REPARAMETERIZATION OF AUTOREGRESSIVE DISTRIBUTED LAG TO VECTOR ERROR CORRECTION MODEL TO STUDY YOUTH UNEMPLOYMENT IN KENYA

SHEM OTOI ODHIAMBO SAM

SCHOOL OF MATHEMATICS
COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES
UNIVERSITY OF NAIROBI

A thesis submitted in fulfillment for the award of the Degree of Doctor of Philosophy in Social Statistics.

July, 2019

DECLARATION

Candidate:

I declare that this is my original work and has not been submitted for examination in any other university.

Signed .... Date..................

Shem Otoi Odhiambo Sam

Reg: I80/50770/2016

Supervisors:

This thesis has been submitted for examination with our approval as university supervisors.

Prof. Ganesh P. Pokhariyal

School of Mathematics

University of Nairobi

Signed .... Date..................

Prof. Moses M. Manene

School of Mathematics

University of Nairobi

Signed .... Date..................

Dr. Isaac C. Kipchirchir

School of Mathematics

University of Nairobi

Signed .... Date..................

ACKNOWLEDGMENT

All glory to God for sound health of mind and body, and strength to complete this thesis. The whole process would not have been successful without continuous guidance, insights and hands-on mentorship by my supervisors Prof. Ganesh P. Pokhariyal, Prof. Moses M. Manene and Dr. Isaac C. Kipchirchir. The support of School of Mathematics fraternity at critical moments during the journey cannot be underestimated. Appreciation goes to Prof. Weke, Dr. Nelson and Dr. Mwaniki. I acknowledge my parents Mr. and Mrs. Pst. Mathews Sam, uncle John Williams and grandmother Yuanita Awuor for believing in me. I thank my wife Celestine Sam and children Chryseis and Achilles for peace of mind and conducive studying environment at home. I do thank Morel Fourman for his benevolence and life skills coaching.

ABSTRACT

The research provides statistical basis for assessing and prioritizing investment policies, initiatives and projects to maximise youth employment by scrutinizing influence of macroeconomic variables. The macroeconomic variables considered are gross domestic product (GDP), external debt (ED), foreign domestic investment (FDI), private investment(PI), youth unemployment(YUN), literacy rate (LR), and youth population (POP). The research approach taken uses predictive analytics such as impulse response functions and variance decomposition from vector error corrections model (VECM) and cointegration regression in autoregressive distributed lag (ARDL) to identify key determinants of youth unemployment to prioritize investment. This research analyzes reparameterization of ARDL to VECM through cointegration of time series. First, the time series data undergo logarithm transformation to reduce outlier effects and have elasticity interpreted in terms of percentage. The study scrutinizes the effects of macroeconomic shocks on youth unemployment in Kenya. For this purpose, the Augmented Dickey-Fuller test is conducted to assess stationarity of the variables used. Then Johansen Cointegration test is carried out to establish the rank at which the series are cointegrated. The unit root test has been performed on YUN, GDP, ED, FDI, PI, and LR, and POP to assess stationarity. The cointegrated dynamic ARDL model is estimated using ordinary least squares (OLS) and effects of variables and their lags interpreted. The results reveal that Gross Domestic Product (GDP) and its second lag have negative effect on youth unemployment, that is, one unit increase in (GDP) and GDP lag 2 reduce youth unemployment by 0.207922% and 0.2052705% respectively. Also, one unit of External Debt (ED) and ED lag 2 reduce youth unemployment by 0.07303% and 0.009116% respectively. Furthermore, unit increase in one year lag of youth literacy rate reduces youth unemployment by 0.0892691%. Lastly, lag one and three of population reduce youth unemployment by 0.2590455% and 4.3093119% respectively. The Johansen Cointegration Analysis has revealed three long run relationships which can be interpreted as a GDP effect; External Debt effect and Foreign Direct Investment effect relations. A structural VECM has been described through restrictions taken from the Cointegration Analysis. Based on the results of the Impulse-Response Function and variance decomposition analyses of the Structural VECM, it is concluded that GDP, literacy level, population, and FDI shocks have significant effects on Kenyan youth unemployment in the long run. On the superiority of the two models, whereas ARDL captures the influence of past shocks through coefficients of lags, VECM predicts the effects of current shocks and resulting movement of variables more than 10 unit steps ahead. Also, Granger causality present in ARDL does not exist in reparameterized VECM. The F-test and t-test reveal that the two models are significant at 95% confidence level. However CUSUM test shows that the estimated ARDL is more stable.

ABBREVIATIONS

ADEA	Association for Development of Education in Africa
ADF	Augmented Dickey-Fuller Test
AIC	Akaike Information Criterion (AIC)
AEO	African Economic Outlook
AfDB	African Development Bank
ARDL	Autoregressive Distributed Lag
CIDP	County Integrated Development Plan
ECM	Error Correction Model
ED	External Debt
FDI	Foreign Direct Investment
FPE	Final Prediction Error
GDP	Gross Domestic Product
ICT	Information Communication and Technology
ILO	International Labour Organization
IMF	International Monetary Fund
LR	Literacy Rate
MLE	Maximum Likelihood Estimation
MoEST	Ministry of Education Science and Technology
UNDP	United National Development Programs
OECD	Organization for Economic Co-operation and Development
PI	Private Investment
SME	Small Medium Entreprises
SC	Schwartz Criterion
SSE	Sum of Squared Errors
VAR	VectorAutoregression
WGI	World Governance Indicator

TABLE OF CONTENTS

List of Tables

List of Figures

1 INTRODUCTION

1.1 Background of Information

Following the 2016 priority of the African Heads of State, and like other African Countries, Kenya has adopted the demographic dividend policy. Like many other countries Kenya has a national vision (Vision 2030), Kenya has become medium income country, with a roadmap of successive five year Medium Term Plans (MTPs) laid out to achieve the national vision. National MTPs, citizen priorities and government policies are further reflected in County Integrated Plans (CIDPs) of county governments. Annual Development Plans (ADPs) and Annual Work Plans (AWPs) capture detailed annual investment and development plans. For Kenya to reap the rewards of demographic dividend, short, medium, and long term policies on public investment must be carefully selected. Investment in labour intensive sectors to reduce youth unemployment supports gains in rewards of Demographic Dividend.

In all countries, more youth are discouraged than employed. This suggests that the youth unemployment challenges has been underestimated AfDB (2012). Youth accounts for 35% of Kenya's population and 'represent 67% of unemployment Census (2009). The youth population growth is 4% while national growth rate in 2.9%. As such youth population growth is higher than national population growth. Studies have linked youth unemployment to social evils and political instability. The unemployment creates a culture of idleness and group-up tendencies. The "youth bulge" is a source of concern for governments and development partners Okojie (2003) observed that unemployment drives youth to engage in illegal activities like touting, stealing, armed robbery, dealing in drugs, gambling and prostitution. Far from social evils like crime, African governments have another concern: political stability. According to Urdal (2006) youth bulges increase the risk of three forms of political violence: rioting, civil war, and terrorism. The unemployed youth bulge becomes political pond by politicians and extremist groups for devious activities ADEA (2014). The Kenyan political history is livid with political violence and intolerance. In the recent past, Kenya experienced continued sporadic Islamic extremists terrorist attacks among other internal organized crime involving youth as perpetrators. These social problems are linked to economic situation of the youth which makes them unproductive.

Unemployment is a situation in which persons of sound health of mind and body are willing and ready to work but are unable to get job opportunity (Davidson, 1998). Notably, customs vary across countries. It is notable that because customs and behavior vary across countries and the description of the term youth also varies (United Republic of Tanzania, 1995). Whereas the united nations considers persons in 15-24 age bracket to be youth, the Kenya constitution sets youth age to be 18-35 years. In most African countries, persons aged 15-24 years are engaged in pursuit of education in various levels: upper primary, secondary, university and other tertiary colleges. Therefore, in this case, youth unemployment rate is the proportion of persons aged between 18-35 years and actively looking for employment but unable to find work expressed as percentage of working population. Whoever is not engaged in gainful employment or income generating activity and not searching for work does not fall under the category of unemployed.

The Kenya report, GOK (2014) ascribes youth unemployment to population growth, few job opportunities, and inadequate relevant skills Harvey et al (2002); also Okojie (2003) noted that youth unemployment is affected by cultural practices, gender, little to no access to information regarding accessible opportunities and negative ethnicity. These factors are discussed in this research.

1.1.1 Types of Unemployment

According to resource materials considered in this study, many scholars have classified types of unemployment depending on a number of socio-economic circumstances. According to Hameed et al (2012) unemployment can be classified as follows: seasonal unemployment, disguised unemployment, frictional unemployment, classical unemployment, Structural unemployment, and cyclical unemployment.

1.1.2 Global Perspective

Generally, the growing number of unemployed youth cohort is a big challenge to many countries (Kabaklalri et al 2011). Youth employment is a concern of quality and quantity to both medium and low income countries, African Development Bank, AfDB (2012). The OECD

(2010) estimates 85 million unemployed youth around the world. Whereas the global unemployment rate is 11.7 %, youth unemployment rate is 23.6 %. Okojie (2003) and Kabaklarli et al (2010) observed that unemployment rate for young people is twice their older counter part rate, studies conducted in Nigeria and Turkey respectively.

The high rates stated above are attributed to global crises affecting GDP, FDI, PI, and ED.

1.1.3 Continental Perspective

The demographic dividend is the benefit associated with declining fertility rates and increasing persons in the working age brackets. It has been deemed a solution to many problems being experienced by African countries. The African Union adopted a common position on the Post 2015 Development Agenda that resulted in the incorporating demographic dividend in the 2030 Agenda for Sustainable Development. In this way, the African Union formulated the theme "Harnessing the Demographic Dividend through investments in Youth" for the year 2017 (NCPD,2014).

In Africa, youth in vulnerable employment are in the majority where more are unemployed than working AfDB (2012). Davidson (1998) estimated global youth unemployment rate to be 27.3%, while in Africa it stood at 36.7%. In addition, Chigunta (2002) estimated that 50% of most African countries population is youth. He observed that youth between 15-25 years in Africa are 122 million with the median age between 16-20 years. Mills (2011) noted that the African young cohort is most likely to live in poverty stricken urban conditions, characterized by unemployment.

As such, the phenomenon described by various scholars as "youth bulge" is a source of concern not only to African governments but also development partners who incessantly embark on programs to bolster youth employment Okojie (2003). In Africa wage employment accounts for less than 15% of youth in the labour market AfDB (2012). Youth unemployment is concentrated between age 20-24 because ages 15-19 are still in school and about one-third of economically active youth in Africa are unemployed Okojie (2003).

According to Okojie (2003), Africa has been confronted with multidimensional challenges including drought, poverty, endemic diseases and underlying phenomenon of youth unemployment. Okojie (2003) noted that youth unemployment has become a major concern for African governments and pointed out that youth unemployment rate is twice that of adults in Africa. The situation is similar in other countries like Turkey where youth unemployment double that of adults,Kabaklarli et al (2011).

1.1.4 Regional Perspective

The population in East Africa is largely youthful. As a result of high fertility rate the population of young people is 55 per cent ,AfDB (2012).

The common barriers affecting employment in the region include: inadequate jobs, lay offs, labor costs and high wage demands.

1.1.5 Kenyan Perspective

Based on statistics from Ministry of Education, Science and Technology (MoEST), (Sessional Paper No.14, 2012 on Reforming Education and Training Sectors in Kenya), 1.2 milion youth enter labour market without formal training or skills. At age 24 less than 11% have formal training. As such education is a major determinant in youth unemployment in Kenya. Ponge (2013) found that graduates in Kenya are largely unemployed and unemployable due to lack of skills resulting from disconnect between education and labour market demand. The study concluded that youth unemployment rate needs urgent attention.

The variables used are youth unemployment rate, GDP, external debt, FDI, PI, literacy rate and youth population. Youth literacy rate and youth population are defined as;

Education/Literacy Rates:

In this case education is taken as youth literacy. The number of persons aged between 15-34 years with basic education and above, expressed as a percentage of population of people in the same age group.

Youth Population (POP):
The total number of persons aged between 15-34 in a country.

1.2 STAEMENT OF THE PROBLEM

There are currently limited studies in Kenya using Mathematical models on youth unemployment. More so there are no appropriate conventional Mathematical models that can isolate macroeconomic variable of interest and predict how they could influence unemployment over time. Also the statistical strength of ARDL and VECM have to be established to know when to use either model or reparameterized model. This research scrutinizes statistical frameworks of both models to be applied whenever necessary. The gap exists in investigating unemployment data patterns and possible interventions by investigating empirical relationship among youth unemployment, gross domestic product, population, foreign direct investment, external debt, private investment, and literacy rates.

1.3 Objectives of the study

The broad objective of this study is to develop a suitable model explaining data patterns of youth unemployment in Kenya. More precisely , the specific objectives are:

i. Identify key factors which affect youth unemployment in Kenya.

ii. Determine the short and long run effects of GDP, population growth, ED, FDI, private investment, and education on youth unemployment in Kenya.

iii. Develop a suitable model for youth unemployment in Kenya and determine equilibrium level.

iv. Comparison of short-run and long-run results.

v. Validate the models using empirical data.

1.4 Significance of the Study

The study provided extent of the effects of undertaken variables on youth unemployment so that adequate policies may be formulated by both public and private sector. The research found that there are few (if any) mathematical models explaining dynamics, patterns and trends of youth unemployment in Kenya (Otoi and Pokhariyal, 2016). This study therefore provides basis and open ground for future statistical modeling of youth unemployment in Kenya. The VECM simulation proposed in this study provides further insights into the dynamics and patterns of youth unemployment in Kenya.

In addition, the study is the first of its kind to model the effects of literacy level/education on youth unemployment. It is important to pave way for future research studies on influence of education on unemployment in general.

Also, scholars are undecided as to whether ARDL and VECM have identical framework of results, or which one of the two is a better model for estimating detrended time series data model. The study, through reparameterization, answered this question following t-type and F-type among other tests. As such, the study filled both policy and knowledge gaps and set ground for further research in the area of both mathematical models and youth unemployment.

2 LITERATURE REVIEW

2.1 Theoritical Perspective

This chapter dealt with previous studies on factors of unemployment in various countries, and consists of the theoretical and empirical reviews.

GDP:

GDP is the annual value of finished goods and services in a country. The study assumed that increase in GDP reduces youth unemployment, whereas decreases in GDP increases youth unemployment. This is a negative relationship.

There is a positive relationship between employment rates and countries economic development Okun (1962). On this, Dimian (2011) as quoted by Kabaklalri et al (2011) reported that youth unemployment rate has negative impact to country's gross domestic products. Kabaklarli et al (2011) noted that when the economic activity is healthy and developing, employment as well as youth employment will be better. However, economic meltdown and crises affect employment as well as youth unemployment negatively. According to Msigwa and Kipesha (2013), youth employment results into increased aggregate demand as well as increase in capital formation. Further on this ILO (2011) pointed out that youth are likely to spend a higher percentage of their income on goods and services, which boost the countries' aggregate demand: thus economic growth. In addition ILO (2011) explained that employed youth who received higher salaries saved and invested or deposited them in banks. The argument followed that, the savings result in increase in pool of capital which can be used to finance SME and start small businesses thereby boosting a counties economic development. In addition, Kabaklarli et al (2011) observed that young people possess a marginal propensity to consume more than adults, therefore, increasing unemployment rate in young people negatively affects consumption, total investment and as a result GDP.

On a different note, studies showed that youth employment reduces the social costs its leads to reduction of violence, criminal activities, gambling, drug addiction as well as prostitution McLean et al, (2009), as quoted by Msigwa and Kipesha, (2013). From previous studies considered herein, increase in GDP is found to reduce youth joblessness.

2.1.1 Socio-demographic Variables

They are also referred to as moderating variables. These variables are not economic in nature but play role in youth unemployment. They include education, population growth, age and gender.

Education/Literacy Rate

For the purposes of this study, education is taken as youth literacy rate.The number of persons aged between 15-34 years with basic education and above, divided by the population of people of the same age group at a particular time multiplied by 100. Table 2.1 depicts the number of young people entering job market annually. They are segmented by level of education.

Table 2.1: Number of youth entering job market annually

Level of Education	**Numbers**
No education	200,000
Drop out of primary	300,000
Complete Primary	250,000
Drop out of secondary	180,00
Complete Secondary	250,000
Drop out of tertiary education	45,000
Complete Tertiary education	155,000
Total	**1,380,000**

Population Growth

In its definition, population refers to the total number of persons of a country at a given time. The study assumed population growth leads to increase in unemployment. Youth represents an important cohort of the Kenya's population. The number of youth almost tripled from 4.94 million in 1979 to 13.67 in 2009 (Census, 2009).

When youth population grows beyond a margin the labour market can absorb, the youth unemployment increases and results in backlog of unemployment (Maqbool et al, 2013).

Gender

According to literature reviewed, young women in rural Kenya account for large numbers of the unemployed and the situation is the same in the urban areas. This is attributed to the fact that most communities in Kenya prefer to educate the boy child at the expense of girl child (Kenya Country Report 2014). This contributes to young womens' unemployment.

Age

This study revealed that, generally, active education takes place between six and nineteen years. Therefore, between fifteen and nineteen years unemployment is low as youth in this cohort are in school. Unemployment is higher among secondary school leavers and tertiary institutions' graduates.

2.1.2 Intervening Factors

This study views global economy and ICT as intervening factors in youth unemployment.

Information Communication and Technology

According to Okojie (2003) ICT is paying homage to employment sector for African youth through computer shops, internet service providers and providers and phone shops. However these jobs do not reach youths in rural areas who are avast to new technology.
Lately, there are also vibrant social media outlets like Facebook, Twitter, YouTube, WhatsApp among others which are all accessible on cell phone through mobile phone data subscriptions. These outlets have helped the youth to network among themselves. They get information on new opportunities in social media, market their skills and services locally and globally.

Also, many universities and tertiary colleges have established business incubators for young people (Ponge 2013). The incubators nature young people to come up with prototypes that can attract investments and eventual commercialization. The institutions include, Jomo Kenyatta University of Agriculture and Technology, Kenyatta University, University of Nairobi, Strathmore University, and Mount Kenya University.

