UNIVERSITY OF RWANDA

COLLEGE OF BUSINESS AND ECONOMICS

SCHOOL OF ECONOMICS

Econometric Modeling for Consumer Price Inflation in East Africa Community: interdependency assessment and prediction guideline

Submitted by Davy UWIZERA

Reg No: 215032645

To the Department of Economics

In partial fulfillment of the requirements for the Master's Degree of Science in Economics

Degree Awarded by the University of Rwanda

Supervisor: Dr. Ildephone MUSAFIRI

Gikondo, June 2016

DECLARATION

I, Davy UWIZERA, hereby declare that this dissertation entitled "***Econometric Modeling for Consumer Price Inflation in East Africa Community: interdependence assessment and prediction guideline***" is my own research work and it has never been submitted anywhere for the award of any degree.

Davy UWIZERA

Signature of the graduate …… Date 17th June, 2016

ACKNOWLEDGEMENTS

I take this opportunity to thank my family and friends for the support throughout this master's program.

I am grateful for the college of business and economics faculty for the great knowledge acquired from them.

I appreciate the time and guidance of Dr. IldephonseMUSAFIRI throughout this thesis.

Abstract

The negative impact of high inflation on the economic health of country has pushed different countries to come up with economic strategies that might keep inflation rate at a low level. However, the positive trend in inflation rate continues to defy solution in Sub-Saharan Africa in recent years; leaving Sub-Saharan countries the option of building up strong predictive skills as they deal with the challenges of economic manipulations of indicators influencing inflation. In East Africa Community, where there is a suggestion of using one currency, inflation might be a big threat, accented with South Sudan a new member of East Africa which is among countries with the highest inflation rate in the world, as of May, 2016. In this research, we assess the interdependency of consumer price inflation rate of elderly members of East African Community, to look at the credibility of one currency in the area in terms of consumer price inflation. We also make an econometric modeling analysis that provides a guideline in consumer price inflation prediction of East Africa Community Member states. The prediction performance of Ordinal Least Square multiple linear model, Support Vector Machine, K Nearest Neighbor, Penalized Linear Models and Autoregressive Integrated Moving Average have been compared to assess the more reliable model in the area. The results of the study show that there is a strong interdependence of consumer price inflation rate in old members of East Africa Community, and there is no difference in the levels of inflation in these countries, giving an edge for one currency suggestion in the area. For the prediction side, Support Vector Machine turned out to be the best model outperforming all other models on accurate prediction in the whole area.

Keywords

Consumer Price Inflation, East African Community, Econometric modeling

JEL Classification Codes: C31, C51, E31, F00, F53

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ACRONYMS AND ABBREVIATIONS

ARIMA	: Autoregressive integrated moving average
BM.GSR.GNFS.CD	: Imports of goods and services (BoP, current US$)
BN.KLT.DINV.CD	: Foreign direct investment, net (BoP, current US$)
BX.KLT.DINV.CD.WD	: Foreign direct investment, net inflows (BoP, current US$)
BX.KLT.DINV.WD.GD.ZS	: Foreign direct investment, net inflows (% of GDP)
BDI	: Burundi
CPINF	: Consumer Price Inflation
CPINFBDI	: Burundi Consumer Price Inflation
CPINFKE	: Kenya Consumer Price Inflation
CPINFRWA	: Rwanda Consumer Price Inflation
CPINFUGA	: Uganda Consumer Price Inflation
CPINFTZA	: Tanzania Consumer Price Inflation
EAC	: East African Community
EU	: European Union
FR.INR.DPST	: Deposit interest rate (%)
FR.INR.LEND	: Lending interest rate (%)
FR.INR.RINR	: Real interest rate (%)
KEN	: Kenya
KNN	: K nearest neighbors
NE.CON.GOVT.KD.ZG	: General government final consumption expenditure (annual % growth)
NE.CON.GOVT.ZS	: General government final consumption expenditure (% of GDP)
NE.CON.PETC.ZS	: Household final consumption expenditure, etc. (% of GDP)
NE.CON.PRVT.CD	: Household final consumption expenditure (current US$)
NE.CON.PRVT.PC.KD.ZG	: Household final consumption expenditure per capita growth (annual %)
NE.IMP.GNFS.KD.ZG	: Imports of goods and services (annual % growth)
NE.IMP.GNFS.ZS	: Imports of goods and services (% of GDP)
PA.NUS.FCRF	: Official exchange rate (LCU per US$, period average)
PLM	: Penalized linear models
RDA	: Rwanda
SL.UEM.TOTL.FE.ZS	: Unemployment, female (% of female labor force)
SL.UEM.TOTL.MA.NE.ZS	: Unemployment, male (% of male labor force) (national estimate)
SL.UEM.TOTL.MA.ZS	: Unemployment, male (% of male labor force)
SL.UEM.TOTL.ZS	: Unemployment, total (% of total labor force)
SVM	: Support vector machines
TM.VAL.AGRI.ZS.UN	: Agricultural raw materials imports (% of merchandise imports)
TZA	: Tanzania
UGA	: Uganda
VIF	: Variance Inflation Factor

CHAPTER I

1. INTRODUCTION

1.1 Background

East Africa community (EAC) is made by five countries located in the Sub-Saharan Africa, Rwanda, Kenya, Uganda, Tanzania and Burundi, with South Sudan yet to join as the youngest state in the community. Since the beginning of the community, different agreements were made in between members to support economic development in the area by promoting different sectors such as easy trade, infrastructure etc. In fact, in Rwanda, Kenya and Uganda, it only requires a national ID to travel from one country to another, no need for a passport or visa application. Opening the doors for each other produced different opportunities for each country member. But it is also believed that this might have increased macro-economic interdependency between members of this community.

Even though the community has been doing fine economically on different things, inflation remains a threat for the recent years, South Sudan leading with 290.3%, followed by Uganda, Tanzania, Kenya, Rwanda and Burundi with 5.4%, 5.1%, 5%, 4.4% and 2.2 % respectively as of May, 2016 (IECONOMICS). With an old suggestion of having a common currency in EAC countries, inflation is one of the things that should be assessed carefully to understand the behavior of each country inflation rate and the relationship between them. This can be done by making a descriptive assessment on the behavior of inflation between these countries, the impact of specific factors affecting inflation in EAC, and a prediction analysis with different models for distributions heterogeneity assessment and reliable model selection.

Predicting and assessing inflation is very important for every nation in the world; it plays an important role in monetary policy. In every country, central banks try to pursue an intermediate money target and assess future inflation. Appropriate inflation prediction is important for central banks' policy making and strategies as the impact of monetary policy tends to level off with the economy between one to two years, Duisenberg (1999).

And since high inflation normally results in the negative impact on economic health of country, it is always good to plan ahead of time to ensure the best policy is chosen that will be effective and efficient. With respect to this, different studies have been conducted about information on the factors driving inflation pressures and future yield curve regarding inflation. Some of those researches include Mishkin (1988), Tzavalis and Wickens (1996) ,Jochum and Kirchgassner(1999) and Hansen (2011).

In this line we want to make an econometric analysis of consumer price inflation (CPINF) in EAC by making a descriptive assessment of interdependence of EAC CPINF and its prediction analysis by comparing the prediction performance of different models.

1.2 Statement of the Problem

Different researches have shown that high inflation has a negative impact on the economic growth of a country. Hence, keeping inflation at a low level has always been promoted by countries to ensure economic health, each year strategies are taken to manipulate the economic variables that might help in keeping a low inflation. In fact high inflation normally results in higher borrowing costs for businesses, risks of wage inflation, business uncertainty, falling of real incomes etc… Hence when keeping inflation at a low level fails, prediction analysis remains the best way to deal with inflation effect by planning ahead for any country.

This applies for communities and unions of countries which use the same currency too. However for this case, controlling inflation requires much more than for one country since the behavior of one country's economy is much likely to influence the economy of others due to the shared currency. The case of Ukraine in EU support the statements stated above.

