.126378	0.033513	0.148418	0.092323	0.511085

Determinant resid covariance (dof adj.)	5.88E-12
Determinant resid covariance	2.52E-12
Log likelihood	481.9150
Akaike information criterion	-10.82896
Schwarz criterion	-8.850425