Global Economy

The study found out that global economy plays critical role in employment in general. Increase in global economy growth reduces unemployment. Global economic crises have even more negative effects on youth employment as compared to adult employments. According to Kabaklarli et al (2011) the 2009 global financial crisis affected youth and adults employment disproportionately. The examples include USA where general unemployment rate is 9.4%

while youth unemployment rate is 17.6%. The same scenario is reflected in UK, Spain and Turkey. This provided a clear indication of dynamics of global economy and youth unemployment.

2.2 Empirical Review

From the considered literature, GDP is deciphered as the major youth unemployment determinant. The relationship is based on Okun's law. In his ground breaking paper, Okun (1962) described a coefficient that gives the rate of change of real GDP for a measure of unemployment rate. According to Okun's Law, an increase in the economic growth rate by 3%, above the normal rate, reduces the unemployment rate by 1%. This requires that the rate of GDP and potential growth ought to be equal.

Gul et al (2012) noted unemployment rate and rate of inflation have a negative relationship in a countrys' economy. It is acknowledged clearly, the lower the unemployment in an economy, the higher the rate of inflation.

Also Eita and Eshipala (2001) used data between 1971 and 2007 and revealed that incases where real GDP is below potential GDP unemployment increased.
Kabaklarli et al (2011) employed long term cointegration analysis from an econometric analysis perspective to determine the effects of the discussed variables of youth unemployment as is done by Eita and Eshipala (2001). The results indicated that inflation and productivity had positive effects on youth unemployment rate despite the fact that GDP and investment had negative effects on the long-run.

From the considered literature, the most commonly used variables of youth unemployment are population, GDP, FDI, PI, and ED. In this study, we consider Education as one of the key determinants. All studies considered mentioned the importance of education in youth unemployment but none provided a mathematical model on it. It is for these reasons that this study adopted the same variables. On mathematical models Hassler and Wolters (2010) considered cointegration analysis within an ARDL framework and the review of cointegration tests based on ECM regression paying particular attention to linear time series without detrending. The results indicated that conditional ECM model is superior to unconditional one.

For stationarity test and cointegration to solve challenges of spurious regression Engle and Granger (1987), Johansen (1991), Phillip (1991), Phillip and Hansen (1990), and Phillips and Loretan (1991) offer solutions. The tests are further simplified by Dickey and Fuller (1978) and (1981) in ADF unit root test. The reparameterized ECM and ARDL are suggested by Pesaran and Shin (1998) under exogeneity and not endogeneity conditions without detrending to have numerical identity. The study sort to find statistical framework identity when variables are detrended appropriately.

The results from ARDL and VECM underwent to t-type and F-type tests as suggested by Banerjee et al (1998) and Boswijk (1994). The results of Hassler and Wolters (2006) revealed that t-type cointegration test can replace F-type cointegration test since they are equally appropriate in comparative analysis.

3 COINTEGRATION

3.1 Introduction

The information included in this section is useful in the next chapters in the explanations of further analysis of models. The series used in this section are tested for unit roots to verify their stationarity at level, that is, without detrending. Correlation test is done to establish causality-effect in the subsequent sections. The models considered by this study required cointegration tests of macroeconomic variables conducted earlier in this chapter. The purpose of the tests is to ensure validity, replicability and ultimately generalization of the results.

The study uses secondary national macroeconomic data from World Bank and Kenya National Bureau of Statistics. This fell under the category of document analysis of existing data.

The macroeconomic annnual data covering 1979-2015 period. The used data include GDP, YUN, FDI, PI, ED, Population and Youth Literacy Rate (Education). The previous studies used the variables discussed interchangeably, with exception of education (Youth Literacy Rate) which has not been used before. The variable are used interchangeably by Maqbool et al. (2013), Kabaklarli et al. (2011), Eita and Eshipala (2010), and Valadkhani, (2003). The research subjected the data to time series analysis, more specifically, autoregressive distributed lag model (ARDL) and Error Correction Model (ECM) analysis. The study used R-language,software for analysis. The results are then subjected to statistical interpretation from which conclusion and recommendations are drawn. Cointegration being the methodological framework of both ARDL and VECM, all the exploratory data analysis relevant to chapter 4, 5 and 6 are conducted here and referenced for use in the mentioned chapters. Also stationarity test, cointegration test, serial correlation test are done in this chapter for use in the next chapters.

Exploratory Data Analysis

3.2 Diagnostic Tests

3.2.1 Multicolinearity Test

This test is carried out to establish the relationship of the variables used. The study used the correlation matrix. In testing linear relationship between the explanatory variables, correlation matrix determined the strengths of variables in the model. It enabled the researcher to know which variable to drop from the equation.

3.2.2 Normality, Heteroscedasticity and Serial Correlation Tests

The study used Breusch-Pagan-Godfrey, Jarque-Bera statistic, Breusch-Godfrey LM Test, and cumulative sum test in testing for heteroscedasticity, normality, serial correlation and stability respectively. These are done to ensure that the coefficients of the estimate are efficient, consistent and reliable in making inference.

3.2.3 Multiple Correlation

The study conducted the correlation tests on the variable to establish existing relationship in such a way that the granger-causality foreseen in ARDL and VECM regression are considered valid. The relationship between the dependent variable Youth Unemployment (YUN) and each of the six independent variables (GDP, ED, FDI, PI, LR, POP) is scrutinized using R software correlation test at 1% CL. The test is done on the variables at level. The results are tabulated below.

Table 3.1: Simple correlation test results

R-code	**Result**
>cor(YUN,GDP)	0.294
>cor(YUN,ED)	0.44
>cor(YUN,FDI)	-0.094
>cor(YUN,PI)	0.48
>cor(YUN,LR)	0.22
>cor(YUN,POP)	0.27

In the result, the correlation of dependent variable is tested with each independent variable. From the table, it is evident that none of the variables explained more than 50% variation in youth unemployment. To get more robust results, study subsequently computed the multiple regression equation to verify correlation coefficient R-squared and adjusted R squared, and commented on the elasticity of each coefficient. By computing the multiple regression, it is possible to comment on the magnitude of adjusted R to determine extent of the variation in YUN being explained by independent variables.

3.2.4 Multiple Correlation test

The study further conducted multiple correlation test by performing multiple regression equation and observing adjusted R. The multiple regressions is done as follows: In (3.11) the standard errors are presented in parenthesis.

$$\begin{aligned} YUN =& 0.169736GDP_{\{0.036054\}} + 0.0388ED_{\{0.008756\}} + 0.001253FDI_{\{0.001431\}} + 0.014296PI_{\{0.006801\}} \\ &+ 0.02003LR_{\{0.009138028\}} - 0.114258POP_{\{0.053802\}} \end{aligned} \tag{3.1}$$

According to the test, the multiple R-squared is 0.8156 and the adjusted R squared is 0.7787. It revealed that 81.56% of variations in YUN is explained by the explanatory variables. The standard errors are in the parenthesis.

Table 3.2: Multiple correlation test results

Variable	**Estimate**	**Std. Error**	**t - value**	**p - value**
Intercept	14.808212	3.833368	3.86	0.000556 ***
GDP	0.169736	0.036054	4.708	5.31e-05 ***
ED	0.038800	0.008756	4.431	0.000115 ***
FDI	0.001253	0.0001431	8.76	0.0388176 **
PI	0.014296	0.006801	2.10	0.050627*
LR	0.02003	0.009138028	2.19	0.051986*
POP	-0.114248	0.053802	-2.123	0.042075 *

3.2.5 Stationarity Tests

Series are considered stationary when mean and variance are time invariant. The stationarity test is conducted to ensure that variable used did not exhibit spurious regression. The ADF test is used to assess stationarity status of each variable at level. The study checked whether the variables contained unit roots and stochastic trend. Establishing random walk using correlogram is not considered sufficient because it could be deceptive since a stochastic trend is a feature of time series with unit roots. So the study used Augmented Dickey-Fuller Test to test stationarity. It is noted that when the variables are not of the same stationary order, cointegration test is conducted in the next chapters to construct ARDL and VECM models. The results for stationary test are presented in table below.

Table 3.3: Stationarity tests results of variables

Variable	**p-value**	**Tau**	**1st diff Tau**		**Lag**
			p-value	**Tau**	**Order**
YUN	0.99	3.0023	0.01	-4.9328	3
GDP	0.04211	-3.6641	NA	NA	3
ED	0.02464	-3.9058	NA	NA	3
FDI	0.01	-5.7999	NA	NA	3
PI	0.6188	-1.8831	0.01	-5.573	3
LR	0.6146	-1.894	0.01	-5.0686	3

The results presented in the table indicate that the variables are I(0) and I(1) stationary. These results are necessary for the study to proceed and carry out cointegration analysis, ARDL and reparameterize to VECM.

3.2.6 Time plot

The time plots for all variables in level are generated to complement ADF stationarity test as illustrated below:

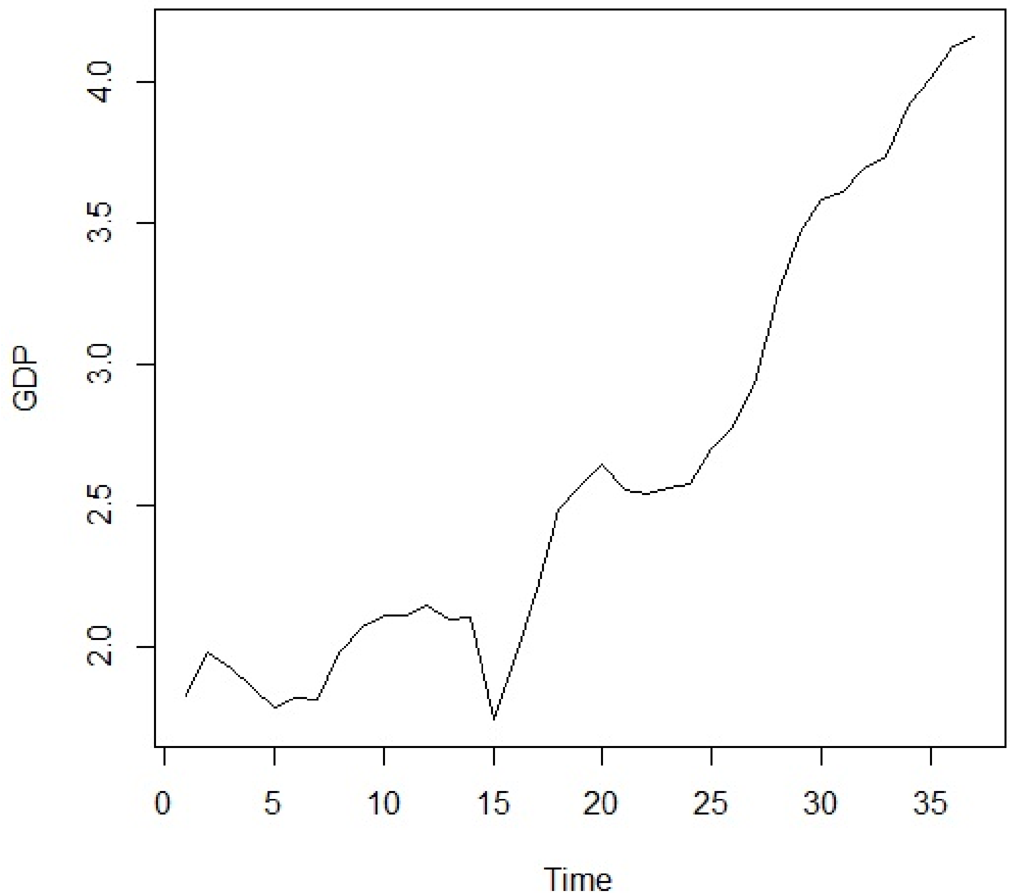

Figure 3.1: Time Plot of GDP

Time plot of GDP in billion US dollars verses time in years

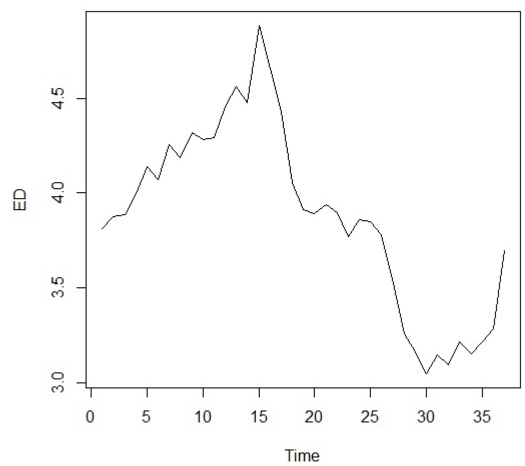

Figure 3.2: Time Plot of ED

Time plot of external debt in billions US dollars verses time in years

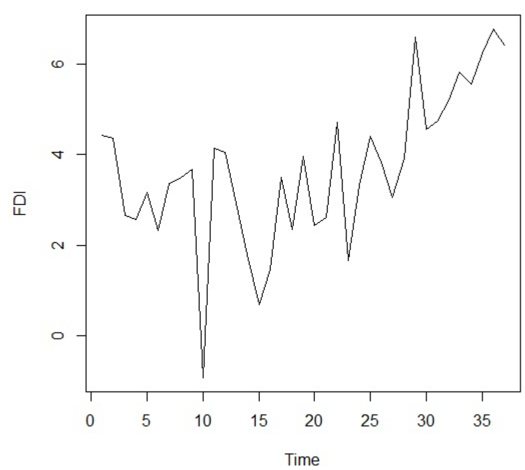

Figure 3.3: Time plot of FDI

Time plot of FDI in billions US dollars verses time in years

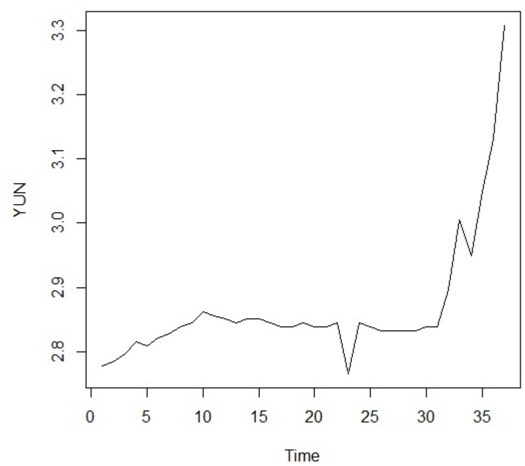

Figure 3.4: Time Plot of YUN

Time plot of YUN in millions verses time in years

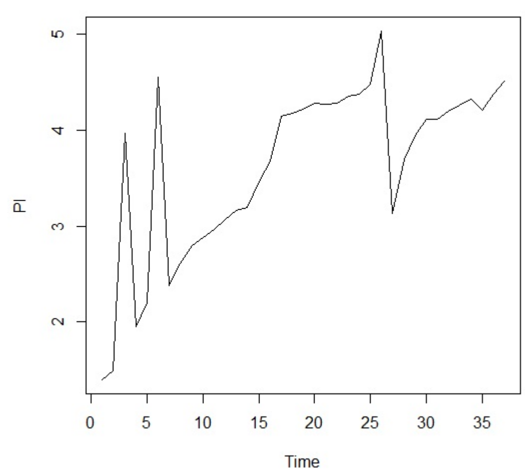

Figure 3.5: Time plot of PI

Time plot of PI in billions US dollars verses time in years

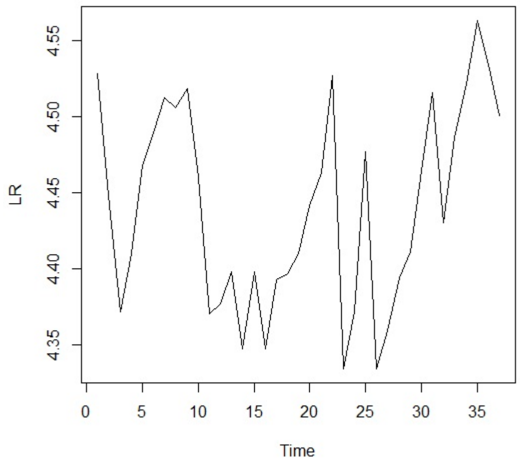

Figure 3.6: Time plot of LR

Time plot of literacy rate in percentage verses time in years

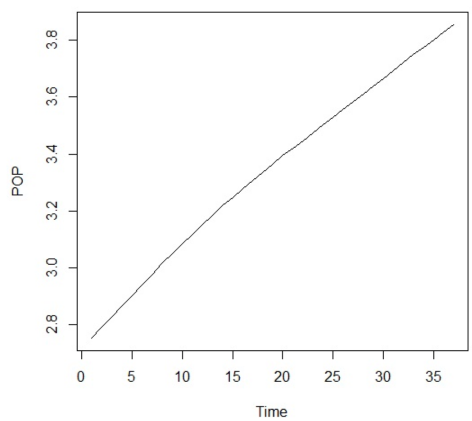

Figure 3.7: Time plot of POP

Time plot of population in tens of millions verses time in years

From the time plots, series considered are $I(0)$ and $I(1)$ at level. Results from stationarity test indicated that the variables would not drift far from each other over time.

3.3 Choosing lag length

In time series data used in the study, it is (and is always) possible to choose lag length based on results of Dickey-Fuller stationarity test. The ADF test in this study indicated lag order as indicated in table 3.3. However, the study went further to conduct independent lag selection as shown in the table below. In table 3.3, above stationarity of the variables is established at lag of order 3. When the data VAR is applied to the variables to determine the most appropriate lag, the results are obtained as shown on the table below. The study then chose FPE suggesting 3 lags in line with stationarity results.

Table 3.4: Lag selection criteria results

Lag	**AIC**	**HQ**	**SC**	**FPE**
1	-3.633133e+01	-3.553215e+01	-3.364367e+01	1.893163e-16
2	-3.837335e+01	-3.687489e+01	-3.333399e+01	5.824986e-17
3	NaN	NaN	NaN	-1.300749e-47 *

Serial Correlation Test

The independence of observations and efficiency of labour market is tested using serial correlation test which is done using Peirce-Box test in R. It is also a diagnostic test for residuals of OLS.

Table 3.5: serial correlation test results

Variable	chi-squared stat	p-value
GDP	142.16	0.2475
ED	138.05	0.2565
FDI	39.219	0.1566
YUN	30.138	0.8133
PI	51.501	0.1287
LR	21.126	0.02023
POP	155.99	0.0475

H_0: NO serial correlation up to lag 10.

vs

H_1: The is serial correlation up to lag 10.