With the shared vision of development in EAC countries, inflation is one of the things that might be looked at jointly once one country's inflation might be influencing other countries' inflation. Moreover, given the suggestion of using one currency in all country members of the community, we believe inflation is one of the major things that have to be looked at to assess the credibility of this suggestion. Hence in this line we would like to make an econometric analysis of CPINF by assessing its interdependency between countries making EAC as we demonstrate one of the best ways that can be used to predict it by comparing the prediction performance of different models commonly used in regression.

1.3 Objectives

1.3.1 General Objective

The general objective is to conduct an econometric analysis of CPINF in East Africa. In this line, we want to assess CPINF country interdependency in EAC and compare prediction performance of four models, using different features that will be clearly outlined in chapter 3 section 5, Variables Specification.

1.3.2 Specific Objectives

1) Make a dependency and trend analysis of CPINF in EAC countries
2) Assessing the variation of the impact of features considered on CPINF in EAC countries
3) Finding out the best model among OLS, SVM, KNN, PLM and ARIMA in predicting CPINF for each country in EAC
4) Finding out the best model among OLS, SVM, KNN, PLM and ARIMA in predicting CPINF weighted average of all EAC countries

1.3.3 Research questions:

1) How is the dependency and trend of CPINF in EAC countries?
2) How is the variation of the impact of features considered on CPINF in EAC countries?
3) What's the best model among OLS, SVM, KNN, PLM and ARIMA in predicting CPINF for each country in EAC?
4) What's the best model among OLS, SVM, KNN, PLM and ARIMA in predicting CPINF average of all EAC countries?

1.4 Scope of the study

The study has been conducted on five different countries which are members of EAC. It bases its analysis on the records stored at World Bank databank, covering a period of 25 years (1990-2015). We believe the period chosen gives accurate information of the case. Moreover, a period of 25 years is sufficient enough to assess potential major changes and inclinations.

The study will be conducted within the period of 5 weeks. Four weeks of field work for a qualitative understanding of the problem, one week retrieving and cleaning all required data from World Bank, two weeks of analyzing the data and two weeks of writing the final report.

1.5 Significance of the study

The results of this study will be prolific to different parties including the researcher himself. Given the desire of east Africa community to use same currency, the results of this study contribute in policy making for this particular case and give more insight into inflation interdependency in the participating countries with its possible application in implementing one currency in the area. Moreover, it will add to the existing knowledge about modeling CPINF in EAC.

To the researcher, the research has enabled him to gain more knowledge and skills in research area. The research will also help other masters' students and researchers who will be interested in writing their research work in the same field.

CHAPTER 2

2. LITTERATURE REVIEW

2.1 Introduction

This section explores existing theories and concepts related to the topic under the study. Theories are discussed respective to the proposed objectives in order to answer our research questions. It highlights the concepts behind the main indicator in the research and methods for evaluating our case study.

The majority of literature review will consist of the theory and the application of the methods used by different researches to tackle problems similar to the ones we have in this research. We will begin with our main indicator.

2.2Inflation

Despite different definitions associated with inflation it can be simply explained as a sustained general rise in price levels Johnson (1972). Usual temporal oscillations of prices make the inflation to be measured on a long time period that could offset the bias that can be associated with temporal rise of prices. Inflation normally has a negative impact on the economic health of the country, most of countries faced inflation problem in 1980's.

2.2.1 Sourced of Inflation

Inflation word widely has been on a positive trend for many years. Economists have been facing the challenge of reviewing pressures that spark large variations in levels of price. Different researches have been done and enormous literature review has been written about inflation.

According to Teriba and ajayi (1974), inflation causes can be put in three categories: monetarist, cost push theories and excess demand theories explanations.

2.2.2 Monetarist Inflation

Monetarist view of inflation traces its strong roots from the quantity theory which is based on theories developed in early decades of the twentieth century by different great economists like Fisher Irvin, Alfred Marshall, A.C Pigou and Keynes.

In recent years, there has been a renovation of the quantity theory in what is known as monetarism, spearheaded by different economist such as Milton Friedman, Karl Brunner, Allen Meltzer and Philip Cagan, Humphrey (1974). The monetarists argue that the important determinant of aggregate spending is money supply; looking at increases in the money supply as important causes of inflation.

2.2.3 Cost Push Inflation Theories

The theoretical concepts of this view are derived from the studies of institutional contexts by which prices and wages are regulated. The determination of wages is considered to have a vital importance in understanding inflationary pressures. One of the target subjects is the role of trade union; it is of much interest due to its impact on inflation and national income money value. Using all-out strike (protesting by refusing to work), trade unions can pressurize institutions to increase wages for labor. However, this is only likely to happen when they have strong control over labor supply.

Increase in wages normally results in the increase in the cost of production when the marginal increase of labor production with respect to wage can't offset the cost incurred over the former margin of wages. This discrepancy is usually passed on to end consumers in the form of higher prices, resulting in the cost push inflation. The mark-up for profits can also be regarded as another cost push inflation source through profit seeking incentives.

2.2.4 Excess Demand Theories

Excess demand is when the demand for goods and services and their substitutions in the market surpass them. This leads to higher prices than usual of these goods and services as interested buyers compete to win the little that is available on the market, hence creating edge for the sellers to increase prices. The concept of demand and supply with the associated price behavior is very well outlined in Keynesian analysis.

2.3 Effects of Inflation

It is argued that sometimes that inflation can't be avoided during growth, and so many developing countries usually experience some sort of high inflation. However, inflation history keeps a record of many countries that experienced growth without suffering from inflation and those who suffered from inflation without clear growth, Friedman (1973).

Hinshaw (1972) emphasized that inflation is not good for economic health and should be avoided by giving different reasons stated below:

- Money usefulness impairment. In fact it can completely destroy it once the rate turns out to be too high, resulting in a failure of monetary term contracts and the matching of units of money and its expected value in units of goods and services.

- It hinders distribution of wealth and income in an effective way. Given fixed wages and their holders such as pensioners and salary earners, the value of their income is valued in terms of the goods they can purchase with it. Hence, since the wages are not adjusted all of a sudden, inflation devalue their income. Moreover, it makes lenders to lose while borrowers gain.

- Balance of payment crisis. Contraction of sales and competitiveness are weak when there are high prices and costs pressures. Hence, people are reluctant to buy goods from an economy with high inflation and on the other hand willing to sell them in that kind of economy.

- Finally, it distorts and destabilizes the economy as a whole. People are motivated to buy and accumulate some inventories speculating on the inflation state on a specific period. On the other hand there is a misallocation of resources due to excessive bidding real capital assets and bad effect on output and depressing. However, since inflation motivates capital investment it is argued by different economists that it can sometimes serve as a sparker for effective economic development.

2.4 Prior work on related researches

Different researches have been done on analyzing and predicting inflation. Below we outline different researches conducted in this area, by stating methods used and problem solved.

Hassan et al, (2012), conducted a research on the prediction of UK rate of inflation and changes in its direction using univariate and multivariate singular spectrum analyses. Major features used were consumer price indices, real-time chain-weighted gross domestic product price index series. The study compared the out-of sample prediction results of univariate and multivariate spectrum analyses with other models such as the activity-based non-accelerating inflation rate of unemployment Phillips curve, auto-regressive AR(p) model, the dynamic factors model and random-walk models and naive forecasting method, to find out that multivariate singular spectrum analysis outperforms all other models used in the analysis.

Scheide et al(2000), conducted a research on the prediction of inflation in Euroland; their general objective was test if price gap has an impact on consumer price inflation. But they also looked at other factors such as raw material prices and unit labor costs, the results showed that these feature have also some explanatory power.

Ferdinand(2011), conducted a research on the prediction of inflation in Ghana using ARIMA model. Monthly inflation variable was chosen and Box-Jenkins approach was used.

Pourgerami and Maskus (2002) analyzed the effects of inflation on the predictability of price changes, the researchers found out that the higher the inflation the higher the difficulty in predicting prices.

Tilmann and Thordis (2010), conducted a research on the prediction of inflation by implementing an ensemble prediction system and post-process techniques. In this research surveys were interpreted as ensemble forecasts like ensemble models in weather forecasting.

Fama (1975), conducted also a research on predicting inflation using short-term interest rates.

Levi and Makin (1980), conducted a research on inflation uncertainty and the Philips curve. Due to the coincidence of high rates of inflation and unemployment rates, the researchers try to make a rethinking of the Phillips curve.

Etuk (2012) conducted a research on inflation prediction, using SARIMA model the researcher estimated inflation rate in Nigeria.

Stock and Watson (2001), conducted a research on output and inflation forecasting. The research examines old and new evidence on the predictive power of asset prices for inflation and real output growth.

Wright (2008), worked on inflation prediction using Bayesian model averaging. The researcher showed how the used beats equal-weighted averaging forecast of linear regression models in predicting US inflation. Gary and Dimitris (2012), looked at forecasting inflation based on generalized Philips curve using model averaging.

Carstensen and Hawellek (2003), conducted a research of inflation forecasting in Germany. The researcher analyzes the forecasting ability of the term structure for future inflation.

Baghestani and AbuAl-Foul (2010) worked on prediction inflation under asymmetric loss. The researcher compares Federal Reserve and private inflation forecasts in terms of directional accuracy in the period of 1983-2002. It finds that Federal Reserve forecasts are more accurate.

Onder (2004), conducted a research on inflation forecasting in Turkey. The researcher analyzed the predictive performance of the Philips curve in a high inflation emerging market country, with ARIMA model as an alternative model. The results showed the Philip curve outperforms forecasts based on other macroeconomic variables.

Bokil and Schimmelpfennig (2006), conducted a research on Pakistan's inflation. The researchers used three different approaches: leading indicators based model with broad money growth and private sector credit, univariate time series model and vector autoregressive model.

Thomakos and Bhattacharya et al (2005), did a research on India's inflation prediction. The researchers assessed the predictive performance of linear models, ARIMA and VAR models using root mean square error RMSE. Their analysis results showed that VAR outperforms other models in India inflation forecasting on the selected time period.

Hafer and Hein (1990), conducted a research on inflation prediction using interest rate and time series models. The results show that univariate time series models give better results than interest-rate based models.

Stock and Watson (2010) looked at modeling inflation after the crisis focusing on the United States. Despite the observed fall of the rate of price inflation during this time, the researchers argue that it is hard to get this observation into a useful forecasting equation due to different reasons such changes in Fed policy and credibility. They proposed a parametrized model where inflation deviation from a stochastic trend reacts stably on the unemployment recession gap.

Downes et al (1991) demonstrated inflation modeling in a small economy, using cointegration approach, to separate long run inflation from the short run one. They found that wages, productivity and unemployment and the price of traded goods are significant variable in estimating inflation.

Aleksander(2000) presented two aggregate models of inflation in Poland, the first model was Bayesian based used to analyze turning points, and the second one VAR structured to identify long-run relationships between wages, prices, labor productivity and unemployment

Siliverstovs and Bilan (2005) modeled Ukrainian inflation by analyzing its relationship with money growth, wage growth, and devaluation expectations, they found that devaluation expectations is the most important factor driving high price levels, while money supply growth has negligible impact on inflation.

2.5 Gap

In the explored existing literature, especially in Africa, they are a few researches which might have looked at the impact of neighboring countries' inflation rate trends on a specific country's inflation; even as they assess various effects of other economic variables on it rather than focusing directly on money supply.

To our own knowledge, this research is the first which is specifically going to fill this research gap in East Africa region. It assesses the credibility of one currency in the area in terms of consumer price inflation (CPINF) by conducting an interdependence analysis of CPINF in EAC member states as it demonstrates one of the best ways that can be used to predict them by comparing the prediction performance of different models commonly used in regression. The prediction assessment is done using different economic variables rather than focusing directly on money supply.

CHAPTER III

3. METHODOLOGY

3.1 Introduction

In this chapter we present different methods that will be used in this research. It covers research design, model specification, conceptual frame work, data and data source, and data tools used. The application and details of most of the models discussed below can be found in Marwara (2013).

3.2 Research Design

The research is quantitative and qualitative. We will achieve our main objective using statistical data from World Bank databank. Qualitative analysis was only done for personal understanding of the researcher and broad insight of problems solved in this research. The methods used for achieving our specific objectives are listed below.

Using Pearson correlation computations and heat-map we will quantify and visualize the relationship between inflation rate of countries in EAC. Each country series will be assessed against other countries' CPINF. To get an insight about the inflation rate pattern in EAC countries, the test of significance will be applied on correlation coefficients to find out Pearson significant correlations between considered countries. Further assessment of interdependency will be done by carrying out bivariate normality test, variance homogeneity, means equality, and Granger causality tests. Granger causality tests will be used to assess further the relationship between CPINF series of a country and lagged CPINF series of other countries. Quantile computations and measures of central tendency and Man-Kendall trend test, will be used for trend analysis.

We will use OLS to assess the marginal impact of features used on CPINF. This will be done by estimating and interpreting coefficients of features used to estimate CPINF using OLS. Features will be selected from the list of independent variables listed in variables specification section, using VIF and linear stepwise regression selection with backward and forward direction.

To assess the best model among OLS, SVM, KNN, PLM and ARIMA in predicting CPINF for each country in EAC, we will use mean absolute error and root mean square error of the

predictions results of the models stated above for each country; features used for prediction will be selected using the method described above, variance inflation factor stepwise and stepwise regression.

And lastly, a weighted average of inflation rate for all countries making EAC will be computed and we will assess the best model among OLS, SVM, KNN, PLM and ARIMA in predicting weighted average CPINF as stated in three, given this kind of scenario. The weight used for inflation rate average computation, will be the ratio of the time series mean over the standard deviation.

3.3Conceptual Frame Work

In section we present econometric models used and the algorithm behind it. Econometric models are the implementation of statistical and artificial intelligence models in assessing the relationship and pattern between in different economic variables and to help in predicting and classifying them. Models presented below are used in pattern detection and relationship assessment Wang et al (2010).

3.3.1 OLS Multiple Linear Model

It models linear relationship between the dependent variable and the independent variables by fitting a line of estimates which minimizes the sum of square error in the whole data set of independent variables against the dependent variables. It makes a number of assumptions.

Let ε be the residuals and X the independent variables and Y dependent variable.
Assumptions:

- $E(\varepsilon|X=x)=0$
- $(X_i, Y_i), i=1,2,...,n$ are independent and identically distributed
- $(E(Y^4)<\infty, E(X^4)<\infty)$, No very large outliers
- $E(\varepsilon_t, £_{t-i})=0$, No autocorrelations
- ε are normally distributed with mean 0 and variance $\pm^2$

Estimation:

Let Y be a vector of our dependent variable, X a matrix of independent variables, B a vector of coefficients and E error terms as illustrated below.

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} X \begin{bmatrix} 1 X_{11} X_{21} \dots X_{n1} \\ 1 X_{12} X_{22} \dots X_{n2} \\ . \quad . \quad . \quad . \\ . \quad . \quad . \quad . \\ . \quad . \quad . \quad . \\ 1 X_{1p} X_{2p} \dots X_{np} \end{bmatrix} B = \begin{bmatrix} 0 \\ 1 \\ . \\ . \\ . \\ n \end{bmatrix} E = \begin{bmatrix} £_0 \\ £_1 \\ . \\ . \\ . \\ £_n \end{bmatrix}$$

Then, the multiple linear model is of the form, $Y = XB + E$ and below is the process taken to estimate B by minimize the sum square error,

$$Y = XB + E$$

$$E = Y - XB$$

$$E^2 = (Y - XB)^2$$

$$\frac{\partial E^2}{\partial B} = \frac{\partial (Y - XB)^2}{\partial B}$$

$$= \frac{\partial (Y^2 - 2B^{'} X^{'} Y + B^{'} \mathrm{X}^{'} \mathrm{XB})}{\partial B}$$

$$\rightarrow -2X^{'}Y + 2X^{'}XB = 0$$

$$\rightarrow X^{'}XB = X^{'}Y$$

$$\rightarrow B = (X^{'}X)^{-1} X^{'}Y$$

3.3.2 Penalized linear models

Different from OLS as discussed above these models apply constrained minimization. That is, $= \arg\min \ (\sum_{i=1}^{N} (y_i - (X_{ij} \ _i)^2)$, subject to $K(\ ) \le t$. Estimates are estimated by minimizing sum square error subject to a penalty on the model coefficients, $K(\ )$. Below is the table of commonly used penalized linear models and the associated penalty.