According to this test, there is no serial correlation up to lag 10 in GDP,ED, FDI, YUN and PI whose p-values are greater than 0.05%. However, literacy rates and population had serial correlation up to lag 10 since their p-values are less than 0.05%.

3.4 Hypotheses

H_1:ARDL is statistically superior to VECM and more suitable for statistical modeling of youth unemployment.

H_2:ARDL is NOT statistically superior and not more suitable for modeling youth unemployment than VECM.

H_3: GDP, FDI, POP, PI, ED, and LR have +ve short run effects on youth unemployment.

H_4: GDP, FDI, POP, PI, ED, and LR have -ve long run effects on youth unemployment.

3.4.1 Cointegration Analysis

If Dickey-Fuller test reveals nonstationarity we proceed to determine cointegration between the variables. Nonstationary time series variables should not be used in regression models, to avoid the problem of spurious regression. However, there is an exception to this rule: If Y_t and X_t are nonstationary variables, then we expect their difference, or any linear combination, such as

$$e_t = Y_t - \beta_1 - \beta_2 X_t \tag{3.2}$$

where e_t is error term, β_1 and β_2 are coefficients.

I(0) denotes stationary at level (without differencing) and I(1) denotes stationary after first difference.

In this case, Y_t and X_t are said to be cointegrated.

The basic idea behind cointegration analysis is that a group of nonstationary variables might individually wander extensively in such a way that they do not drift too far apart from one another, given that the difference is stationary. That is, individually they are time series with unit roots but a particular linear combination of them is stationary. We outline the definition and estimation for cointegrated vectors as follows:

An $(mx1)$ vector time series is defined as cointegrated in (c, b) order if each of the series is $I(d)$ process, that is, a nonstationary process with d unit roots, whereas a certain linear combination of the series αY_t is an $I(c-b)$ process for some nonzero $(mx1)$ constant vector,α. The vector Y_t considered in this study is (7×35) seven variables with 35 observations: y_{t1}, y_{t2}, y_{t3}, y_{t4}, y_{t5}, y_{t6} & y_{t7} representing the variables used. The Vector Error correction Model of order $I(1)$ is given by, equation 3.3.

$$C(L)\Delta \ln Y_t = \theta + \gamma Y_{t-1} + e_t \tag{3.3}$$

Where C(L) is a 7 by 35 matrix with p is maximum lag, Δ is the first difference, θ is vector of intercept, γ is a 7×35 matrix, and e_t is the error term vector.

Therefore if the regression of two or more series which are individually integrated yield residuals with lower order of integration; they are said to be cointegrated. There is the special case of cointegration in which the linear combination of series integrated of the same order is stationary.

Most of the cointegration tests are based on this special case in which series integrated of order one yields a linear combination which is stationary.

3.5 Phillips-Ouliaris Test

The cointegration test helped to check that variables do not drift apart over time i.e have equilibrium relationship. The Phillips-Ouliaris test helps to establish the order of cointegration between two variables in turns.

Table 3.6: Philip-Ouliaris Cointegration result of paired variables

R-CODES	VARIABLES	P-VALUE	Po-demeaned	Differential Order
po.test(diff(log(cbind(YUN,GDP))))	YUN and GDP	0.02140	-24.8909	1
po.test(diff(log(cbind(YUN,ED))))	YUN and ED	0.01	-34.185	1
po.test(diff(log(cbind(YUN,FDI))))	YUN and FDI	0.01	-30.4839	1
po.test(diff(log(cbind(YUN,PI))))	YUN and PI	0.01	-31.1498	1
po.test(diff(log(cbind(YUN,LR))))	YUN and LR	0.01	-29.2731	1
po.test(diff(log(cbind(YUN,POP))))	YUN and POP	0.01	-33.1063	1
po.test(diff(log(cbind(GDP,ED))))	GDP and ED	0.04044	-21.7589	1
po.test(diff(log(cbind(GDP,PI))))	GDP and PI	0.02022	-25.2477	1
po.test(log(cbind(GDP,LR))	GDP and LR	0.02079	-25.0744	1
po.test(diff(log(cbind(GDP,POP))	GDP and POP	0.01094	-28.0382	1
po.test(diff(log(cbind(ED,PI))))	ED and PI	0.01273	-27.5002	1
po.test(diff(log(cbind(ED,LR))))	ED and LR	0.01832	-25.8183	1
po.test(diff(log(cbind(ED,POP))))	ED and POP	0.01	-30.7157	1
po.test(diff(log(cbind(PI,LR))))	PI and LR	0.01	-51.4342	1
po.test(log(cbind(PI,POP)))	PI and POP	0.01	-35.1818	0
po.test(diff(log(cbind(LR,POP))))	LR and POP	0.01	-42.1315	1
Po.test(log(cbind(GDP,FDI)))	GDP and FDI	0.0446	-21.2072	0
Po.test(log(cbind(ED,FDI)))	ED and FDI	0.05377	-20.1868	0
Po.test(log(cbind(FDI,PI)))	FDI and PI	0.04558	-21.0765	0
Po.test(log(cbind(FDI,LR)))	FDI and LR	0.0344	-22.5618	0
Po.test(log(cbind(FDI,POP)))	FDI and POP	0.01542	-26.691	0
po.test(diff(log(cbind(YUN,GDP))))	YUN and GDP	0.02140	-24.8909	1
po.test(diff(log(cbind(YUN,ED))))	YUN and ED	0.01	-34.185	1
po.test(diff(log(cbind(YUN,FDI))))	YUN and FDI	0.01	-30.4839	1
po.test(diff(log(cbind(YUN,PI))))	YUN and PI	0.01	-31.1498	1
po.test(diff(log(cbind(YUN,LR))))	YUN and LR	0.01	-29.2731	1
po.test(diff(log(cbind(YUN,POP))))	YUN and POP	0.01	-33.1063	1
po.test(diff(log(cbind(GDP,ED))))	GDP and ED	0.04044	-21.7589	1
po.test(diff(log(cbind(GDP,PI))))	GDP and PI	0.02022	-25.2477	1
po.test(log(cbind(GDP,LR))	GDP and LR	0.02079	-25.0744	1
po.test(diff(log(cbind(GDP,POP))	GDP and POP	0.01094	-28.0382	1
po.test(diff(log(cbind(ED,PI))))	ED and PI	0.01273	-27.5002	1
po.test(diff(log(cbind(ED,LR))))	ED and LR	0.01832	-25.8183	1
po.test(diff(log(cbind(ED,POP))))	ED and POP	0.01	-30.7157	1
po.test(diff(log(cbind(PI,LR))))	PI and LR	0.01	-51.4342	1
po.test(log(cbind(PI,POP)))	PI and POP	0.01	-35.1818	0
po.test(diff(log(cbind(LR,POP))))	LR and POP	0.01	-42.1315	1
Po.test(log(cbind(GDP,FDI)))	GDP and FDI	0.0446	-21.2072	0
Po.test(log(cbind(ED,FDI)))	ED and FDI	0.05377	-20.1868	0
Po.test(log(cbind(FDI,PI)))	FDI and PI	0.04558	-21.0765	0
Po.test(log(cbind(FDI,LR)))	FDI and LR	0.0344	-22.5618	0
Po.test(log(cbind(FDI,POP)))	FDI and POP	0.01542	-26.6911	0

Explanation

The test revealed that all the variables are cointegrated at order $I(0)$ and $I(1)$ as shown in the Table 3.6 above. Out of 21 relations, 6 are $I(0)$ while 15 are $I(1)$ The Phillip-Ouliaris did not have the ability to conduct cointegration test of more than two variables. The test is further limited in identifying coefficients of cointegration necessary for cointegrating vectors and equations. With these limitations the study conducted Johansen Cointegration test.

3.6 Johansen Cointegration Rank Test

The study noted that Johansen (1988) cointegration rank test had the ability to perform cointegration test for more than two variables. In table 3.7 below, cointegration test is conducted for six variables at 95% confidence level. The Johansen test used two test statistics: Trace test and maximum eigen value statistics. In both cases, we reject the null hypothesis at the first instance when the critical value is greater than test statistics. See table 3.7 for results.

These are the eigenvalues: 8.601594e-01, 8.137679e-01, 6.928169e-01, 3.964401e-01, 3.675151e-01,
2.562996e-01, 1.461890e-01, -9.150666e-17.

The largest eigen value is 8.798487e-01

Table 3.7: Values of test statistic and critical values of test:

R (cointegration rank)	**Test Stats**	**10%**	**5%**	**1%**
$r \leq 6$	5.37	7.52	9.24	12.97
$r \leq 5$	10.07	13.75	15.67	20.20
$r \leq 4$	15.58	19.77	22.00	26.81
$r \leq 3$	17.17	25.56	28.14	33.24
$r \leq 2$	40.13	31.66	34.40	39.79
$r \leq 1$	57.15	37.45	40.30	46.82
$r \leq 0$	66.89	43.25	46.45	51.91

$H_0 : r \leq 3$, the test statistics is greater than critical value at 95% C.L, that is, $28.14 > 17$. The study rejected the null hypothesis and concluded that there are $r = 3$ cointegrating relations at 5% confidence level. The test statistics revealed the existence of three cointegrating relations combining the seven series. This imply existence of long-term equilibrium.

3.7 Identification of Cointegration Relationship

In Johansen-(2001)-cointegration test with constant, the eigenvalues (lambda) are ordered according to columns of variables. The eigenvalues are normalized to the first column. The maximum eigenvalue corresponded to the column with a cointegration relationship. These are the eigenvalues:

8.601594e-01, 8.137679e-01, 6.928169e-01, 3.964401e-01, 3.675151e-01, 2.562996e-01, 1.461890e-01, -9.150666e-17

The largest eigen value is 8.601594e-01 corresponding to the GDP column. Johansen Test recommends that the largest eigen values correspond to columns with cointegrating vectors. However, the columns of the variables (as cointegrating relations) are tested for stationarity using ADF test. The stationary equations formed the Cointegrating vectors.

These are the Cointegrating relations tested for stationarity: The columns in bold in table 3.8 are found to be stationary and formed the 3 cointegrating relationship.

Table 3.8: Columns of cointegration relationships :

Variable	GDP	ED	FDI	YUN	PI	LR	POP
GDP	**1.000000**	**1.0000**	**1.0**	1.00	1.0000	1.0000	1.00000
ED	**-0.68979268**	**-0.7225719**	**0.2363211**	2.3730471	0.21836910	0.79938274	0.2874472
FDI	**-0.8050538**	**-0.3684427**	**-0.0130395**	0.3261325	-0.06307162	0.02078336	-0.1283847
YUN	**1.68667026**	**8.0728399**	**-1.0932464**	-31.3622929	0.37462436	-2.37548733	-3.0544291
PI	**-0.03626714**	**1.404787**	**0.3033717**	-1.2633971	-0.39335741	0.16457778	0.4907323
LR	**-1.11845144**	**-8.3423590**	**-2.2610056**	-18.5692818	0.72578376	0.22781790	2.8297209
POP	**-2.20928252**	**2.3409439**	**-3.0604127**	8.4011665	-1.30575355	-1.14352444	-2.1788867
CONST	**13.40430269**	**-13.5878323**	**20.5798943**	118.3169565	-0.96791357	2.71781603	-2.8024775

In Johansen cointegration test, eigenvectors, normalised to first column provides the cointegrating relationships. The columns in bold formed the cointegration vectors. To identify the cointegration of the vectors, the study tested the cointegrating vectors for stationarity using ADF test as shown below.

3.7.1 Identified Cointegrating equations

From the results above, GDP, ED, and FDI columns formed cointegrated series as depicted in the equations below:

The Cointegrating equation for GDP is:

$$\begin{aligned} GDP_t =& 1GDP_{t-1} - 0.68979268ED_t - 0.80505386FDI_t + 1.68667026YUN_t - 0.03626714PI_t \\ & - 1.11845144LR_t - 2.20928252POP_t + 13.40430269 \end{aligned} \tag{3.4}$$

The ADF test found GDP to be stationary with a p-value 0.02089.

The elements of vector GDP are:

2.524116, 2.698738, 3.951576, 3.851734, 3.041053, 3.620336, 2.634182, 2.680132, 2.448953, 6.221083, 2.164391, 2.059987, 2.727771, 3.698701, 3.818731, 3.549144, 2.156470, 3.554189, 2.365964, 3.573837, 3.215634, 1.433391, 4.008154, 2.658323, 1.726258, 2.379012, 3.302157, 2.999891, 1.035730, 2.756817, 2.451243, 2.341810, 1.853017, 2.098202, 1.653895, 1.398835, 1.700126

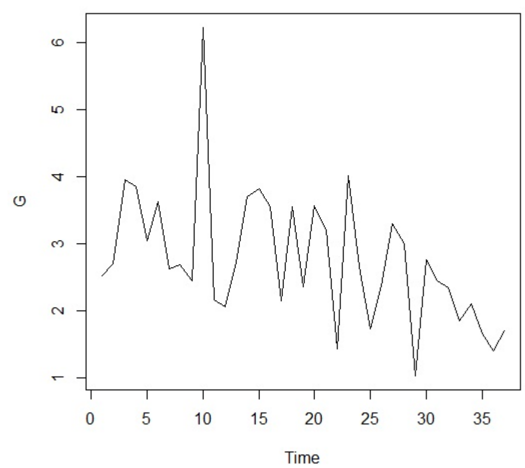

Figure 3.8: Time plot of GDP cointegrating relation of GDP column

Time plot of GDP in billions of US dollars verses time in years

The Cointegrating relation is:

$$
\begin{aligned}
FDI_t =& 1GDP_t + 0.2363211ED_t - 0.0130395FDI_t - 1.0932464YUN_t + 0.3033717PI_t \\
&- 2.2610056LR_t - 3.0604127POP_t + 20.579843
\end{aligned}
\tag{3.5}
$$

The equation is found to be stationary with a p-value of 0.01

Elements of FDI time series data: 1.973719, 2.227484, 2.995824, 2.115802, 1.917993, 2.485004, 1.674090, 1.772463, 1.818279, 1.934157, 2.005040, 1.996232, 1.882481, 1.889753, 1.529902, 1.773386, 1.882343, 2.007700, 1.929557, 1.876347, 1.664351, 1.385123, 1.875700, 1.638691, 1.465568, 1.940326, 1.339290, 1.588009, 1.689717, 1.650812, 1.502314, 1.635726, 1.387293, 1.482823, 1.255313, 1.319256, 1.305664

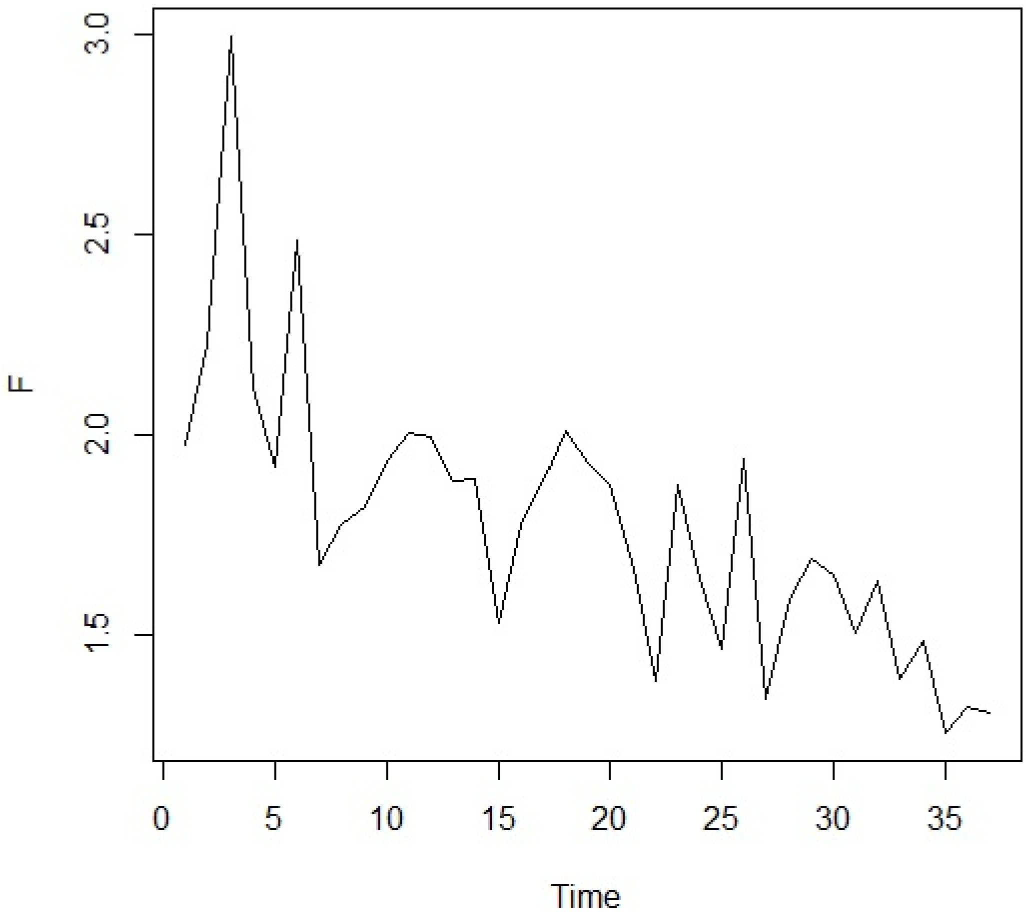

Figure 3.9: Time plot of cointegrating relation of FDI column

Time plot of FDI in billions of US dollars verses time in years

The P-value for ED cointegrating equation is 0.05

$$
\begin{aligned}
ED_t =& 1GDP_t - 0.7225719ED_t - 0.3684427FDI_t + 8.0728399YUN_t + 1.4047877PI_t \\
&- 8.342359LR_t + 2.3409439POP_t - 13.5878323
\end{aligned}
\tag{3.6}
$$

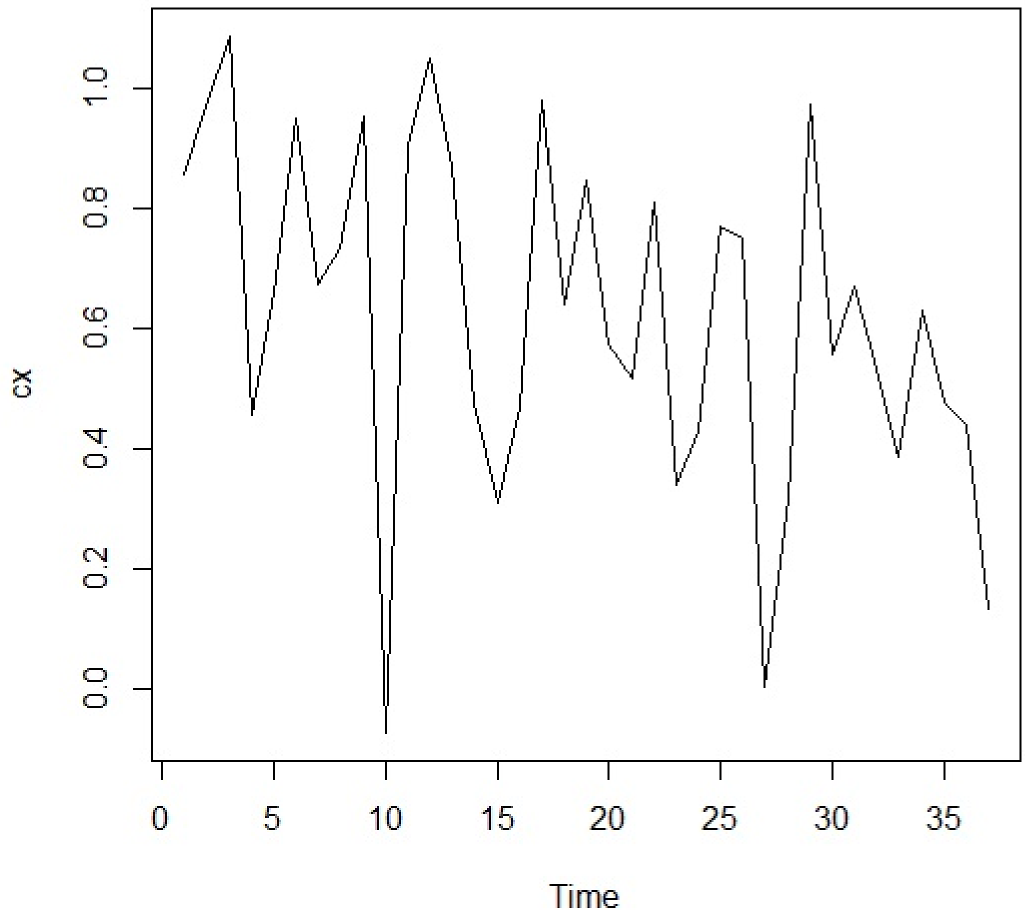

Figure 3.10: Time Plot of Cointegrating relations of ED column

A time plot of External debt in billions of US dollars and x axis time in years

The ADF stationarity test is conducted on all the Cointegrating vectors and a constant from Johansen test. The result in the table below show that only cointegrating relation corresponding to GPD, ED and FDI are stationary with a p-value of 0.0209 and 0.01 respectively.

The study tested all the cointegrating relation columns for stationarity and presented the results in table 3.9. From the p-value presented in table 3.9 GDP, ED and FDI columns are found to be stationary.

Table 3.9: ADF test results for cointegrating relations:

Cointegrating Relations Column	Eigen Value(Lambda)	P-value
GDP	8.601594e-01	0.0209
ED	8.137679e-01	0.05
FDI	6.928169e-01	0.01
YUN	3.964401e-01	0.09
PI	3.675151e-01	0.08793
LR	2.562996e-01	0.4754
POP	1.461890e-01	0.5683
Constant	9.150666e-17	0.99

3.7.2 Linear cointegration Equation

The study established existence of linear cointegration among the variables used, therefore the research proceeded to estimate linear cointegration equation because the cointegrated variables would not drift far away from each other. The variables exhibit a characteristic of co-movement of the macroeconomic variables over time. The linear cointegration equation of GDP, ED, FDI, YUN, PI, LR and POP is estimated using the equation below

$$\ln YUN_t = \beta_0 + \beta_1 \Delta \ln GDP_t + \beta_2 \Delta \ln POP_t + \beta_3 \Delta \ln FDI_t + \beta_4 \Delta \ln ED_t + \beta_5 \Delta \ln PI_t + \beta_6 \Delta \ln LR_t + \epsilon_t \tag{3.7}$$

Table 3.10: Results of estimated linear cointegration equation:

Variable	**Estimate**	**Std. Error**	**t value**	**P-value**
Intercept	1.322215	0.634925	2.082	0.045925 * *
GDP	0.356204	0.044585	7.989	6.44e-09 ***
ED	0.269000	0.029884	9.001	5.00e-10 ***
FDI	0.002441	0.006745	0.362	0.719919
PI	0.011790	0.012108	0.974	0.337989
LR	0.154216	0.129145	1.194	0.241783
POP	-0.350383	0.084511	4.146	0.000255 ***

The estimated linear cointegration equation is:

$$
\begin{aligned}
YUN_t = {} & 1.3222151_{\{0.634925\}} + 0.356204_{\{0.044585\}}GDP_t + 0.269_{\{0.029884\}}ED_t + 0.002441_{\{0.006745\}}FDI_t + 0.011 \\
& + 0.154216_{\{0.129145\}}LR_t - 0.350833_{\{0.084511\}}POP_t
\end{aligned}
\tag{3.8}
$$

From the regression in table 3.10: Adjusted R-squared is 0.8407 6. The F-statistics is 32.66 on 6 while the p-value is 7.322e-12 or 0.001. The results indicated that 84.076% of variation in youth unemployment is explained by GDP, external debt, foreign direct investment, private investment, literacy level, and the youth population.

The computed F-value: 32.66 and F-critical value: 3.29. The null hypothesis is rejected and the equation is declared statistically significant.

4 AUTOREGRESSIVE DISTRIBUTED LAG (ARDL)

4.1 Introduction

The estimated ARDL is subjected to t-type and F-type test as recommended by (Uwe and Wolters, 2006). The diagnostics test including initial tests of data set used in development of the model. Lag length is chosen, stationarity test, multiple correlation test, serial correlation test and CUSUMSQ test are conducted. The diagnostic tests conducted are residual versus fitted, normality of fitted residual, scale and location of residual, residual and leverage. These tests are conducted to establish statistical superiority of the model. The study postulated that ARDL is superior to VECM, thus diagnostic tests.
ARDL is estimated using OLS. CUSUM and CUSUMSQ test stability of ARDL. The general form of ARDL model with both dependent and independent variables lagged up to p and q respectively is given by;

The form of the model is given below,

$$YUN = f(GDP, FDI, PI, ED, LR, POP) \tag{4.1}$$

The general form of ARDL model with both dependent and independent variables lagged up to p and q respectively is described as

$$Y_t = \delta + \sum_{i=1}^{p} \theta_i Y_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j} + \epsilon_t \tag{4.2}$$

where, Y_t is the dependent variable, X_t independent variable δ is the impact multiplier, θ_i is the distributed lag weight of Y_t, ϵ_t is the error term, p is the lag length of Y_t, q is the lag length of X.

4.1.1 Assumptions of ARDL

(i) $t = q + 1, \cdots, T$, Y_t and X_t are stationary random variables

(ii) ϵ_t is independent of current past and future values of X_t.

(iii) $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = \sigma^2$, $cov\ (\epsilon_t, \epsilon_s) = 0, t \neq s$, $\epsilon_t \sim (0, \sigma^2)$

4.1.2 Choosing the length of lags p and q

The study uses final prediction error (FPE) as recommended by Akaike (1971), for vector AR(m) lag selection. The FPE equation is given in 4.3

$$FPE = \det\left\{1 + \left(\frac{qk}{T}\right)\right\} SSE \tag{4.3}$$

Our equation of interest is given by:

The ARDL being autoregressive (AR), has previous values of YUN_t among explanatory variables.

$$\ln YUN_t = \beta_0 + \beta_1 YUN_{t-1} + \beta_2 \ln GDP_t + \beta_3 \ln POP_t + \beta_4 \ln FDI_t + \beta_5 \ln ED_t + \beta_6 \ln PI_t + \beta_7 \ln LR_t + \epsilon_t \tag{4.4}$$

Incase every variable is lagged up to $t-p$ for independent and $t-q$ for dependent variable.

$$\begin{aligned}\ln YUN_t = \beta_0 &+ \sum_{i=1}^{q} \beta_1 \Delta \ln YUN_{t-i} + \sum_{j=1}^{p} \beta_2 \Delta \ln GDP_{t-j} + \sum_{j=1}^{p} \beta_3 \Delta \ln POP_{t-j} + \sum_{j=1}^{p} \beta_4 \Delta \ln FDI_{t-j} \\ &+ \sum_{j=1}^{p} \beta_5 \Delta \ln ED_{t-j} + \sum_{j=1}^{p} \beta_6 \Delta \ln PI_{t-j} + \sum_{j=1}^{p} \beta_7 \Delta \ln LR_{t-j} + \epsilon_t\end{aligned} \tag{4.5}$$

4.1.3 Advantages of ARDL Model

(i) ARDL captures dynamic effects of lagged x_s and y_s

(ii) Serial correlation can be eliminated if sufficient number of lags of x_s and y_s are included.

(iii) ARDL can be transformed into infinite distributed lags with lagged x_s only which extend into infinite past.

4.1.4 Estimation of Parameters of ARDL

An ARDL(k, n) model where k are parameters of lagged coefficients and n is the number of coefficients is estimated as described by Sargan (1964) as shown in the equation below. This is in agreement with standard asymptotic normal theory, Hendry, Pagan and Sargan (1984) and Wickens and Bruesch (1988) [42].

$$K(\alpha^{2'} - \alpha^{3'}) = O_p\left(\frac{1}{T}\right) \tag{4.6}$$

where, $O_p\left(\frac{1}{T}\right)$ is $A^{2'} - A^{3'}$,

$A^{2'}$ are coefficients of values of parameters,

$A^{3'}$ are coefficients of lagged parameters,

$\alpha^{\mathbf{2'}}$ is a vector of a matrix $A^{2'}$,

$\alpha^{\mathbf{3'}}$ is a vector of a matrix $A^{3'}$,

T is total time

and

K is a square matrix of order equal to the number of unknown lagged coefficients in the equation.

K is viewed as obtained by taking the Frobenius matrix product and picking only those rows and columns corresponding to unknown coefficients. K is partitioned as shown below,

$$K = \begin{pmatrix} K_{11} & K_{12} & \cdots & K_{1n} \\ K_{21} & K_{22} & \cdots & K_{2n} \\ K_{31} & K_{32} & \cdots & K_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & K_{nn} \end{pmatrix}$$

4.1.5 Inference on long run and short run coefficients

Lag operator notation:

Let

$$L^p Y_t = Y_{t-p} \tag{4.7}$$

where $p = 1, 2, ., $ and L is the lag operator.
Re-write ARDL model given by equation (4.2) in terms of lag operator and factoring out X_t and Y_t,

$$(1 - \theta_1 L^1 - \theta_2 L^2 \theta_p L^p) Y_t = \delta + (\delta_0 + \delta_1 L + \delta_2 L^2 + \delta_p L^q) X_t + \epsilon_t \tag{4.8}$$

From Okun's Law (1962):

$$\Delta YUN_t = \delta + \theta_1 \Delta YUN_{t-1} + \delta_0 G_t + \delta_1 G_{t-1} + \epsilon_t \tag{4.9}$$

Rewriting and factoring the like terms,

$$(1 - \theta_1 L) \Delta YUN_t = \delta + (\delta_0 + \delta_1 L + \delta_2 L^2) G_t + \epsilon_t \tag{4.10}$$

and

$$\Delta YUN_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 \beta_{t-2} + e_t \tag{4.11}$$

solving for β_s interms of θ_1 and δ_s
From (4.13)

$$\Delta YUN_t = (1 - \theta_1 L)^{-1} \delta + (1 - \theta_1 L)^{-1} (\delta_0 + \delta_1 L + \delta_2 L^2) G_t + (1 - \theta_1 L)^{-1} \epsilon_t \tag{4.12}$$

and from (4.14)

$$\Delta YUN_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 \beta_{t-2} + e_t \tag{4.13}$$

But
$(1 - \theta_1 L)^{-1} = 1 + (\theta_1 L) + (\theta_1 L)^2 + (\theta_1 L)^3$
since

$(1-\theta_1 L)^{-1} = 1 + (\theta_1 L) + (\theta_1 L)^2 + (\theta_1 L)^3 + \cdots$

then

$$\begin{aligned} (1-\theta_1 L)^{-1}(\delta_0 + \delta_1 L + \delta_2 L^2) &= (1 + (\theta_1 L) + (\theta_1 L)^2 + (\theta_1 L)^3 + \cdots)(\delta_0 + \delta_1 L + \delta_2 L^2)) \\ &= \delta_0 + (\theta_1\delta_0 + \delta_1)L + (\delta_0{\theta_1}^2 + \theta_1\delta_1 + \delta_2)L^2 \end{aligned} \tag{4.14}$$

From (4.15), (4.16) and (4.17)

$\alpha = (1-\theta_1 L)^{-1}\delta$

$\beta_0 = \delta_0$

$e_t = (1-\theta_1 L)^{-1}\epsilon_t$

$\beta_1 = \delta_0\theta_1 + \delta_1$

$\beta_2 = \delta_0{\theta_1}^2 + \delta_1\theta_1 + \delta_2 = \delta_2 + (\delta_0\theta_1 + \delta_1) = \delta_2 + \theta_1\beta_1$

To determine the effects of change in individual independent variable on dependent variable Multiplier analysis is used. From equation (4.2),

$$Y_t = \delta + \alpha Y_{t-i} + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \epsilon_t \tag{4.15}$$

The multipliers are given by,

$\beta_s = \dfrac{\partial Y_t}{\partial x_{t-s}} = s$ period delay multiplier

$\displaystyle\sum_{j=0}^{q} \beta_j$, is the total period interim multiplier

Or

From (4.15) above we have:

$$C(L)Y_t = \delta + \beta(L)X_t + \epsilon_t \tag{4.16}$$

where

$$C(L) = 1 - \theta_1 L - \theta_2 L^2 - \theta_p L^p \tag{4.17}$$

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_q L^q \tag{4.18}$$

Equations (4.17) and (4.18) are coefficients expressed as lag orders of dependent and independent variables of (4.16).
The equilibrium multiplier (long- run effects of a change in x) in the ARDL model generally is,

$$\sum_{i=0}^{\infty} \alpha_i = \frac{B(1)}{C(1)} = A(1) = \frac{\sum_j^q \beta_j}{1 - \sum_i^p \theta_i} \tag{4.19}$$

4.1.6 Stationary Test

From previous studies the research found that a series is stationary when its mean and variance are time invariant, i.e integrated of order zero. Also, nonstationary series have infinite variance asymptotically and therefore any inference made will be invalid due to both spurious regression problems.

The study uses ADF test to check whether X_t is stationary at level, i.e $I(0)$. In a case where the AR process is of a higher order, say AR(k) the series is differenced (k) times and the test is done on k^{th} order of the variable. This is done like in variables at level, that is, I(0).

The test is generally based on the AR(1) process:

$$Y_t = \rho Y_{t-1} + v_t \tag{4.20}$$

is stationary when $|\rho| < 1$

When $\rho = 1$; it is nonstationary random walk process $Y_t = Y_{t-1} + v_t$
Hence, to test for stationarity we examine the value of ρ, i.e. we test whether $\rho = 1$ or $\rho < 1$ the unit root test.To formalize this procedure, we deal with AR(2) model:

$$Y_t = \rho Y_{t-1} + v_t$$

Where $v_t \sim N(0, \sigma^2)$ is *iid*.
We can test for nonstationarity by testing $H_0 : \rho = 1$

$$H_A : \rho < 1.$$

We subtract Y_{t-1} from both sides to obtain:

$$\begin{aligned} Y_t &= \rho Y_{t-1} + v_t \\ Y_t - Y_{t-1} &= (\rho - 1)Y_{t-1} + v_t \\ \Delta Y_t &= \gamma Y_{t-1} + v_t \end{aligned}$$

Where $\gamma = \rho - 1$ and, $\Delta Y_t = Y_t - Y_{t-1}$.

The hypothesis can then be written in terms of γ or ρ as follows:

$$H_0 : \rho = 1 \Longleftrightarrow H_0 : \gamma = 0$$

$$H_1 : \rho < 1 \Longleftrightarrow H_1 : \gamma < 1$$

In particular, we estimate the following three ADF models:

(i) A random walk (no trend and no constant-drift)

$$\Delta Y_t = \omega Y_{t-1} + e_t \tag{4.21}$$

(ii) ADF with intercept but no trend

$$\Delta Y_t = \alpha_t + \omega Y_{t-1} + e_t \tag{4.22}$$

(iii) ADF with intercept and trend

$$\Delta Y_{t-1} = \alpha_t + \delta_t + \omega Y_{t-1} + e_t \tag{4.23}$$

We test: H_0: *Non stationary (unit root)*

H_1: *Stationary*

4.2 Estimated ARDL Equation

From the lag selection criterion above, FPE allowed the study to choose three lags of each variable with exception of one lag for dependent variable as required in ARDL in econometric models. All preliminary data test and diagnostic tests are conducted on the equation of interest. Thus the equation to be estimated was:

4.3 ARDL Equation

$$
\begin{aligned}
YUN = & \alpha YUN_{t-1} + \beta_1 GDP_t + \beta_2 GDP_{t-1} + \beta_3 GDP_{t-2} + \beta_4 GDP_{t-3} + \gamma_1 ED_t + \gamma_2 ED_{t-1} \\
& + \gamma_3 ED_{t-2} + \gamma_2 ED_{t-3} + \theta_1 FDI_t + \theta_2 FDI_{t-1} + \theta_3 FDI_{t-2} + \mu_1 PI_t + \mu_2 PI_{t-1} + \mu_3 PI_{t-2} \\
& + \mu_4 PI_{t-3} + \pi_1 LR_t + \pi_2 LR_{t-1} + \pi_3 LR_{t-2} + \pi_4 LR_{t-3} + \phi_1 POP_t + \phi_2 POP_{t-1} + \phi_3 POP_{t-2} \\
& + \phi_4 POP_{t-3} + \epsilon_t
\end{aligned}
\tag{4.24}
$$

The multiple correlation test, the stationarity test and lag length conducted in exploratory data analysis allowed the research to estimate the ARDL equation (4.24) above. Also the Phillip Oulliaris cointegration test and Johansen cointegration test are additional supplementary data tests before estimating ARDL.