Table 3. 1: PLM Models and their Associated Penalties

Method	Ridge	Adaptive LASSO	Elastic net	LASSO
penality	$\sum_{j=1}^{p}(\ \ _j^2) < t_1$	$\sum_j^p(\|\ \ _j\|/\|\ \ _j\|) < t$	$\sum_{j=1}^{p}\|\ \ _j\| < t_1 \; and \sum_{j=1}^{p}\ \ _j^2 < t_2$	$\sum_{j=1}^{p}\|\ \ _j\| < t$

Taking example of ridge, below is the minimization of the sum square error subject to the penalty associated to ridge.

Assumptions:

- $E((Xj)_{1\le j\le p-1}) = 0$
- $length((Xj)_{1\le j\le p-1})=1$

$SSE_a(\ \ ) = \sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p-1} X_{ij}\ \ _j)^2 + a\sum_{j=1}^{p-1}\ \ _j^2$, hence using langrage multiplier,

$$\frac{\partial}{\partial\ \ _l} SSE_a(\ \ ) = -2\langle Y - X\ \ , Xl\rangle + 2a\ \ _l$$

$$\rightarrow -2\langle Y - X\ \ , Xl\rangle + 2a\ \ _l = 0, 1 \le l \le p-1$$

$$\rightarrow -Y^tX + \hat{\ }_a^t(X^tX + aI) = 0$$

$$\rightarrow\ \ _a(= X^tX + aI)^{-1}X^tY$$

3.3.3 Support Vector Machines

It is a kind of supervised linear model used for classification and regression; originally it was designed to maximize the margin between classes, Boser et al (1992).

Algorithm by Gunn(1998),

Linear regression:

Consider estimating set of data $S = \{(x^i, y^i)\}$, $i = 1, 2, ..., k$ $x \in$ ^{n}, $y \in$, with a linear function $f(x) = \langle \mu, x \rangle + b$ with $\langle ., . \rangle$, μ, x and b meaning as stated in section 3.4. The support vector machine is the regression obtained by minimizing $¤(\mu, ¶) = \frac{1}{2} \| \mu \|^2 + C \sum_i (¶_i^- + ¶_i^+)$, where C a pre-specified value and $¶_i^-, ¶_i^+$ are slack variables which represent lower and upper constraints. Using a loss function $L_£(y) = \begin{cases} 0 & \text{for} \quad | f(x) - y | < £ \\ | f(x) - y | - £ & \text{otherwise} \end{cases}$ the optimal solutions is given by,

$$max_{~,~^*} W(~, ~^*) = max_{~,~^*} - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} (~_i - ~_i^*) (~_j - ~_j \langle x_i, x_j \rangle + \sum_{i=1}^{k} ~_i (y_i - £) - ~_i^* (yi + £)$$

, with constraints $0 \le ~_i, ~_i^* \le C, i = 1, 2, ..., k$ and $\sum_{j=1}^{k} (~_i - ~_i^*) = 0$, where $~_i$ and $~_i^*$ are the Langrage multipliers, solving for the above.

optimal μ and b are given by $\sum_{j=1}^{k} (~_i - ~_i^*) x_i$ and $-\frac{1}{2} \langle \bar{\mu}, (x_r + x_s) \rangle$.

Non-linear regression:

Is given by the regression function $f(x) = \sum_{SVs} (\bar{~}_i - \bar{~}_i^*) K(x_i, x) + \bar{b}$ where,

$\langle \bar{\mu}, x \rangle = \sum_{i=}^{k} (~_i - ~_i^*) K (x_i, x_j)$ and $-\frac{1}{2} \sum_{i=1}^{k} (~_i - ~_i^*)(K(x_i, x_r) + K(x_i, x_s))$, where $K$is the kernel function.

3.3.4 K-Nearest Neighbors

It is model used in classification and regression. Given a number of neighbors, the algorithm works by classify the target point by referring to the closest neighbors occurrence of a specific label Sameer (2005).The difference between the points can be calculated using any valid mathematical formulas normally used for distance calculation between two points Colannino and Toussaint (2005).Neighbors (K) can be predefined by the user or the algorithm can even use local density points Shi and Albada (2007). Choosing a small K can result in modeling the noise and a big K can consequently make the algorithm consider different elements of different classes

as one class. Hence it is very crucial to understand how to set K, when using K-nearest neighbors Daniel and Yeung (2006).

K nearest neighbors have been successful in solving different problems, classification and regression challenges. It is a non-parametric method and below is how the algorithm works for weighted KNN using backward elimination Colin and Peter (2010):

- Get the training data
- Get testing data
- Set K to some value
- Normalize data in range 0 to 1
- Apply Backward Elimination
- For each testing example in the testing data set:

  Find the K nearest neighbors in the training data set based on the Euclidean distance

  Predict the observation by finding the average of observations or estimates associated with the K nearest neighbors
- Calculate the accuracy as,

  Accuracy using R-squared = (SST-SSE)/SST as usual

3.3.5ARIMA

In simple explanation, these are time series models used to predict a variable based on its lag values and errors. The detailed explanations and mathematics behind ARIMA models can be found in Ky (2007).

3.4 Model Specification

In this section we outline the econometric model specification of the models involved in our research as discussed in the preceding section.

OLS:

$Y=BX+\varepsilon_i$, where y is the independent variable, B a vector coefficients minimizing the difference between the fitted observed values with X using linear approximation, X a matrix of independent variable values and ε_i.

SVM:

$Y=\langle \mu, X \rangle + b$, where $\langle .,. \rangle$ is the dot product, μ is a measure of flatness, X a matrix of input pattern and b a constant

KNN:

$f(x)=\frac{1}{K}\sum_{i}^{l} f^{'}(x_i)$, where 1 to $_l$ is the indices of x_i nearest neighbors, $f^{'}(x_i)$ observations or estimates associated to nearest neighbors and K number of neighbors considered.

PLM:

$Y = \quad X + £$, $\quad = \arg\min\ (\sum_{i=1}^{N}(y_i - (X_{ij}\ _i)^2)$, subject to $K(\ ) \le t$. are the coefficients satisfying the constraints and X independent variables values used to estimate Y.

ARIMA:

$y_t = a + \sum_{i=1} b_i y_{t-i} + \sum_{j=0} c_i £_{t-j}\ i = 1, 2, \ldots$ where y_t, $y_{t-1}, £_{t-j}$, b and c, and a are: series, series lags, errors and errors lags, coefficients associated with series lags and errors then a a constant, normally regarded as a drift.

3.5 Variables Specification

In this section we introduce variables to use in our research and clarify between independent variables and dependent variables in our research.