Every independent variable is lagged 3 times while the dependent variable is lagged once. The parameters are found as follows:

Table 4.1: Results of estimated ARDL

Variable	Estimate	Std. Error	t - value	P - value
Intercept	3.1881443	1.1867684	2.731	0.06930 **
Lag(YUN,-1)	0.6580734	0.2268072	2.90	0.1090 .
GDP	-0.2079220	0.040073	-5.19	0.0406 ***
Lag(GDP,-1)	0.3848136	0.0917367	4.19	0.0796 *
Lag(GDP,-2)	-0.2052705	0.0673422	-3.05	0.1548 *
Lag(GDP,-3)	0.1208412	0.0257892	-4.69	0.03694***
ED	-0.073030	0.0149475	-4.89	0.00235 ***
Lag(ED,-1)	0.0214684	0.0110151	1.95	0.1369 .
Lag (ED,-2)	-0.009116	0.00325426	-2.91	0.09087*
Lag(ED,-3)	0.0094806	0.0125552	0.76	0.4718
FDI	0.0020017	0.0010008	1.99	0.3159 .
Lag(FDI,-1)	0.0025032	0.0012111	2.07	0.1787*
Lag(FDI,-2)	0.0006147	0.001649	0.37	0.7234
Lag(FDI,-3)	0.0027486	0.0010519	2.61	0.1145 *
PI	0.0107310	0.0054604	1.97	0.1946 .
Lag(PI,-1)	0.0069095	0.0067884	1.018	0.3385
Lag(PI,-2)	0.0144567	0.0059603	2.403	0.0539 *
Lag(PI,-3)	0.0083734	0.0073051	1.146	0.2848
LR	0.0116937	0.003686	3.17	0.07589 *
Lag(LR,-1)	-0.0892691	0.0373312	-2.391	0.0438 **
Lag(LR,-2)	0.0557574	0.0456325	1.222	0.2565
Lag(LR,-3)	0.0169212	0.0389834	0.434	0.6757
POP	2.0186805	2.0022926	1.0813	0.4396
Lag(POP,-1)	-0.2590455	0.036858	-7.03	0.9457**
Lag(POP,-2)	2.2498754	3.3935272	0.663	0.5260
Lag(POP,-3)	-4.3093119	4.1772794	-1.032	0.3324

Where significance codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Adjusted R-squared = 0.9406

F-value computed = 7.909 on 25 and 8 DF, F-value critical = 2.34

p-value = 0.002519

$H_0 = \beta_1 = \beta_2 = \cdots = \beta_k$ where $k = 7$

$H_A : H_0$ is not true

The null hypothesis is rejected and the equation is declared statistically significant at 95% confidence level.

The estimated ARDL equation with standard errors in the parenthesis was:

$$
\begin{aligned}
YUN = {} & 3.1881443_{\{1.1867684\}} + 0.6580734_{\{0.2268072\}} YUN_{t-1} - 0.2079220_{\{0.040073\}} GDP_t \\
& + 0.3848136_{\{0.0917367\}} GDP_{t-1} - 0.2052705_{\{0.0673422\}} GDP_{t-2} + 0.1208412_{\{0.0257892\}} GDP_{t-3} \\
& - 0.073030_{\{0.0149475\}} ED_t + 0.0214684_{\{0.0110151\}} ED_{t-1} - 0.009116_{\{0.00325426\}} ED_{t-2} \\
& + 0.0094806_{\{0.0125552\}} ED_{t-3} + 0.0025032_{\{0.0012111\}} FDI_{t-1} + 0.0006147_{\{0.001649\}} FDI_{t-2} \\
& + 0.0107310_{\{0.0054604\}} PI_t + 0.0069095_{\{0.0067884\}} PI_{t-1} + 0.0144567_{\{0.0059603\}} PI_{t-2} \\
& + 0.0083734_{\{0.0073051\}} PI_{t-3} + 0.0116937_{\{0.003686\}} LR_t - 0.0892691_{\{0.0373312\}} LR_{t-1} \\
& + 0.0557574_{\{0.0456325\}} LR_{t-2} + 0.016921_{\{0.0389834\}} LR_{t-3} + 2.0186805_{\{2.0022926\}} POP_t \\
& - 0.2590455_{\{0.036858\}} POP_{t-1} + 2.2498754_{\{3.3935272\}} POP_{t-2} - 4.3093119_{\{4.1772794\}} POP_{t-3}
\end{aligned} \tag{4.25}
$$

Table 4.2: The long run coefficient for selected ADRL Model

Regressors	Coefficient	Prob
Intercept	4.8447	0.041
GDP	0.09148447	0.05
ED	-0.0512	0.036
FDI	0.01196	0.02
PI	0.0615	0.04
LR	-0.0154	0.01
POP	1.762635	0.01

The long run relationship between the variables ad youth unemployment is depicted in the table 4.2 above. Of interest is the result that increase in GDP increases youth unemployment in the long run. Also increase in population, private investments, foreign direct investment in the long run increase youth unemployment. However, literacy rate and external debt reduce youth unemployment in the long run.

The figures below illustrate the fitted, actual, standardized and normality of estimated residuals

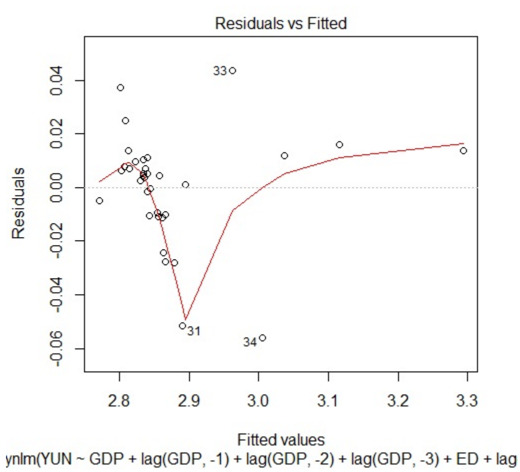

Figure 4.1: Fitted vs Actual Residuals of ARDL

The figure of Residual Vs Fitted plot shows the assumption of linearity and symmetry (along the zero line) of distribution of residual data points is upheld.

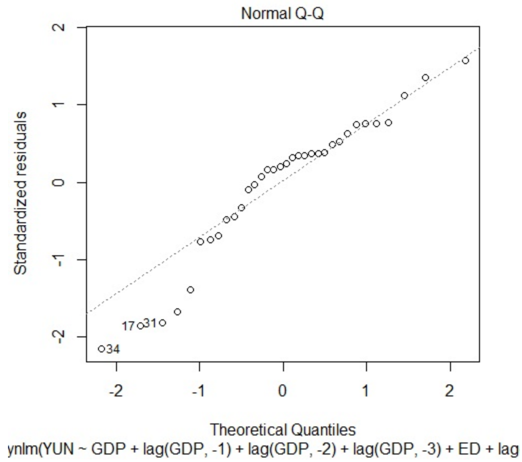

Figure 4.2: Normality of fitted residuals

From figure 4.2 above the normal probabilty plot depict that the errors of residuals are normally distributed.

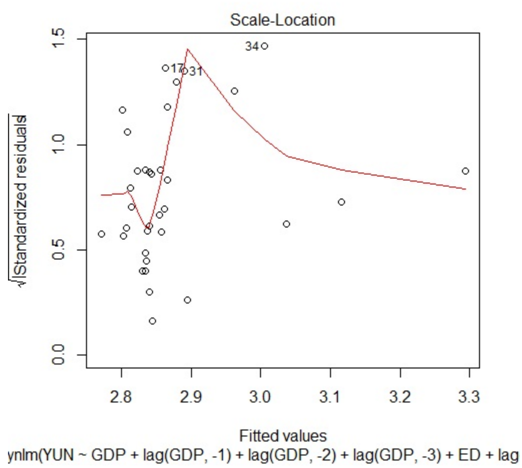

Figure 4.3: Scale and Location of Residual Data Points

The distribution of the standardized squares of residuals is relatively horizontally symmetrical strengthening the assumption of linearity of the fitted residuals.

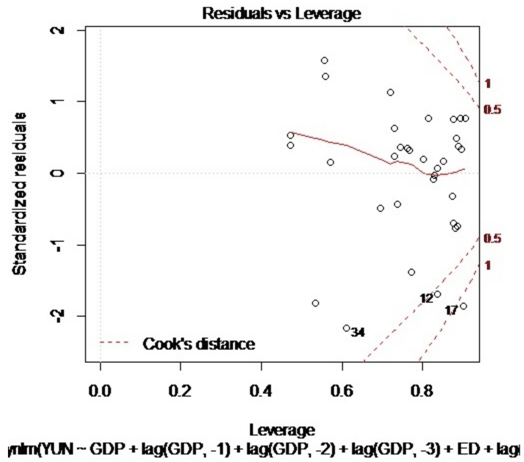

Figure 4.4: Residual vs Leverage

In figure 4.4 above the data points of residuals fell within the cooks red dashed lines. There are no outliers but 17 fell almost at the boundary described by the red dashed lines. This indicated that from the upper right and lower right corner data points of residuals the estimated model is stable and within limits.

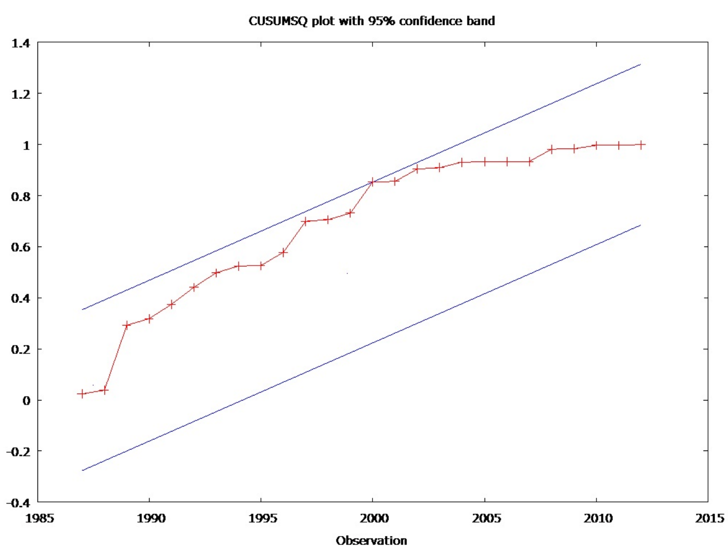

Figure 4.5: Cummulative Sum of Squares of Recursive residuals of ARDL

The figure 4.5 above indicates that the movement or deviation of the mean of the fitted model is well within the boundary of 95% confidence level. The model is found to be very stable.

4.4 ARDL Results Explanation

The results tabulated in table 4.1 indicated that GDP and its two-year lag had negative effect on youth unemployment, that is, one unit increase in GDP and GDP lag 2 reduced youth unemployment by 0.207922% and 0.2052705% respectively. This is explained as decrease in unemployment due to increase in productivity. Also, one unit of ED and ED two year lag reduced youth unemployment by 0.07303% and 0.009116% respectively. The external debt effects on youth unemployment occurs when money borrowed from foreign institution is invested in employment intensive sectors to create more jobs. Furthermore, unit increase in one year lag of youth literacy level reduced youth unemployment by 0.0892691%. This is explained that government intervenes to address backlog in unemployed youth in the previous year for political reasons. The same reason is given for lag one and three of population which reduced unemployment by 0.2590455% and 4.3093119% respectively. From table 4.2, in the long run, increase in GDP causes increase in youth unemployment by 0.09148447%. It is explained that GDP growth in the country is “ jobless growth” .

5 VECTOR ERROR CORRECTION MODEL

5.1 Introduction

In this chapter the ARDL is reparameterized to VECM structure and new parameters are estimated. The cointegrating relations discussed earlier in chapter three are restricted to obtain long run cointergrating relations. The impulse response functions and various decomposition are used to predict behaviour pattern of variables over time when subjected to unit shocks.

5.1.1 Reparameterization of ARDL to VECM

The conditions for the standard reparameterization of ARDL to VECM is that the variables must be cointegrated, that is $I(1)$, $I(2)$, $I(3)$ and their linear combination must be $I(0)$. Where this is not the case, the variables are differenced and cointegration test is run again.

ARDL(1,1)

$$Y_t = \delta + \theta_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + e_t \tag{5.1}$$

Assumption:Y and X are cointegrated at equilibrium $Y_t = Y_{t-1}, X_t = X_{t-1}$ and $e_t = 0$

Collecting like terms,

$$\left.\begin{aligned} &Y_t - \theta_1 Y_{t-1} = \delta + \delta_0 X_t + \delta_1 X_{t-1} \\ &Y_t(1-\theta_1) = \delta + \delta_0 X_{t-1} + \delta_1 X_{t-1} \\ &Y_t = \tfrac{\delta}{(1-\theta_1)} + \tfrac{(\delta_0+\delta_1)}{(1-\theta_1)} X_t \end{aligned}\right\}\text{forming new parameters from equilibrium}$$

This represents the function: $Y_t = \beta_1 + \beta_2 X_t$
where

$$\beta_1 = \frac{\delta}{(1-\theta_1)}, \beta_2 = \frac{(\delta_0+\delta_1)}{(1-\theta_1)} \tag{5.2}$$

β_1 and β_2 in (5.2) are the new parameters.
This is a long-run relationship.
We now manipulate (5.1) to derive ECM

$$Y_t - Y_{t-1} = \delta + (\theta_1 - 1)Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + v_t$$

Add $-\delta_0 X_{t-1} + \delta_0 X_{t-1}$

$$\Delta Y_t = \delta + (\theta_1 - 1)Y_{t-1} + \delta_0 X_t + (\delta_0 + \delta_1)X_{t-1} + v_t \tag{5.3}$$

Multiply (5.3) by $\frac{(\theta_1 - 1)}{(\theta_1 - 1)}$ and manipulate

$$\Delta Y_t = (\theta_1 - 1)\left[\delta(\theta_1 - 1) + Y_{t-1} + \frac{(\delta_0 + \delta_1)}{(1 - \theta_1)}X_{t-1}\right] + \delta_0 X_t + v_t \tag{5.4}$$

Reorganizing the equation using the definition of β_1 and β_2 we obtain

$$\Delta Y_t = \alpha(Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + \delta_0 \Delta X_t + v_t \tag{5.5}$$

Where,

$\alpha = 1 - \theta_1$, (10) is the error correction that embeds the cointegration relationship. And,

- $Y_{t-1} - \beta_1 - \beta_2 x_{t-1}$ shows deviation from its long run value of $\beta_1 + \beta_2 X_{t-1}$, that is error in the previous period where the downward arrows indicate the direction of deviation from equilibrium which are then reverted back to equilibrium in the model.
- If error term is positive such that $Y_{t-1} > \beta_1 + \beta_2 X_{t-1}$, $Y_t$falls
- If error term is negative such that $Y_{t-1} < \beta_1 + \beta_2 X_{t-1}$, Y_t rises $(\theta_1 - 1)$ is the correction of $\Delta\ Y_t$ to the error.

Since VECM and ARDL have different assumptions but are cited by Hassler and Wolters (2010) to be identical and equivalent upon reparameterization, the study held that two models are not numerically identical, although there is reversion of error to maintain equilibrium. We thus used data to validate this claim.

5.1.2 Estimation of VECM using OLS

The VECM is estimated using OLS framework like done previously in ARDL Model. We use maximum eigenvalue or trace test to identify the number of cointegrating relations then we proceed to estimate VECM of variables; Youth Unemployment rate, Population Growth, GDP, FDI, ED, PI, and LR.

$$\ln \Delta Y_t = \pi Y_{t-1} + \lambda_1 \Delta \ln Y_{t-1} + \cdots + \lambda_{p-1} \Delta y_{t-(p-1)} + \epsilon_t \tag{5.6}$$

where the Y_t consists of the above variables, Δ is the order of difference, p is a 7 by 1 intercept vector, λ is a 7 by 35 constant matrix,ϵ_t is a 1 by 35 in line with Bayesian criteria.

5.1.3 Estimation of VECM using Maximum Likelihood

Since the variables for the study are found to be normally distributed, Sargan (1964) for normal distribution is used like in (9) above.

The structure of VECM is presented as:

$$\Delta Y_t = \gamma + \sum_{i=1}^{p} \beta_1 \Delta Y_{t-i} + \Omega ECM_{t-1} + \epsilon_t \tag{5.7}$$

Where, ΩECM_{t-1} is the error correction component in the structure;

Ω measures the speed of correction of deviation from equilibrium;

Y_t denotes a vector of variables in the model ;

γ vector of constants;

β vector of parameters containing short-run information;

p the maximum lag; and ϵ_t is the vector of white noise.

$$\begin{aligned}\Delta \ln YUN_t =& \gamma + \sum_{i=1}^{p} \beta_1 \Delta \ln YUN_{t-1} + \sum_{i=1}^{p} \beta_2 \Delta \ln GDP_{t-i} + \sum_{i=1}^{p} \beta_3 \Delta \ln POP_{t-i} + \sum_{i=1}^{p} \beta_4 \Delta \ln FDI_{t-i} \\ &+ \sum_{i=1}^{p} \beta_5 \Delta \ln ED_{t-i} + \sum_{i=1}^{p} \beta_6 \Delta \ln PI_{t-i} + \sum_{i=1}^{p} \beta_7 \Delta \ln LR_{t-i} + \alpha ECM_{t-i} + \epsilon_t\end{aligned} \tag{5.8}$$

Where,

β_i $i = 1, 2, 3, 4, 5, 6, 7$

α, adjustment to equilibrium;

p, maximum lag;

ϵ_t, vector of error term.

When the variables of a VECM are not cointegrated, we use a vector autoregressive (VAR) model.

The Philip-Ouliaris test conducted earlier in Chapter 3 table 3.6 revealed cointegration of differential order I(0) and I(1). Also, the Johansen Cointegration test in table 3.7 indicated existence of three (3) cointegrating relations. The study proceeded to estimate VECM. Estimating of VECM starts from checking stationarity of variables as done earlier to test cointegration of the series. The Johansen cointegration test are either maximum eigen value test or trace test. The research also considered Johansen cointegration test and presented results in table 5.1

Table 5.1: Johansen Cointegration Trace Test Results

R	**Test Stats**	**10%**	**5%**	**1%**
$r \leq 6$	5.37	7.52	9.24	12.97
$r \leq 5$	15.44	17.85	19.96	24.60
$r \leq 4$	31.02	32.00	34.91	41.07
$r \leq 3$	48.18	49.65	53.12	60.16
$r \leq 2$	88.31	71.86	76.07	84.45
$r \leq 1$	145.46	97.18	102.14	111.01
$r \leq 0$	212.35	126.58	131.70	143.09

According to trace test, the first incident at which the critical value at 95% confidence level is greater than test statistics is when $r = 3$, i.e. $53.12 > 48.18$. This implied the existence of three Cointegrating relations.

The results are compared with Johansen cointegration eigen value test in table 3.7 and the results are similar as indicated below

In maximum eigen value test, the first incident at which the critical value at 95% confidence level is greater than test statistics is when $r = 3$, i.e. $28.14 > 17.17$. This implied the existence of three Cointegrating relations.

5.2 Estimating VECM using Trace Function

5.2.1 Unrestricted cointegrating equations

The unrestricted equations in appendix III are subjected to restrictions imposed in table 5.2 to obtain three long run relations described as "GDP Effect" and "ED Effect" and "FDI effect". Consequently, the cointegrating relations helped economic interpretation of YUN, GDP ,ED ,FDI, POP, PI, and LR.

From Johansen cointegration, we identified the three relations with error correction terms (ECT) which adjust the system to the equilibrium.
The unrestricted VECM equations are presented as follows:

$$\begin{aligned} GDP_t\ Effect = {} & 1.0GDP_t + 1.06O35e^{-15}ED_t - 3.209238e^{-17}FDI_t + 6.514101e^{1}YUN_t + 1.1577e^{1}PI_t \\ & - 3.148514e^{1}LR_t + 4.683578e^{1}POP + ECT \end{aligned} \tag{5.9}$$

$$\begin{aligned} ED_t\ Effect = {} & 0.0GDP_t + 1.0ED_t - 4.132979e^{-16}FDI_t - 1.201289e^{2}YUN_t - 1.998009e^{1}PI_t \\ & + 5.423654e^{1}LR_t - 8.622218e^{1}POP_t + ECT \end{aligned} \tag{5.10}$$

$$\begin{aligned} FDI_t\ Effect = {} & 0.0GDP_t + 1.077263e^{-15}ED_t + 1.0FDI_t + 2.26711e^{2}YUN_t + 3.840023e^{1}PI_t \\ & - 1.002592e^{2}LR_t + 1.660608e^{2}POP_t + ECT \end{aligned} \tag{5.11}$$

Table 5.2: Restrictions imposed on the structural VECM

EQUATION	GDP	ED	FDI	YUN	PI	LR	POP
DGP effect	1 (0.000)	0 (0.000)	0 (0.000)	17.71 5.06	3.15 0.90	-8.56 3.424	12.73 2.8288
ED Effect	0 (0.000)	1 (0.000)	0 (0.000)	-8.88 (3.552)	-453 (2.265)	14.74 4.211	-23.44 (4.261)
FDI Effect	0 (0.000)	0 (0.000)	1 (0.000)	16.75 (3.722)	10.44 (2.324)	-7.41 (2.47)	12.47 (4.156)
$x^2 = 12.59$, P-value = 0.003154							

Restricted cointegrating equations with vector error correction terms (ECT):

$$GDP_t\ Effect = GDP_t + 17.71YUN_t + 3.15PI_t - 8.56LR_t + 12.73POP_t + ECT_{t1} \quad (5.12)$$

$$ED_t\ Effect = ED_t - 8.88YUN_t - 4.53PI_t + 14.74LR_t - 23.44POP_t + ECT_{t2} \quad (5.13)$$

$$FDI_t\ Effect = FDI_t + 16.75YUN_t + 10.44PI_t - 7.41LR_t + 12.47POP_t + ECT_{t3} \quad (5.14)$$

5.3 VECM Unrestricted

The error correction terms (ECT) in the restricted equations (5.12), (5.13), and (5.14) above are represented as

$$\underline{\alpha}\underline{\beta}' = \prod \quad (5.15)$$

Where α and β , are coefficient matrix $\prod$ is the vector error correction term (ECT) matrix with each variable coefficient.

Matrices are:

$$\alpha = \begin{bmatrix} -0.081618248 \\ 0.505343993 \\ 3.923971897 \\ 0.182982408 \\ 0.182982408 \\ 0.222520093 \\ 0.002655281 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 1.00000000 \\ -0.09335818 \\ -0.10318207 \end{bmatrix} \beta' = \begin{bmatrix} 1.00000000 & 0.09335818 & -0.1031820 \end{bmatrix}$$

$$\alpha\beta' = \prod = \begin{bmatrix} -0.081618248 & 0.0076197315 & 0.0084215401 \\ 0.505343993 & -0.0471779975 & -0.0521424410 \\ 3.923971897 & -0.3663348906 & -0.4048835563 \\ 0.182982408 & -0.0170829053 & -0.0188805042 \\ 1.021564156 & -0.0953713745 & -0.1054071077 \\ 0.222520093 & -0.0207740718 & -0.0229600846 \\ 0.002655281 & -0.0002478923 & -0.0002739774 \end{bmatrix}$$

The matrix $\prod$ is obtained by multiplying loading matrix α with the transpose of β obtained from Johansen cointegration test

Lagged Unrestricted VECM equations

1. $GDP\ Effect = 0.68478311_{0.1423} - 0.081618248_{0.0278} + 1.041681964_{0.026932}GDP_{t-1} + 0.397612298_{0.00193}ED_{t-1} - 0.0255150492_{0.00867459}FDI_{t-1} + 0.9278922307_{0.045612}YUN_{t-1} + 0.005592738_{0.004908}PI_{t-1} - 0.149172387_{0.03625}LR_{t-1} - 31.8398511_{15.3447}POP_{t-1}$

2. $ED\ Effect = -3.18319209_{1.06} + 0.50534399_{0.25} - 1.184821983_{0.39}GDP_{t-1} - 0.436486222_{0.17}ED_{t-1} + 0.0617421278_{0.02}FDI_{t-1} - 2.458858892_{0.19}YUN_{t-1} + 0.063671436_{0.03}PI_{t-1} + 0.253072107_{0.08}LR_{t-1} + 53.4281154_{35.62}POP_{t-1}$

3. $FDI\ Effect = 22.32580046_{7.44} + 3.923971897_{1.31} - 2.498460158_{0.832}GDP_{t-1} - 3.055893269_{1.02}ED_{t-1} - 0.5656133375_{0.188}FDI_{t-1} - 12.045049342_{8.03}YUN_{t-1} - 0.318886184_{0.21}PI_{t-1} - 2.609287260_{0.87}LR_{t-1} - 11.58130717_{7.72}POP_{t-1}$

Equations of second lag

1. $GDP\ Effect = -0.06385096_{0.0213}GDP_{t-2}0.224104318_{0.064}ED_{t-2} - 0.0265371095_{0.00665}FDI_{t-2} + 0.87400097_{0.583}YUN_{t-2} + 0.0352681753_{0.0186}PI_{t-2} - 0.13053551_{0.0522}LR_{t-2} + 3.150141_{1.75}POP_{t-2}$

2. $ED\ Effect = -0.60590768_{0.22}GDP_{t-2} - 0.099116362_{0.33}ED_{t-2} + 0.0414003710_{0.014}FDI_{t-2} - 2.41348040_{1.609}YUN_{t-2} - 0.0696685665_{0.0279}PI_{t-2} + 0.23082382_{0.0769}LR_{t-2} + 29.359769_{7.34}POP_{t-2}$

3. $FDI\ Effect = 1.00925248_{0.3342}GDP_{t-2} - 2.565540917_{0.855}ED_{t-2} - 0.7770933637_{0.31}FDI_{t-2} + 16.26597627_{7.394}YUN_{t-2} + 0.7382019274_{0.295}PI_{t-2} - 3.31568913I_{0.947}LR_{t-2} + 191.920192_{38.384}POP_{t-2}$

Equations of third lag

1. $GDP\ Effect = -0.64496693_{0.21}GDP_{t-3} - -0.292881979_{0.12}ED_{t-3} - 0.0092672922_{0.0038}FDI_{t-3} + 0.69493831_{0.28}YUN_{t-3} + 0.0325996781_{0.02}PI_{t-3} + 0.519181892_{0.26}LR_{t-3} + 9.4242655_{3.49}POP_{t-3}$

2. $ED\ Effect = 0.29177615_{0.07}GDP_{t-3} + 0.124253977_{0.02}ED_{t-3} + 0.0178582429_{0.0017}FDI_{t-3} - -2.95338510_{1.181}YUN_{t-3} - 0.0315248454_{0.01212}PI_{t-3} - 0.261408639_{0.07468}LR_{t-3} + 8.9361943_{3.7234}POP_{t-3}$

3. $FDI\ Effects = -17.61866858_{4.4046}GDP_{t-3} - 12.511462711_{4.03595}ED_{t-3} - 0.2764977772_{0.5529}FDI_{t-3} - 10.53903805_{8.54055}YUN_{t-3} + 0.7370746046_{0.3704}PI_{t-3} - 4.059738566_{1.6239}LR_{t-3} + 4.53959332_{2.4375}POP_{t-3}$

In order to conduct t-test and F-test on the estimated VECM long run equations, the reparameterized VECM is transformed into VAR model. From the system of VAR equations the one with youth unemployment as dependent variable is picked and estimated and results

Table 5.3: VECM F-test and T-test estimation results

Variables	Estimate	Std Error	T	P-value
GDP.dl1	-0.11523	0.0384113	-3.0025	0.0225340*
ED.dl1	-0.01824	-0.00698	2.613	0.122945*
FDI.dl1	0.347005	0.153209	2.265	0.034780 *
YUN.dl1	-0.46469	-0.11614	0.4001	0.000751***
PI.dl1	-0.08384	0.02396	-3.4991	0.0049536*
LR.dl1	-0.044997	0.019639	-2.291	0.032943 *
POP.dl1	1.031723	0.215777	4.781	0.000114 ***
GDP.l1	0.093550	0.026691	3.505	0.002229 **
ED.l1	0.008972	0.006957	1.290	0.211898
FDI.l1	-0.16458	0.05486	-3.694	0.495518**
YUN.l1	-0.24948	0.07118	-3.5049	0.0481276*
PI.l1	0.080744	0.032297	2.50047	0.0938807.
LR.l1	0.032324	0.024420	1.324	0.200551
POP.l1	-0.065884	0.026735	-2.464	0.022910 *
constant	1.053876	0.40378	0.298	0.0969139.

presented in the table above. From the results the computed F-statistics > critical F-value. Also p-value of coefficients estimated revealed their significance. As such the estimated VECM is found to be significant.

Adjusted R-squared: 0.8007

F-statistics = 10.37 on 15 and 20 DF, p-value: 0.01

F-value critical = 2.85 < 10.37F-value computed the null hypothesis is rejected and equation declared statistically significant.

The CUSUM chart of VECM presented above shows that unlike ARDL the model is unstable. The movement of deviation of mean extended beyond the boundaries at 95% confidence level.

5.4 Variance Decomposition (VD)

In the study analysis, three years and below is considered as short run while the rest is long run and therefore two, three and ten years indicate short-term and long-run respec-

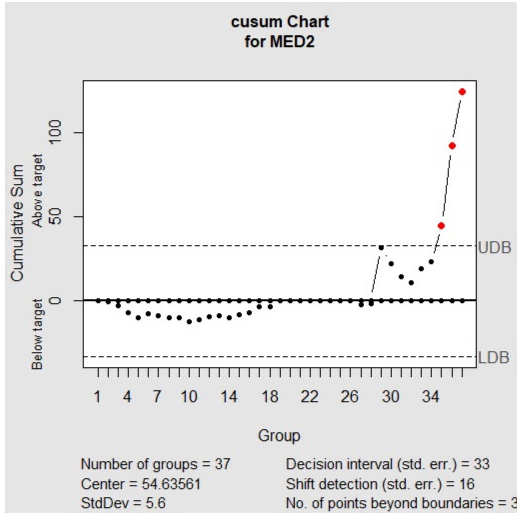

Figure 5.1: Cummulative Sum of Recursive Residuals of VECM

tively. The figure 5.2 below illustrates the propositions of the reaction over time of YUN due to its own innovations in relation to disturbance of variables in the system. For instance an innovation to a particular variable affects the rest of the variables in the VECM structure.

The VD helps to identify the length of forecast error variance over time as a result of unit shock to YUN_t at $t = 0$. The study plotted variance decomposition of GDP, ED, FDI, PI, LR, POP, to a unit shock in YUN as illustrated below. For a unit exposure of shock to YUN, a lag of three years is experienced followed by a response of steady increase for ten (10) years. However, the exposure of one unit shock of YUN lead to two year lag of no movement followed by exponential increase for a period of 10 years. Contrarily, exposure of one unit shock to YUN led to one year lag of no movement followed by ten (10) year sharp decline in LR and PI.

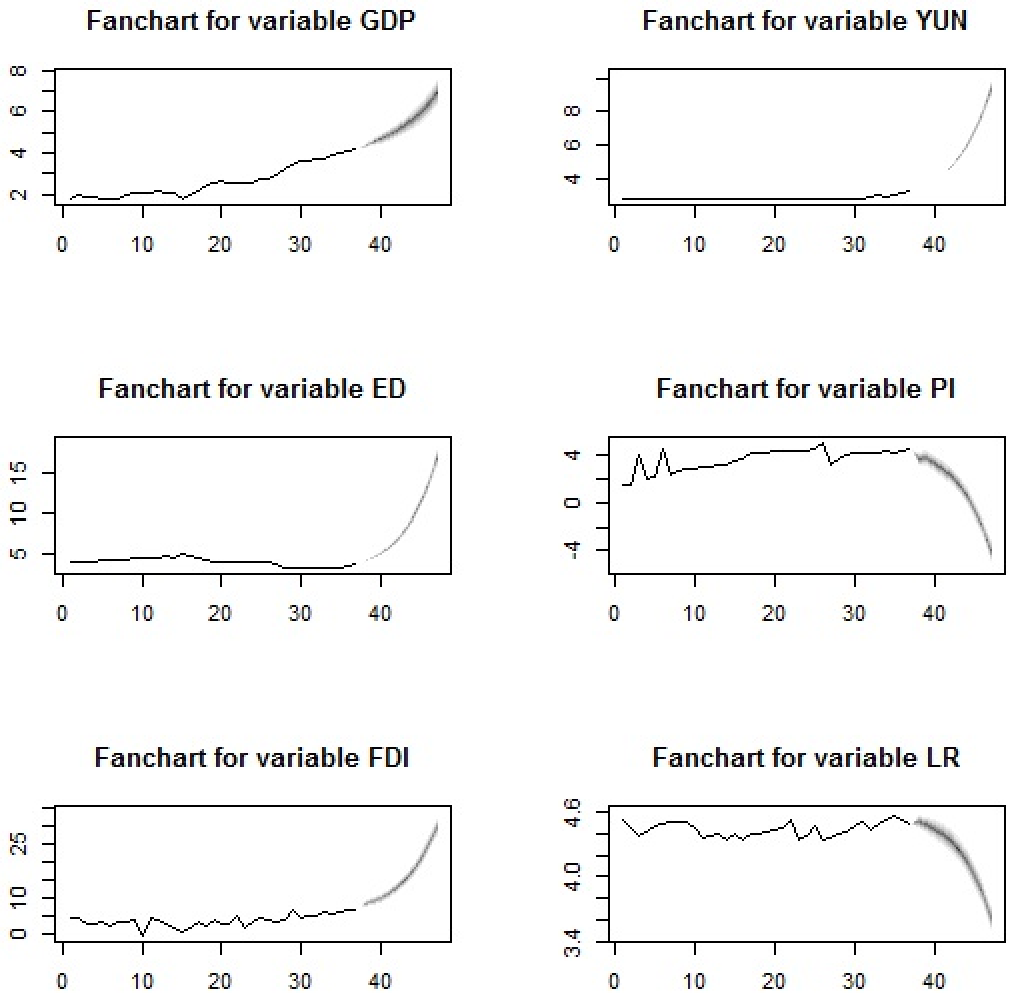

Figure 5.2: Variance Decomposition of variables

5.5 The Estimated Impulse Response Function (IRF)

A unit shock is applied, to each variable, to verify the effect on the dependent variable. The study used adjusted method in setting the ordering of the variables. A positive shock of one standard deviation is applied to the error term to determine the reaction of the dependent variable. The shock is applied to each variable because we are interested in the reaction of the Youth Unemployment to the other variables.

The IRF analysis of the VECM model used has researched long run effects on variables when structural shock of one unit is applied to each variable. The IRF in this study traced the responsiveness of youth unemployment in the VECM to shocks to other variables (GDP, ED, FDI, PI, LR and POP).

We plot impulse response function of YUN to all other variables. The YUN response to all other variables in the VECM system made it possible to conduct economic interpretation related by the long run relation among youth unemployment and other variables as illustrated below.

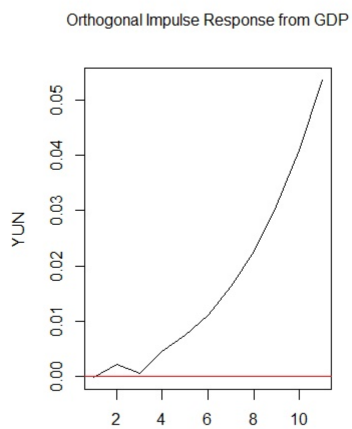

Figure 5.3: Impulse response of GDP to a unit shock of youth unemployment

Explanation

The youth unemployment's response to GDP shock, displayed in Figure 5.3, is interpreted in the short run, and observed that GDP shock gives a negative effect on youth unemployment stretching out to the third period. However, in the long run, it can be verified that GDP shock results in increase in youth unemployment. Its effect is not stationary but in the long run it adjust to equilibrium after 10 years.