Our dependent variable is the CPINF and our initial independent variables are list in the table below:

Table 3.2: List of Independent Variables

Variable	Abbreviation
Household final consumption expenditure, etc. (% of GDP)	NE.CON.PETC.ZS
Household final consumption expenditure per capita growth (annual %)	NE.CON.PRVT.PC.KD.ZG
Household final consumption expenditure (current US$)	NE.CON.PRVT.CD
Imports of goods and services (annual % growth)	NE.IMP.GNFS.KD.ZG
Imports of goods and services (BoP, current US$)	BM.GSR.GNFS.CD
Agricultural raw materials imports (% of merchandise imports)	TM.VAL.AGRI.ZS.UN
Imports of goods and services (% of GDP)	NE.IMP.GNFS.ZS
Official exchange rate (LCU per US$, period average)	PA.NUS.FCRF
Deposit interest rate (%)	FR.INR.DPST
Real interest rate (%)	FR.INR.RINR
Lending interest rate (%)	FR.INR.LEND
Unemployment, male (% of male labor force)	SL.UEM.TOTL.MA.ZS
Unemployment, female (% of female labor force)	SL.UEM.TOTL.FE.ZS
Unemployment, male (% of male labor force) (national estimate)	SL.UEM.TOTL.MA.NE.ZS
Unemployment, total (% of total labor force)	SL.UEM.TOTL.ZS
General government final consumption expenditure (% of GDP)	NE.CON.GOVT.ZS
General government final consumption expenditure (annual % growth)	NE.CON.GOVT.KD.ZG
Foreign direct investment, net inflows (BoP, current US$)	BX.KLT.DINV.CD.WD
Foreign direct investment, net (BoP, current US$)	BN.KLT.DINV.CD
Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS

For each regressed country's consumer price inflation the community, we will also add consumer price inflation of remaining countries as independent variables too. From the independent variables listed above, VIF and Stepwise selection will be applied during the modeling process for final features selection.

3.6 Data and Data source

Our study will focus on CPINF rate and the list of independent variables as listed in conceptual framework as recorded by the World Bank. The chosen data is yearly from 1990 to 2015. The range is long enough to assess trends and variation in the series; it was chosen using the domain knowledge of the researcher and World Bank visualization platform, which showed the range to be more consistent in all countries making EAC than in earlier years.

3.7 Data analysis tools

In order to make analysis of data, figures, tables, graphs, charts these tools and procedures will be involved in our research:

R-programming: statistical programming language that provides a wide range of basic and advanced data analysis capabilities. This statistical programming language will also be used to build an automatic program for our research computations.

Excel: a quantitative analysis comprehensive and flexible software program. It's used to generate tabulated reports, charts, and plots of distributions and trends, as well as generate descriptive statistics and more complex statistical and mathematical analyses.

CHAPTER IV

4. DATA ANALYSIS, FINDINGS AND RESULTS

4.1 Introduction

In this chapter we explore different findings and results drawn from different analyses made on our data set to answer our research questions.

4.2 Dependency and trend analysis of consumer price inflation in EAC

In this section we analyze the dependency and trend of CPINF in EAC members. The dependence of CPINF in these countries could signal a possibility of common policies that can jointly be used to control CPINF in all the countries once there is an implementation of one currency in the area. Weak dependence would imply the opposite.

Trend analysis will give us more insight about the variation and general direction of CPINF in EAC.

4.2.1 Interpolation of the missing values

Before analyzing our data we found that we had some missing values in our data set, hence we used linear interpolation to replace the missing values with interpolated values from available series. Below are the missing value maps of our data before and after interpolation for CPINF.

Missing Values Maps:

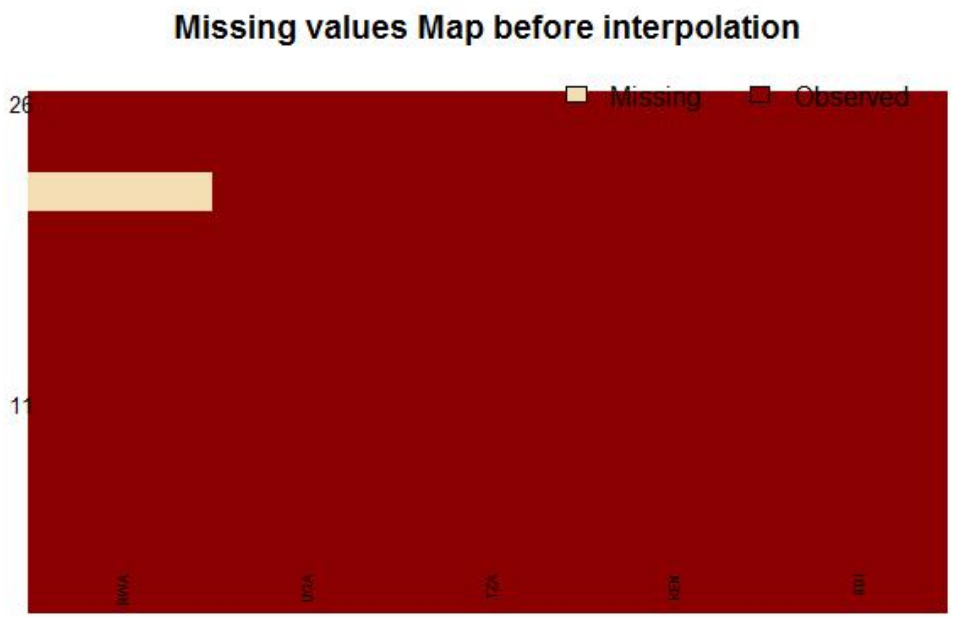

Figure 4.1: Missing Values Map before Interpolation

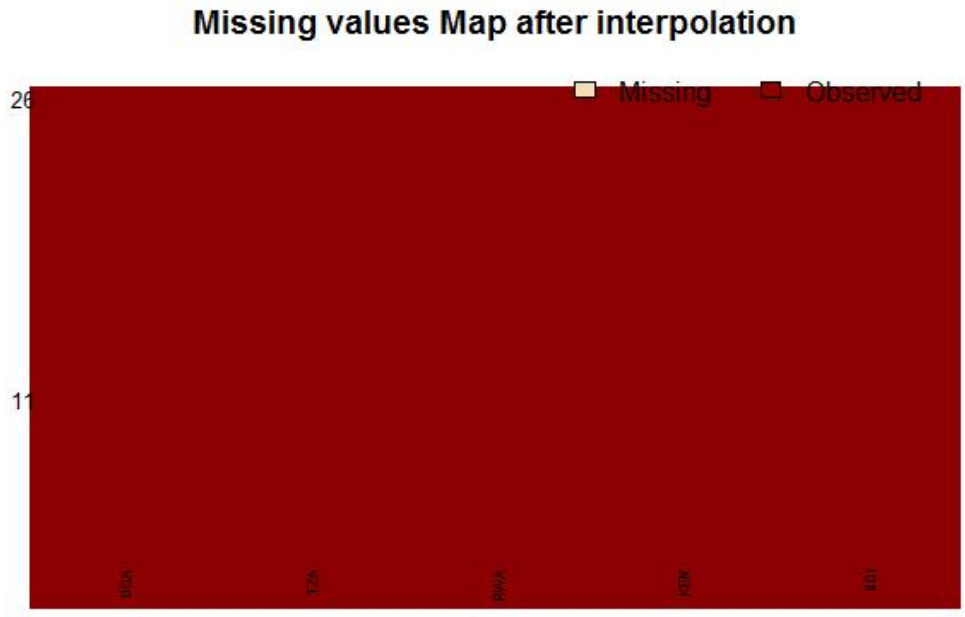

Figure 4.2: Missing Values Map after Interpolation

Interpolation was used in other analyses done in this research where missing values were encountered; mean replacement for missing values was also used to deal with missing values where the majority of data was present.

4.2.2 EAC Countries CPINF Interdependency Analysis

In this section will carry out the analysis of relationship of CPINF between EAC countries to find out how dependent and independent they are of each other. We will use Pearson Correlation coefficients and test their significance, bivariate normality test, variance homogeneity and means equality, and Granger causality tests.

4.2.2.1 Pearson Correlation Coefficients and their Significance

Using Pearson method, correlation values were computed and their associated p-values, this helped us to assess CPINF relationship between EAC countries and how it varies between countries.

Table 4.1: EAC CPINF Correlations Coefficients

	Burundi	Kenya	Rwanda	Tanzania	Uganda
Burundi	1	0.05	0.36	0.2	-0.15
Kenya	0.05	1	0.57	0.54	0.38
Rwanda	0.36	0.57	1	0.41	0.29
Tanzania	0.2	0.54	0.41	1	0.52
Uganda	-0.15	0.38	0.29	0.52	1

Table 4. 2: P-values of EAC CPINF Correlations Coefficients

	Burundi	Kenya	Rwanda	Tanzania	Uganda
Burundi		0.8083	0.0716	0.3175	0.4677
Kenya	0.8083		0.0022	0.0047	0.059
Rwanda	0.0716	0.0022		0.037	0.1483
Tanzania	0.3175	0.0047	0.037		0.0067
Uganda	0.4677	0.059	0.1483	0.0067	

The results show that they are significant correlation of CPINF between EAC countries. Some countries are more correlated with others while others seem to be less responsive to trends and variation in other countries' CPINF. Below is the heat map of the computed correlations.