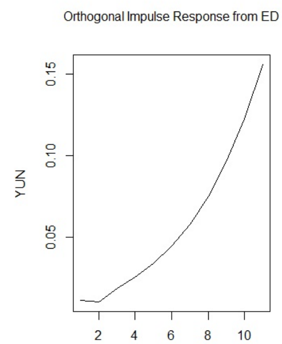

Figure 5.4: Short run and long run effect to youth employment to a unit shock to external debt.

Explanation

The figure above illustrates reaction of external debt on innovation to youth unemployment. In the short run, a constant (neutral) upto the second period but increases upto the tenth period.

From the figure 5.4 above, it is observed that the neutral effect has disappears after 2 years indicating that external debt shock affects youth unemployment over time. The external shock can be compensated by increasing private investment in labour intensive sectors.

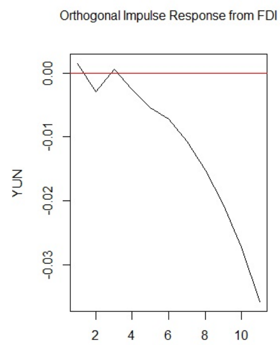

Figure 5.5: Short run and long run effect to youth employment to a unit shock to foreign direct investment(FDI).

Explanation

Figure 5.5 above depicts FDI shock effect on youth unemployment. In the short run, there is a brief intermittent negative and positive effect on youth unemployment up to almost fourth period, and a decreasing negative effect on youth unemployment approximately to the eleventh period is observed. Furthermore, this negative effect has decreased substantially.

In the illustration, it is seen that this negative effect does not disappear and never approaches zero . This result indicating that FDI shock affects youth unemployment over time.

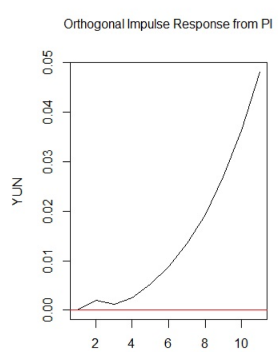

Figure 5.6: Effect of youth employment to a unit shock to private invesment(PI).

Explanation

Figure 5.6 illustrates the effect of private investment shock on youth unemployment. In the figure 5.6, the negative effect of private investment shock on youth unemployment has showed a decrease tendency approaching the fourth period. After this period, the decreasing negative effect of private investment is stationary in short run. The adjustment of a new equilibrium takes approximately 2 years. After this period there a strong positive increase in youth unemployment. The positive effect also supports the direction of relation in the long run. It can be deduced that private investment shock is an important factor for increasing youth unemployment.

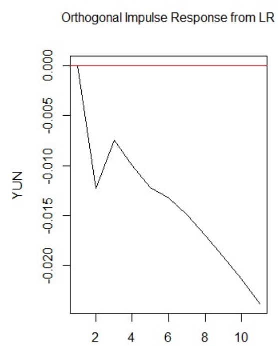

Figure 5.7: Short run and long run effect to youth employment to a unit shock to youth literacy rate(LR)

Explanation

In figure 5.7, the negative effects of Literacy Level on youth unemployment are observable. Up to the second period, the negative effect of Literacy Level shock on youth unemployment reduced significantly. At this period, this negative effect has tended to increase slightly, and reduced steadily afterwards. This result shows that Literacy Level is a very important factor in reducing youth unemployment.

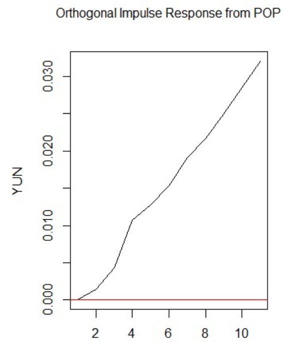

Figure 5.8: Effect to youth employment to a unit shock to youth population(POP)

Explanation

Figure 5.8 illustrates population shock to youth unemployment. The figure 5.8 shows that the population shock increases youth unemployment overtime. The adjustment to equilibrium takes approximately one year. Consequently, this positive effect explained the direction of relation between population and youth unemployment. It can be deduced that the population shock has increased unemployment only for the immediate short run and extensive long run period.

6 DISCUSSIONS, CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

The results obtained are discussed in this chapter. These include unit root test, multiple correlation, serial correlation test, cointegration test and regression analysis, ARDL analysis, VECM analysis, VD, and IRF analyses.

As shown in Table 3.1, a simple correlation test is first conducted. From the results, we see that none of the variables explains over 50% variation of youth unemployment. Individually GDP, ED, and PI had a correlation of 29.4%, 44% and 48% with YUN respectively; which are the highest correlation coefficients, the remaining exhibited much less.

The study subsequently computed the multiple correlation test and established causality-effect on the dependent variable. This is done by assesing the magnitude of R-squared.

After ascertaining that the independent variables are correlated with YUN, the study proceeded to establish stationarity of variables used by means of ADF test. Also, when the variables are not of the same stationary process, we could still perform cointegration test to construct ARDL and VECM. The results in Table 3.6 indicatesstationarity of the variables. The study proceeded to perform cointegration test to ascertain the level of integration of each series.

6.2 Cointegration of the series

To check relationship of the variables overtime we conducted two cointegration tests: First Phillips-Ouliaris test, two Johansen cointegration test. The Phillips-Ouliaris test ascertains order of cointegration between two variables in turns. The Phillips-Ouliaris test reveals that variables are cointegrated as shown in the Table 3.6. Out of 21 relations, 6 are I(0) while 15 are I(1). The Phillip-Ouliaris test has the limitation of NOT conducting cointegration test of more than two variables. Provided the series are either I(0) or I(1), the condition allows one to conduct ARDL estimation for series cointegrated at different levels. The test is further limited in identifying coefficients of cointegration necessary to identify cointegrating vectors and equations. With these limitations the study further conducted Johansen Cointegration test.

Johansen (1988) cointegration rank test has the ability to perform cointegration test for more than two variables. As shown in Table 3.7, we conducted cointegration test for the variables at 95% significant level. The study used both Trace test and Maximum Eigen value test. In either test, we rejected the null hypothesis when the test statistics took on a value below the critical value of a given significance level.

$H_0 : r \leq 3$, the study rejected the null hypothesis and concluded that there are $r = 3$ cointegrating vectors at 5% significant level. From the results 28.14 > 17.17 it implied that we needed a linear combination of three time series to form a stationary series. Thus the test revealed that the cointegration rank r=3. These results indicated that the variables would not wander far away from each other over time. The study therefore proceeded to estimate VECM and ARDL with r=3. The test statistics revealed the existence of three cointegrating relationships among the six variables. With cointegrating vectors, there is basis for VECM and ARDL estimation.There are three cointegrating relationship among the seven variables which implied there is a long-run equilibrium condition binding the levels of the variables together.

6.3 Estimated Linear Cointegration equation

Having established the existence of cointegration among the variables, the study proceeded to estimate linear cointegration equation (3.8) with YUN being the dependent variable.

The estimated equation (3.8) revealed that 86.72% of variation in youth unemployment is explained by GDP, external debt, FDI, PI, literacy level, and the youth population.

6.4 Cointegrating Relations Results

In Johansen Cointegration Test the cointegration relationships are normalized to the first column. The columns corresponding to the maximum eigenvalue(s) form cointegrating relationship. In this study the 3 columns corresponding to the maximum Eigen values are GDP, ED, and FDI as depicted in Table 3.8.

The three equations are further taken through ADF test to ascertain their stationarity. Their respective p-values are recorded in Table 3.9. The three equations (3.4), (3.5) and (3.6) are found to be stationary. Which implies they are actually the cointegrating vectors. Also, corresponding time plots are drawn as illustrated in figures 3.8, 3.9, and 3.10.

6.5 Estimated Auto-regressive Distributed Lag (ARDL) Results

ARDL is a dynamic model whose structure is shown in equation equation(4.24) above study chose lag length using AIC, HQ, SC and FPE as illustrated in table 3.4. The appropriate lag selection is found to be 3 given by FPE. However, the dependent variable is lagged once as shown in equation (4.25).

The results as presented in equation (4.25) with the appropriate number of lags and respective coefficients, showing adjusted R-squared = 0.9406, F-value computed =: 7.909 on 25 and 8 DF, F-value Critical=2.34, p-value: 0.002519.

ARDL equation is significant.

From the results tabulated, the study found out that GDP and its second lag have negative effect on youth unemployment, that is, one unit increase in GDP and GDP lag 2 reduce youth unemployment by 0.207922 and 0.2052705 respectively. This can be explained as decrease in unemployment due to increase in productivity. Also, one unit of ED and ED lag 2 reduce youth unemployment by 0.07303 and 0.009116 respectively. The external debt effect on youth unemployment occurs when money borrowed from foreign institutions is invested in employment intensive sectors to create more jobs.

Furthermore, unit increase in one year lag of youth literacy level reduces youth unemployment by 0.0892691. This is explained that government intervenes to address backlog in unemployed youth in the previous year for political reasons. The same reason is given for lag one and three of population which reduce unemployment by 0.2590455 and 4.3093119 respectively.

6.6 Estimated Vector Error Correction Model Results

After choosing lag length, the study established stationarity of variables using ADF test as explained earlier. The most important condition for estimation of VECM is the cointegration of the series used. The study used Johansen test to establish the number of cointegrating relationships as illustrated earlier. There are three cointegrating relations $r = 3$.

The first incident at which the critical value at 95% confidence level is greater than test statistics is when $r = 3$, i.e. $28.14 > 17.17$. The cointegrating condition is satisfied.

The results of Johansen test revealed three cointegration vectors ($r = 3$) described as "GDP effect", "FDI effect" and "External Debt effect". Consequently, the cointegrating vectors are identified for the economic interpretation of YUN, GDP, ED, FDI, POP, PI, and LR variables.

We then estimate VECM to obtain the three equations with error correction terms which adjust the system to the equilibrium.
The equations are presented in (5.9), (5.10) and (5.11).
The VECM equations are subjected to the restrictions in table 12 to form structured VECM equations. From the unrestricted VECM equations in appendix III, we obtained equations (5.9), (5.10) and (5.11)which are restricted to obtain equations (5.12), (5.13) and (5.14). The latter equations are subjected to ADF stationary test and found to be stationary.
The error correction term matrix (ECT) in equation (5.10) from the VECM restriction on table 5.2 is described in equation (5.10)
With successful construction of VECM, the study proceeded to conduct variance decomposition (VD) and impulse response functions (IRF).

6.7 Variance Decomposition Results

To establish instantaneous causality the research constructed variance decomposition of all the variables used. Figure 5.2 illustrates the propositions of the movement of the dependent variable as a resdult of "own" shocks in relation to variables in the system.

VD determines the extent of movement overtime of variables as explained when a unit shock to YUN_t at $t = 0$. The study thus plotted the impulse response functions of GDP, ED, FDI, PI, LR, POP, to a unit shock in YUN as illustrated below. For a unit exposure of shock to YUN, a lag of three years is experienced followed by a response of steady increase for ten (10) years. However, the exposure of one unit shock of YUN lead to two year lag of no movement followed by exponential increase for a period of 10 years in the following series GDP, ED, and FDI. Contrarily, exposure of one unit shock to YUN led to one year lag of no movement followed by ten (10) year sharp decline in LR and PI.

6.8 IRF

IRF Analysis of the VECM used has researched long run effects on variables when structural shock of one unit is applied to each variable. The IRF in this study traced the responsiveness of youth unemployment in the VECM to shocks to other variables (GDP, ED, FDI, PI, LR and POP).

It is possible to plot the impulse response function of YUN to all other variables. The YUN response to all other variables in the VECM system made it possible to conduct economic interpretation by the long run relation among youth unemployment and other variables.

The youth unemployment's response to GDP shock, displayed in figure 5.3, is interpreted in the short run, and observed that GDP shock gives a negative effect on youth unemployment stretching out to the third period. However, in the long run, it can be verified that GDP shock affects youth unemployment positively. The movement is not stationary overtime. It takes 10 years for the movement to adjust to equilibrium. The observation is consistent with results of GDP in the ARDL.

The effect of external debt innovation on youth unemployment is interpreted as follows: in the short term it is slightly negative but increases unemployment overtime.

From the figure 5.4, it is observed that the neutral effect has disappeared. External debt shock has long run effect on youth unemployment. The external debt shock can be compensated by increasing private investment in labour intensive sectors.

The effect of FDI shock to youth unemployment is depicted in figure 5.5 depicts. In the short run, there is a brief intermittent negative and positive effect up to almost the fourth period, and a decreasing negative effect approximately up to the eleventh period is observed. Furthermore, this negative effect decreased substantially. It is seen that the negative effect does not disappear. The result indicated that FDI shock increases on youth unemployment in the long run. FDI shocks is compensated by external debt effect.

The effect of private investment shock on youth unemployment is illustrated in figure 5.6. In the figure 5.6, the negative effect of private investment shock on youth unemployment has shown a decrease tendency approaching the fourth period. After this period, the decreasing negative effect of private investment is stationary in short run. The adjustment of a new equilibrium takes approximately 2 years. After this period there is a strong positive increase in youth unemployment overtime. The positive effect reveals direction of relation overtime. It is deduced that PI shock is important in reducing youth unemployment.

In figure 5.7, the negative effects of Literacy rate on youth unemployment are observable. Up to the second period, the negative effect of Literacy Level shock on youth unemployment reduced significantly. At this period, this negative effect has tended to increase slightly, and reduced steadily afterwards in the long run. As a result, the labor demand shock doesn't have an important effect on unemployment overtime. Literacy Level is very important factor in reducing unemployment.

The effect of population shock on youth unemployment is illustrated in figure 5.8. The population shock has a positive relationship with youth unemployment overtime. The adjustment to equilibrium takes approximately one year. Consequently, this positive effect has explained the direction of relation between population and youth unemployment. It is deduced that the population shock increases youth unemployment in the immediate short term and extensive long run period.

6.9 Comparative Analysis of ARDL and VECM

The study used F-test and t-test to compare the results of ARDL and VECM. Both models are comparatively significant. The results of ARDL and VECM in table 4.1 and 5.3 respectively gave F-test and t-test results. The analysis of F-test is illustrated in table 6.1 above. Both models estimated are found to be significant. The CUSUM test for ARDL and VECM in figure 4.5 and 5.1 respectively revealed that ARDL is more stable than VECM. The residuals of ARDL are within 95% confidence interval whereas VECM residuals plot extend beyond the boundary. The VECM is most suitable for multivariate studies on unemployment since it isolates cointegrating vectors that select parameters and identifies equation of interest. In this study VECM has identified GDP effect, FDI effect, and ED effect. VECM estimation selected and isolated equations (5.12), (5.13), and (5.14) from (5.9), (5.10), and (5.11)respectively. The IRF and VD from VECM in the study offer easier and simpler way of simulation and focasting.

According to this study, in general, the suitable reparameterized Mathematical model for youth unemployment is the reparameterized VECM developed here in.

Table 6.1: F-test comparison of ARDL and VECM

MODEL	F-COMPUTED	F-CRITICAL	OBSERVATION
VECM	10.37	2.85	Significant
ARDL	7.909	2.34	Significant

Hypothesis Verification

HYPOTHESIS	FINDINGS	POSITIVE EFFECT	NEGATIVE EFFECT
H_1	SATISFIED	N/A	N/A
H_2	SATISFIED	N/A	N/A
H_3	NOT SATISFIED	POP, PI,FDI	GDP, ED, LR
H_4	NOT SATISFIED	POP, PI,FDI	GDP, ED, LR

6.10 Summary and conclusion of the Results of ARDL and VECM

The data is taken through multiple correlation test, Box-Pierce serial correlation test, stationarity test and Johansen cointegration test, and the ARDL is estimated. The results revealed that GDP and its second lag had negative effect on youth unemployment, that is, one unit increase in GDP and GDP two-year lag reduced youth unemployment by 0.207922% and 0.2052705% respectively. This is explained as decrease in unemployment due to increase in productivity. Also, one unit of ED and ED two-year lag reduced youth unemployment by 0.07303% and 0.009116% respectively. The external debt effect on youth unemployment occurs when money borrowed from foreign institutions is invested in employment intensive sectors to create more jobs. Furthermore, unit increase in one-year lag of youth literacy level reduced youth unemployment by 0.0892691%. The reduction is explained by government interventions like Kazi Kwa Vijana, Youth Enterprise Development Fund (YEDP), National Youth Service programs among others to address backlog in unemployed youth in the previous year(s). The same reason is given for one-year lag and three-year lag of population which reduced unemployment by 0.2590455% and 4.3093119% respectively. The Foreign Direct Investment (FDI) and Private Investment (PI) did not have statistically significant effects on youth unemployment but increased by 0.01196 and 0.0615 respectively in the long run. F-test showed that ARDL model estimated in the study is significant while CUSUM test indicated its stability. Except for FDI, the results are consistent with Muhammad et al (2013) which used ARDL. The findings revealed that GDP, POP and FDI are key factors in determining Pakistan labor market relation. In the long-run, increase in GDP caused increase in youth unemployment by 0.09148447%. It is explained that GDP growth in the country is " jobless growth" as found by Ajilore and Yinusa (2011) in Botswana. However, the findings are contrary to Kabaklarli et al (2011) in which 1% increase in GDP growth in Turkey reduced youth unemployment by 3.07%. Also, increase in previous literacy rates by one unit in the long run, reduced youth unemployment by 0.0154% contrary to study by Guillermo et al (2012) in Brazil whose findings revealed that higher educational levels do not compensate for unemployment episode in the past.

With VECM, we analyzed youth unemployment dynamics in Kenya. The study focused on establishing the long run effects of macroeconomic shocks on youth unemployment. In this case, youth unemployment rate, gross domestic product, external debt, private investment, literacy rate, FDI, and youth POP variables between 1979 and 2015 are used. A VECM is estimated. First, the results of the Augmented Dickey Fuller unit root tests used in the study revealed that GDP, ED, and FDI are I(0),that is, stationary in level. However, YUN, PI, LR, and POP are I(1), i.e. stationary after first difference. Then the Johansen cointegration analysis is employed. The results of trace and maximum eigen value cointegration tests revealed the existence of three cointegrating vectors. The three cointegration vectors showed three long run relationships which are interpreted as GDP Effect, ED Effect, and FDI Effect. In addition, the Structural VECM Model is constructed by means of restrictions in the long run impact matrix by analyzing the results of the Cointegration Analysis.

Based on the results of the IRF analysis and variance decomposition analysis of the structural VECM, it is concluded that GDP, literacy level, population, and FDI shocks determines Kenyan youth unemployment. Whereas population, external debt, private investment, and GDP had positive effects, FDI and LR had negative effects on youth unemployment in the long run. Notably, GDP has a short run inverse relationship with youth unemployment. In the second period, the negative effect of Literacy Rate shock on youth unemployment reduced significantly. At this period, the negative effect tended to increase slightly, and reduced steadily afterwards. This shows that Literacy Rate is an important factor in reducing youth unemployment. The population shock has a positive relationship with youth unemployment. Consequently, this positive effect explained the direction of relation between population and youth unemployment. It is deduced that the population shock contributes positively to unemployment immediately and extensive overtime. Results for all variables indicated that adjustment(s) to equilibrium took approximately 3 years.
From the methods employed and results generated, both ARDL and VECM estimated in the study are statistically significant. Both F-test and t-test revealed that the two models are significant. However, CUSUM test revealed that ARDL is the stable model.

6.11 Recommendations

6.11.1 Further Research

- Embedding Granger causality in longrun VECM long run relations without reparameterization and detrending.
- Further study should scrutinize direct link deriving Mathematical relationship among lag length equations, stationarity equations and cointegration equations to shorten the

process of estimating ARDL and VECM.

6.11.2 Policy Recommendation

- Policy makers should focus on interventions that are labour intensive through programs budgeting for investments with proceeds from GDP, ED, and FDI to curb youth unemployment

REFERENCES

References

[1] AwuorPonge, *Graduate Employment and Unemployability in Kenya: Transforming University Education to Cope with Market Demands and the Lessons for Africa*, International Journal of Social Science Tomorrow, Vol.2 No.3s

[2] AH Studenmund - *Addison WesleyUsing Econometrics*: A Practical Guide 2000, pg 418.

[3] Banerjee, A., A. Dolado, J.J, Hendry, D.F., Smith G.W. (1986). *Exploring Equilibrium relationships in econometric through static models*: Some Monte Carlo Evidence. Oxford Bulletin of Economics and Statistics 48: 253-277.

[4] Boswijk, H.P. (1995). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* 63 37-60.

[5] Christiana E.E. Okojie *Employment Creation for Youth in Africa*: The Gender Dimension, Expert Group Meeting on Jobs for Youth: National Strategies for EmploymentPromotion,15-16 January, 2003, Geneva, Switzerland.

[6] Chigunta F. *"The Socio-Economic Situation of Youth in Africa*: Problems, Prospects and Options", Paper presented at the Youth Employment Summit, Alexandria, Egypt, September, 2000.

[7] Davidson P. "Post Keynasian Employment Analysis and the Macroeconomics of OECD Unemployment," *The Economic Journal* (1998)

[8] Diagne, S.B, and Ossebi, *The Cultural Question in Africa: Issues, Questions and Research Prospect*. Dhaka: CODESRIA, 1996

[9] Dickey, D. A., Fuller, W.A, (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427-431.

[10] Dickey, D. A., Fuller, W.A, (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057-1072

[11] Dodd, Carley H. *Dynamics of Intercultural Communication*. 5th ed: Mcgrawhill, 1998.

[12] Engle, R.F and C.W.J. Granger (1987) " Cointegration and Error Correction Representation : Estimation and Testing," *Econometrica Theory*, 11:1131-1147.

[13] Hendry, D.,A.Pagan and J. Sargan (1984), *"Dynamic Specifications,"* Chapter 18 in Handbook of Econometrics, Vol II (ed., Z.Griliches and M. Intriligator), North Holland.

[14] Eita, Joel Hinaunye and Johannes M. Ashipala , Determinants of unemployment in Namibia. *International Journal of Business and Management*, Volume 5(10), pp. 92-104, (2001)

[15] Engle, Robert F. and C. W. J. Granger, Cointegration and error correction: Representation, estimation, and testing. *Econometrica*, Volume 55(2), pp. 251- 276, (1987)

[16] Esra Kabaklarli1,Perihan Hazel ER2,Abdülkadir Buluş *Economic Determinants of Turkish Youth Unemployment Problem*: Cointegration Analysis International Conference On Applied Economics – ICOAE 2011, pp 267-272.

[17] Fujiwara Yoshi Aoyama Hideaki: *A Stochastic Model of Labor Productivity and Employment* ,Kyoto University The Research Institute of Economy, Trade and Industry (2007)

[18] Gok (2010) The Prevention of Organized Crimes Act 2010, Nairobi, Government Printer.

[19] Hamilton, James D. 1994. Time Series Analysis. Princeton, N.J: PrincetonUniversity Press.

[20] Henrik Urdal; A Conflict of generations? Youth Bulges and Political Violence, International quarterly Vol 50 Issue 3 pages 607-629 September 11 2006.

[21] Hofstede, Geert, and Gert Jan Hofstede: *Cultures and Organisations*: Software of the mind, I (2005)

[22] Hofstede, Geert. *Cultures consequences*:International Differences in work related values. Beverly Hills: Sage, (1980)

[23] Jamil, Ishtiaq. *Administrative culture*: a mode of understanding public administration across cultures. Bergen: LOS-senteret, (1993).

[24] Johansen, S.(1991), " Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59:1551-80

[25] James H. Stock and Mark W. Watson Source: Journal of Money, Credit and Banking, Vol. 39, Supplement 1: Conference on "Quantitative Evidence on Price Determination", September 29 and 30, 2005, Co-sponsored by the JMCB and the Federal Reserve Board (Feb., 2007), pp. 3-33

[26] Kalim, R. , Population and unemployment: A dilemma to resolve. *The IUP Journal of Economics*, Volume 2, Issue 3, pp. 7-15. (2003)

[27] Kennedy N. Ocharo 1*, Nelson W. Wawire, Tabitha K. Ng'ang'a, George Kosimbei *Private capital inflows and economic growth in Kenya*;

[28] Kenya Population Cencus 2009

[29] International Journal of Development and Sustainability ISSN: 2168-8662, Volume 3 Number 4 (2014): Pages 810-837 ISDS

[30] MoEST- National Education Sector Plan 2013

[31] Muhammad ShahidMaqbool, Tahir Mahmood Abdul Sattar and M. N. Bhalli*Determinants of Unemployment Empirical Evidences from *Pakistan Economic and Social Review* Volume 51, No. 2 (Winter 2013), pp. 191-207.

[32] National Council of Population and Development Kenya's Demographic Dividend Roadmap *Harnessing The Potential Of Young Persons For Development* (NCPD, 2013)

[33] Okun, A. 'Potential GDP: Its measurement and significance', In Okun, A. (ed.), *The Political Economy of Prosperity*. Washington, DC. (1970)

[34] Pesaran, M.H. (1997), " The Role of Economic Theory in Modelling the Long-Run", The Economic Journal, 107: 178-191.

[35] Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems," Econometrica, 59:283-306

[36] Phillips, P.B.C. and B. Hansen (1990), "Statistical Inference on Instrumental Variables Regression with I(1) Processes," Review of Economic Studies, 57: 99-125

[37] Phillips, P.B.C. and M. Loretan (1991), "Estimating Long Run Economic Equillibria," *Review of Economic Studies*, 58: 407-436.

[38] United Republic of Tanzania, Policy Gaps to Assist Academically Disadvantageous Youth, Policy Brief No.22 ,1995

[39] J. D. Sargan Source (1964) Three-Stage Least-Squares and Full Maximum Likelihood Estimates Author(s): Econometrica, Vol. 32, No. 1/2 (Jan. - Apr., 1964), pp. 77-81 Published by: The Econometric Society

[40] UweHassler and JurgenWolters *Autoregrassive Distributed Lag Models and Cointegration*, Allgemeines Statistisches Archive 0, 0-15 Physica-Verlag0, ISSN 0002-6018

[41] Valadkhani, A. The causes of unemployment in Iran: An empirical investigation. *International Journal of Applied Business and Economic Research*, Volume 1(1), pp. 21-33. (2003)

[42] Wickens, M.R. and T.S. Breusch (1988), "Dynamic Specification, the Long Run Estimation of the Transformed Regression models," *The economic Journal*, 98: 189-205

APPENDIX I: R-CODES USED

Stationarity Test

The R-codes used for Augmented Dickey-Fuller Tests are:

>adf.test(MAED[,4])

>adf.test(diff(MEAD[,4])) # for 1st difference #

>adf.test(diff(diff(MAED[,4]))) # for 2nd difference #

>adf.test(diff(diff(diff(MAED[,5])))) # for the 3rd difference #

Multiple correlation test > CT<-lm(data = MED,YUN ~ GDP + ED + FDI + PI + LR + POP)

> 4 summary(CT)

lm(formula = YUN ~ GDP + ED + FDI + PI + LR + POP, data = MED)

Choosing Lag length > MED<-read.csv(”C://Users//user//Desktop//PhD Docs//MED.csv”,header=T)

> MED2¡-(log(MED))

> library(urca)

> library(vars)

> MED3¡-as.ts(MED2[,2:8],start=1979,stop=2015,freq=1)

> MED3

> VARselect(MED3,lag.max=10,type=”const”)

Testing Serial Correlation

>Box.test(lag=3,type=”Ljung-Box”,MED3[,1])

Cointegration Test

> # Conduct eigen(cointegrationtest):

> cointest¡-ca.jo(MED3,K=2,type=”eigen”,ecdet=”const”,spec=”transitory”)

> cointest

> summary(cointest)

ARDL Estimation

>library(dynlm)

>f<-dyn$lm(YUN ~ lag(YUN,-1)+ GDP

> lm(formula = dyn(YUN ~ lag(YUN, -1) + GDP))

> fy¡-dyn$lm(YUN ~ GDP + lag(GDP,-1) + lag(GDP,-2) + lag(GDP,-3) + ED + lag(ED,-1) + lag(ED,-2) + lag(ED,-3) + FDI + lag(FDI,-1) + lag(FDI,-2) + lag(FDI,-3) + PI + lag(PI,-1) + lag(PI,-2) + lag(PI,-3) + LR + lag(LR,-1) +lag(LR,-2)+ lag(LR,-3) + POP+lag(POP,-1)+ lag(POP,-2) + lag(POP,-3)+ lag(YUN,-1))

¿Plot(fy)

> summary(fy)

VECM Estimation

> MED<-read.csv(”C://Users//user//Desktop//PhD Docs//MED.csv”,header=T)

> MED

> MED2<-(log(MED))

> MED2

> library(urca)

> library(vars)

> VECM<-cajorls(cointest) # convert in vecm

> VECM

>rlm

> VECM.r3 <- cajorls(VECM2, r = 3)

> VECM.r3

Or >VECM(MED3,lag=3,estim=”ML”)

Coefficients of α,β and $\prod$ matrices

> ABP¡-VECM(MED3,lag=3,estim=”ML”)

> ABP

>coefA(ABP)

>coefB(ABP)

>coefPI(ABP)

RCODES FOR IRF and VD > fanchart(vecm.pred)

```
> fanchart(vecm.irf)
> plot(vecm.irf)
> vecm.irf¡-irf(vecm.level,impulse=”ED”,response=”YUN”,boot=FALSE)
> plot(vecm.irf)
> vecm.irf¡-irf(vecm.level,impulse=”FDI”,response=”YUN”)
> plot(vecm.irf)
> vecm.irf¡-irf(vecm.level,impulse=”FDI”,response=”YUN”,boot=F)
> plot(vecm.irf)
> vecm.irf¡-irf(vecm.level,impulse=”PI”,response=”YUN”,boot=F)
> plot(vecm.irf)
> vecm.irf¡-irf(vecm.level,impluse=”LR”,response=”YUN”,boot=F)
> plot(vecm.irf)
```

F-Test

```
¿ MED3.cointest.ols¡-cajools(cointest)
¿ summary(MED3.cointest.ols)
```

Testing Cointegration Relations test > plot(cx)

```
>adf.test(cx)
>cusum(MED3,central=1.6717,std.dev=0.5,decision.interval=1.5*0.5,se.shift=4*0.5,plot=T)
```

APPENDIX II: MACROECONOMICS DATA

Year	GDP (In billion USD)	ED(%)	FDI (Billion USD)	YUN(%)	PI (Billion USD)	LR(%)	POP (in million)
1979	6.23	45.3	1.35	16.1	0.113047	92.6	15.66
1980	7.27	48.1	1.09	16.2	0.178052	85.5	16.27
1981	6.85	48.6	0.21	16.4	0.15706	79.2	16.9
1982	6.43	54.5	0.2	16.7	0.140596	82.3	17.56
1983	5.98	62.7	0.4	16.6	0.125115	87.2	18.24
1984	6.19	58.6	0.19	16.8	0.122659	89.1	18.94
1985	6.14	70.6	0.47	16.9	0.155369	91.2	19.66
1986	7.24	65.8	0.46 17.1	0.157582	90.6	20.39	
1987	7.97	75.2	0.49	17.2	0.193607	91.7	21.14
1988	8.36	72.3	0.005	17.5	0.212636	86.5	21.9
1989	8.28	73.3	0.75	17.4	0.205652	79.1	22.67
1990	8.57	86	0.67	17.3	2.075833	79.6	23.45
1991	8.15	95.8	0.23	17.2	0.170954	81.3	24.24
1992	8.21	87.8	0.08	17.3	0.139101	77.3	25.04
1993	5.75	131.9	2.53	17.3	0.101291	81.3	25.84
1994	7.15	105	0.1	17.2	0.137911	77.3	26.63
1995	9.05	83.8	0.47	17.1	0.197389	80.9	27.42
1996	12.05	57.6	0.91	17.1	0.180734	81.2	28.19
1997	13.12	49.9	0.47	17.2	0.198585	82.3	28.94
1998	14.09	48.9	0.19	17.1	0.235254	84.9	29.7
1999	12.9	51.3	0.4	17.1	0.200165	86.8	30.48
2000	12.71	49.2	0.87	17.2	0.221007	92.5	31.29
2001	12.99	43.4	0.04	17.1	0.244021	76.3	32.13
2002	13.15	47.5	0.21	17.2	0.199056	79.2	33
2003	14.9	47	0.55	17.1	0.245644	88	33.91
2004	16.1	43.8	0.29	17	0.275031	76.3	34.83
2005	18.74	34.7	0.11	17	0.31692	78.2	35.79
2006	25.83	26	2	17	0.40389	81	36.76
8 2007	31.96	23.7	2.28	17	0.518351	82.4	37.75
2008	35.9	21.1	0.28	17.1	0.610939	86.9	38.77
2009	37.02	23.3	0.32	17.1	0.613535	91.5	39.82

APPENDIX III: VECM EQUATIONS

>VECM(MED3,lag=3,estim="ML")

	ECT	Intercept	GDP -1	ED -1	FDI -1
Equation GDP	-0.081618248	0.68478311	1.041681964	0.397612298	-0.0255150492
Equation ED	0.505343993	-3.18319209	-1.184821983	-0.436486222	0.0617421278
Equation FDI	3.923971897	-22.32580046	-2.498460158	-3.055893269	-0.5656133375
Equation YUN	0.182982408	-0.94611649	-0.024498641	0.051325819	0.0095041256
Equation PI	1.021564156	-6.19361034	-1.261771997	-0.393917645	0.0507913559
Equation LR	0.222520093	-1.41451676	-0.312736340	-0.349664394	0.0187072937
Equation POP	0.002655281	-0.01344679	0.002495062	0.000518749	0.0005101554

	YUN -1	PI -1	LR -1	POP -1	GDP -2
Equation GDP	0.927892307	0.005592738	-0.149172387	-31.83985118	-0.06385096
Equation ED	-2.458858892	-0.063671436	0.253072107	53.42811540	-0.60590768
Equation FDI	-12.045049342	-0.318886184	-2.609287260	-11.58130717	1.00925248
Equation YUN	-1.043433361	-0.036279051	0.086388632	4.34574033	-0.13132615
Equation PI	-3.154767765	-0.880367934	1.716372070	-41.19202133	-0.22026607
Equation LR	-0.945325481	-0.023009484	0.001318437	19.82500428	-0.60387525
Equation POP	-0.003571835	-0.001234026	0.002682860	-0.07785184	-0.01227370

	ED -2	FDI -2	YUN -2	PI -2	LR -2
Equation GDP	-0.224104318	-0.0265371095	0.87400097	0.0352681753	-0.13053551
Equation ED	-0.099116362	0.0414003710	-2.41348040	-0.0696685665	0.23082382
Equation FDI	-2.565540917	-0.7770933637	-16.26597627	0.7382019274	-3.31568913
Equation YUN	-0.039732813	0.0023852084	-1.04687305	-0.0072855733	0.34546763
Equation PI	-0.765554918	-0.0446435257	-7.02118374	-0.4202742965	1.02356119
Equation LR	-0.356291932	0.0133722519	-1.10386923	-0.0054200670	-0.25318525
Equation POP	-0.008762699	0.0003751496	-0.02562371	-0.0001803173	0.01723963

	POP -2	GDP -3	ED -3	FDI -3	YUN -3
Equation GDP	3.150141	-0.64496693	-0.292881979	-0.0092672922	0.69493831
Equation ED	29.359769	0.29177615	0.124253977	0.0178582429	-2.95338510
Equation FDI	191.920192	-17.61866858	-12.511462711	-0.2764977772	-10.53903805
Equation YUN	14.546972	-0.23969184	-0.084139451	0.0001179452	-1.00142492
Equation PI	104.904901	-3.72719186	-2.281483916	-0.0855291769	-8.64184149
Equation LR	23.867857	-0.44524744	-0.379317337	0.0133323195	-0.07150677
Equation POP	0.902117	-0.01077106	-0.003753171	0.0006440009	-0.02983720

	PI -3	LR -3	POP -3
Equation GDP	0.0325996781	0.519181892	9.4242655
Equation ED	-0.0315248454	-0.261408639	8.9361943
Equation FDI	0.7370746046	-4.059738566	453.9593321
Equation YUN	-0.0009152199	0.074985757	7.7478142
Equation PI	0.0256973561	-1.602523624	118.8526171
Equation LR	-0.0057323175	-0.051545825	-2.0467466
Equation POP	-0.0002525888	-0.004212471	0.5368143