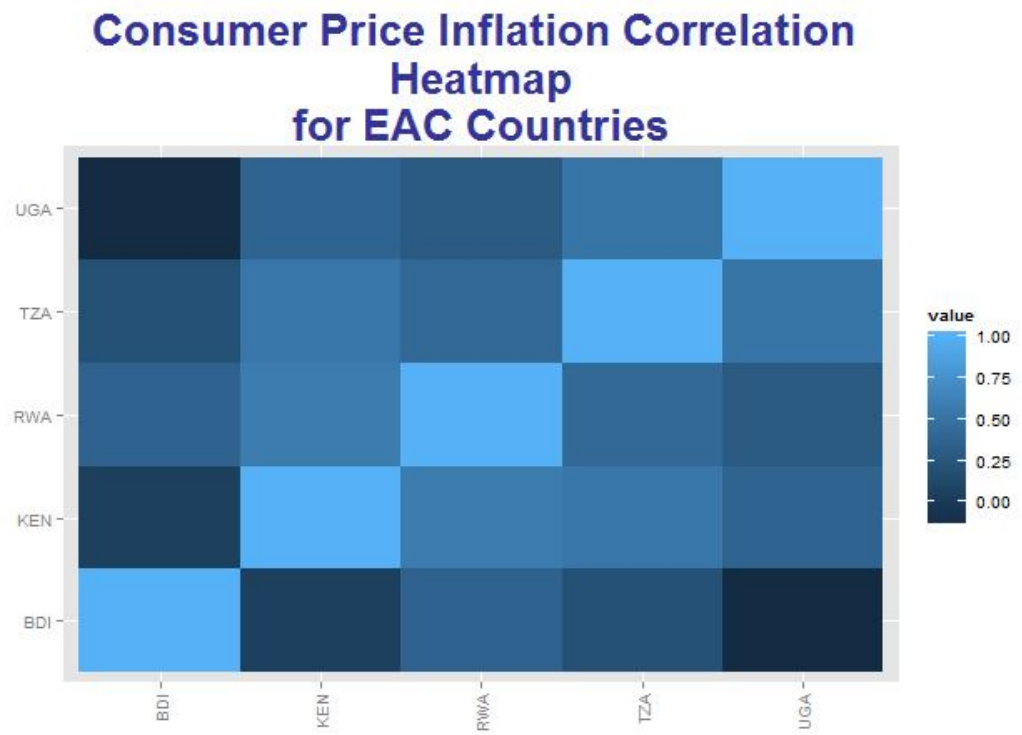

Figure 4.3: Visualization of CPINF Correlation Coefficients in EAC

At 5% level of significance,

Kenya CPINF has a significant positive relationship with the one for Rwanda and Tanzania, with Rwanda having the most significant correlation.

Rwanda CPINF has a significant positive relationship with the one for Kenya and Tanzania, with Kenya having the most significant correlation.

Tanzania CPINF has a significant positive relationship with the one for Kenya, Rwanda and Uganda, with Rwanda having the most significant correlation. Tanzanian CPINF is the one that has a significant correlation with so many countries.

Uganda CPINF rate has a significant positive relationship with Tanzania only, making it to follow Burundi in EAC countries whose CPINF rate is less correlated with other countries in the community. Burundi has no significant relationship of CPINF with other countries of EAC at 5 % level of significance.

Below is the histogram of one minus the p-values of correlation the coefficients computed using Pearson.

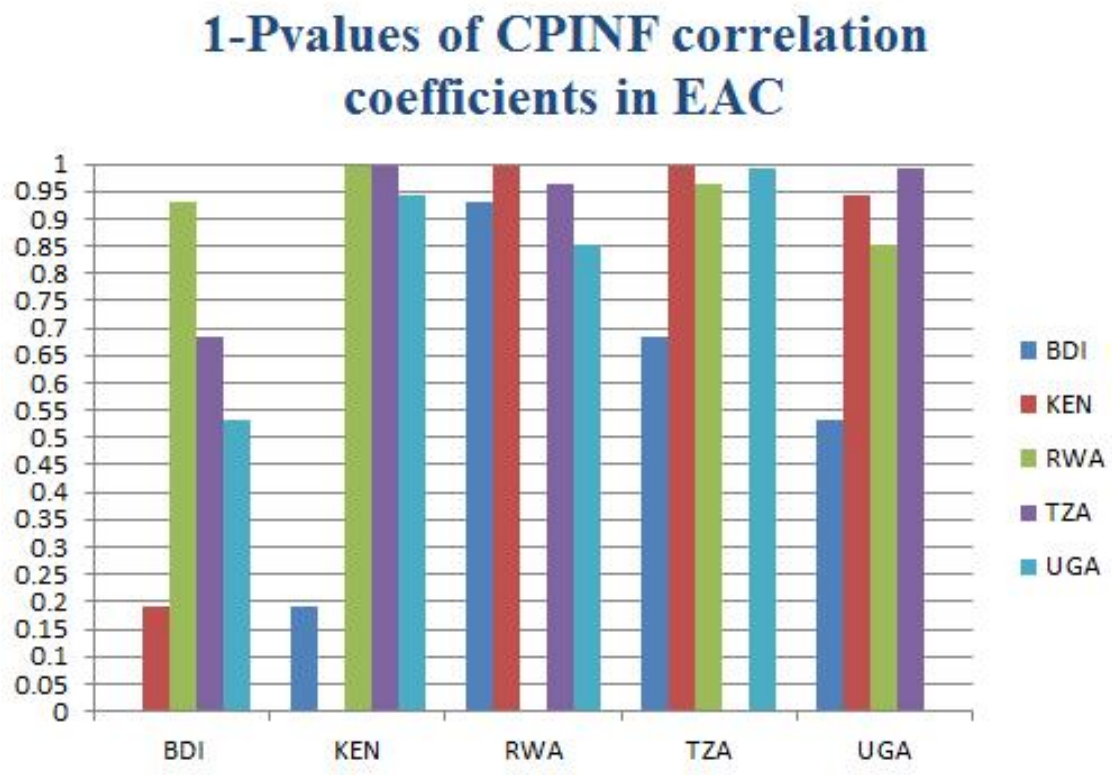

Figure 4.4: Visualization of the Significance of CPINF Correlation Coefficients in EAC

4.2.2.2Bivariate normality test

The absence of Pearson significant correlation coefficients is mistakenly taken as the absence of dependence between variables but it's not always the case. Hence to further test the interdependency of EAC countries CPINF rate, we have used more test rather than only relying on Pearson correlations.

Two random variables X and Y are said to be jointly normal if they can be expressed in the form

$$X = aK + bT,$$

$$Y = cK + dT,$$

where K and T are independent normal random variable.

Applying the bivariate normality test on the CPINF of EAC countries gives the results below,

Table 4. 3: P-values for Tests of Bivariate Normality

	Burundi	Kenya	Rwanda	Tanzania
Burundi				
Kenya	0.2402			
Rwanda	0.4519	0.3975		
Tanzania	0.2568	0.07014	0.2062	
Uganda	0.01304	0.005391	0.1152	0.01032

The results reveal further insight in the interdependency of CPINF in EAC. Ugandan CPINF which was among the least correlated with the one of other countries, turns out to be the one that is the most jointly normal distributed with other countries in EAC. Moreover, Burundi which had no significant correlation coefficient with any country has a jointly significant normal distribution with Uganda at 5% level of significance.

Below is the histogram of one minus p-values of bivariate normality test on EAC CPINF.

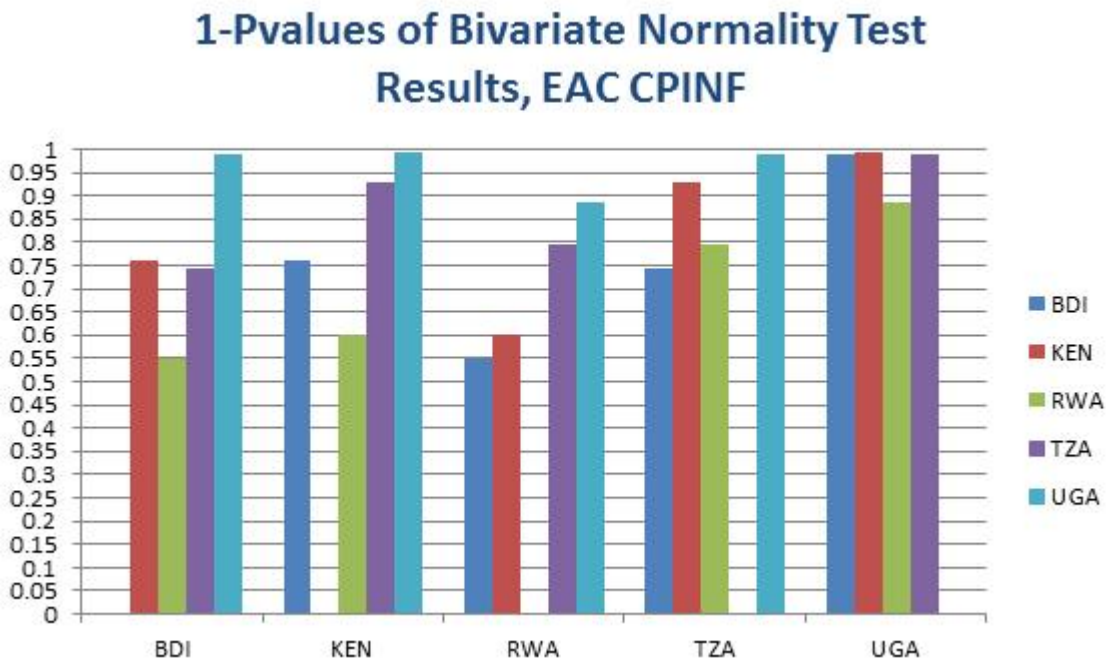

Figure 4.5: Visualization Bivariate Normality Test Significance for EAC CPINF

4.2.2.3 Variance Homogeneity and Means Equality Test

The variance homogeneity is used to test if the variance of groups considered is similar or the same. Once the homogeneity is significant between the series, the samples are assumed to have been drawn from the same distributions. The same applies to means, once the means are not statistically significant, they are assumed to have been drawn from the same population. We have used Fligner-Kileen and ANOVA test to test EAC CPINF variance homogeneity and means equality respectively.

Fligner-Kileen Test:

```
        Fligner-Killeen test of homogeneity of variances

data:  dataset and groups
Fligner-Killeen:med chi-squared = 2.8739, df = 4, p-value = 0.5791
```

Given the p-value obtained, we fail to reject the null hypothesis and conclude that there is homogeneity in the variances of CPINF in EAC countries.

ANOVA Test:

```
Analysis of Variance Table

Response: dataset
Df  SumSq Mean Sq F value Pr(>F)
groups      4   595.9 148.984  1.7831 0.1363
Residuals 125 10443.9  83.552
```

Given the p-value obtained, we fail to reject the null hypothesis and conclude that the difference between the means of CPINF in EAC countries is not statistically significant. This implies that the levels of CPINF in EAC are statistically indifferent at 5% level of significance.

4.2.2.4 Granger Causality Test

Granger causality test, assess the ability of lag information of one series to provide predictive information about another series. Below is the table of all EAC countries CPINF and the corresponding F-statistic and P-value.

Granger causality test, assess the ability of lag information of one series to provide predictive information about another series. Below is the table of all EAC countries CPINF and the corresponding F-statistic and P-value.

Table 4.4: Granger Causality F-statistic and P-value

	F-statistic	p-value
KEN -> BDI	0.46	0.638
RWA -> BDI	1.765	0.198
TZA -> BDI	0.928	0.413
UGA -> BDI	0.214	0.809
BDI -> KEN	2.224	0.136
RWA -> KEN	2.221	0.136
TZA -> KEN	5.75	0.011
UGA -> KEN	11.786	0.000
BDI -> RWA	9.669	0.001
KEN -> RWA	1.066	0.364
TZA -> RWA	0.047	0.954
UGA -> RWA	0.41	0.669
BDI -> TZA	2.93	0.078
KEN -> TZA	10.033	0.001
RWA -> TZA	0.423	0.661
UGA -> TZA	4.603	0.023
BDI -> UGA	0.423	0.661

KEN -> UGA	0.199	0.821
RWA -> UGA	3.878	0.039
TZA -> UGA	0.307	0.739

The test results show that at 5% level of significance, Tanzanian CPINF granger-causes Kenyan's, Ugandan's granger-causes Kenyan's, Burundian's granger-causes Rwandan's, Kenyan's granger-cause Tanzanian's and Rwandan's granger-causes Uganda at a lag length of 2.

Each EAC country member's CPINF rate series granger-causes at least one country's CPINF in the community at lag length 2 and 5 % level of significance.

These results and other relationship tests computed above show that, there is an interesting EAC CIPINF interdependence within country members.

4.2.3 EAC Countries CPINF Trend Analysis

4.2.3.1 Measures of Central Tendency

The table below shows summary of measures of central tendency of CPINF in EAC.

Table 4.5: Summary of CPINF Measures of Central Tendency in EAC

Country	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Std.Dev
Burundi	-1.371	6.551	9.461	11.448	14.521	31.112	8.093
Kenya	1.554	6.617	9.787	12.688	14.346	45.979	9.978
Rwanda	-2.406	3.971	7.431	7.434	10.186	19.637	4.817
Tanzania	4.736	5.976	9.084	13.598	19.756	35.827	9.719
Uganda	-0.288	4.0545	6.871	10.3303	11.547	52.442	11.620

Looking at the table, one can easily notice a big difference between the maximum and the third quartile, resulting in big standard deviation values for each country in the community. This

shows that, there is high variation of CPINF in all EAC country. Below is the corresponding histogram.

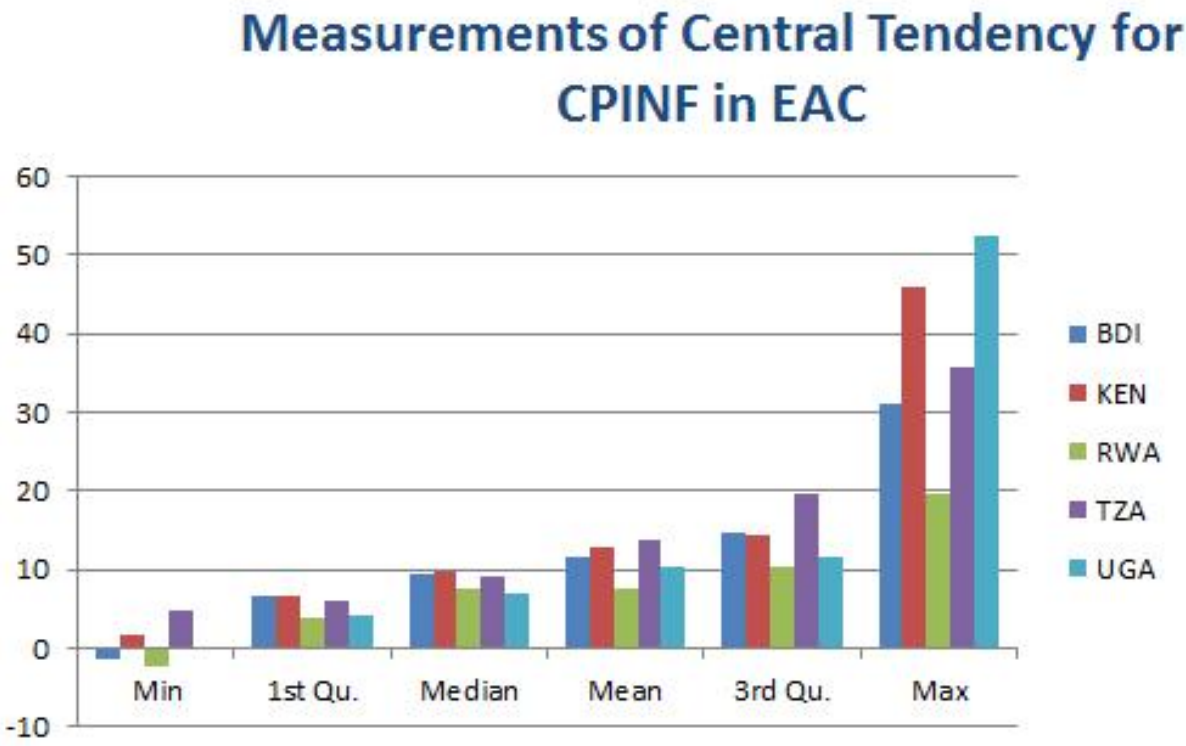

Figure 4.6: Visualization of EAC CPINF Measurements of Central Tendency

4.2.3.2 Trend Test (Mann-Kendall)

It tests if trend in series are statistically significant and the association sign.

```
$Burundi
tau = -0.102, 2-sided pvalue =0.48061

$Kenya
tau = -0.255, 2-sided pvalue =0.070699

$Rwanda
tau = -0.262, 2-sided pvalue =0.064099

$Tanzania
tau = -0.434, 2-sided pvalue =0.0020299

$Uganda
tau = -0.0769, 2-sided pvalue =0.59681
```

Looking at the results of the test, only Tanzania is on a significant negative trend at 5 % level of significance. It is followed by Kenya and Rwanda at 10% level of significance. Despite the mood

of positive trend observed in the area nowadays, in general, EAC CPINF rate has been on a negative trend since 1990.

4.3 Variation of the Impact of Features Considered on CPINF in EAC

In this section we assess the impact of independent variables which will be selected using VIF and stepwise regression, from the list of independent variable considered as described in conceptual framework. Only the features selected for all countries will be assessed for variation.

4.3.1 VIF , Stepwise Feature Selection and Coefficients Analysis

Since real interest rate is the lending interest adjusted for inflation, as mentioned in the conceptual frame work we have used the lagged values of the real interest rate. That is, the interest rate for the previous year was used in the following year for prediction instead of the current real interest rate. In line with this the estimation of the model coefficients has been done in two ways, using the lag values of real interest and then controlling for it. That is removing it to assess the predictive power of the remaining features without it.

We will first use VIF for feature selection and then further apply step wise regression selection on VIF selected features for further selection of features with more predictive power as we mentioned before. Below are the results for each country

Rwanda:

Using the threshold of 5, the features selected for Rwanda using VIF are: NE.CON.PRVT.PC.KD.ZG, TM.VAL.AGRI.ZS.UN, FR.INR.RINR, SL.UEM.TOTL.MA.ZS, SL.UEM.TOTL.ZS, NE.CON.GOVT.KD.ZG, CPINFBDI, CPINFUGA, NE.IMP.GNFS.KD.ZG, NE.IMP.GNFS.ZS, FR.INR.LEND, SL.UEM.TOTL.FE.ZS, NE.CON.GOVT.ZS, BX.KLT.DINV.WD.GD.ZS and CPINFTZA

Stepwise Regression on VIF Selected Features:

With lagged real interest rate at k=1:

Table 4.6: Stepwise regression estimates for consumer price inflation in Rwanda

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	182.675	66.991	2.727	0.021308 *

TM.VAL.AGRI.ZS.UN	0.318	0.195	1.637	0.132774
NE.IMP.GNFS.ZS	-1.320	0.524	-2.517	0.030538 *
FR.INR.RINR	-0.287	0.082	-3.526	0.005487 **
FR.INR.LEND	-2.979	1.235	-2.412	0.036551 *
SL.UEM.TOTL.MA.ZS	-95.632	57.238	-1.671	0.125719
SL.UEM.TOTL.FE.ZS	-37.492	27.410	-1.368	0.201313
NE.CON.GOVT.ZS	-0.50171	0.433	-1.16	0.273049
NE.CON.GOVT.KD.ZG	0.08542	0.061	1.412	0.18832
BX.KLT.DINV.WD.GD.ZS	2.46829	1.200	2.058	0.066653 .
CPINFBDI	0.47701	0.101	4.738	0.000795 ***
CPINFTZA	-0.13678	0.135	-1.014	0.334321

Signif.codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.511 on 10 degrees of freedom
(3 observations deleted due to missingness)
Multiple R-squared: 0.8418, Adjusted R-squared: 0.6678
F-statistic: 4.837 on 11 and 10 DF, p-value: 0.009517

Significant features at 5% level of significance are Imports of goods and services (% of GDP), lagged real interest rate, Lending interest rate (%) and Burundi CPINF, with a marginal effect of -1.31, -0.2872 and 0.47 per unit increase respectively. Foreign direct investment, net inflows (% of GDP) is significant at 10% level of significance with a positive marginal effect.

Without lagged real interest rate:

Table 4.7: Stepwise regression estimates for consumer price inflation in Rwanda, excluding lagged real interest rate

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	30.135	6.20	4.86	0.000126 ***
NE.IMP.GNFS.KD.ZG	0.104	0.030	3.478	0.002685 **
NE.IMP.GNFS.ZS	-1.162	0.275	-4.223	0.000511 ***
BX.KLT.DINV.WD.GD.ZS	2.488	0.850	2.928	0.008977 **
CPINFBDI	0.232	0.086	2.708	0.014396 *

Signif.codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.922 on 18 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.7237, Adjusted R-squared: 0.6622
F-statistic: 11.78 on 4 and 18 DF, p-value: 7.061e-05

Significant features are Imports of goods and services (annual % growth), Imports of goods and services (% of GDP), Foreign Direct Investment, Net Inflows (% of GDP) and Burundi's CPINF with a marginal effect of 0.103, -1.16, 2.487 and 0.2 per unit increase respectively.

Uganda:

Using the threshold of 5, the features selected for Uganda using VIF are: NE.CON.PRVT.PC.KD.ZG, FR.INR.DPST, NE.CON.GOVT.ZS, CPINFKEN, NE.IMP.GNFS.KD.ZG, FR.INR.RINR, NE.CON.GOVT.KD.ZG, TM.VAL.AGRI.ZS.UN, SL.UEM.TOTL.FE.ZS, CPINFRWA, PA.NUS.FCRF, SL.UEM.TOTL.MA.NE.ZS and CPINFBDI.

Stepwise Regression on VIF Selected Features:

With lagged real interest at k=1:

Table 4.8: Stepwise regression estimates for consumer price inflation in Uganda

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-41.164	11.116	-3.703	0.00213 **
NE.IMP.GNFS.KD.ZG	0.180	0.080	2.246	0.04017 *
FR.INR.DPST	2.063	0.201	10.249	3.61e-08 ***
FR.INR.RINR	-0.259	0.135	-1.924	0.07358 .
SL.UEM.TOTL.FE.ZS	2.467	1.526	1.616	0.12682
SL.UEM.TOTL.MA.NE.ZS	3.254	2.635	1.235	0.23576
NE.CON.GOVT.ZS	0.950	0.551	1.723	0.10544
CPINFRWA	0.363	0.287	1.263	0.22571
CPINFKEN	-0.204	0.130	-1.57	0.13734

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 4.037 on 15 degrees of freedom
Multiple R-squared: 0.9012, Adjusted R-squared: 0.8485
F-statistic: 17.1 on 8 and 15 DF, p-value: 3.113e-06

At 5% level of significance, significant features are Imports of goods and services (annual % growth), Deposit interest rate, with a marginal effect of 0.18, 2.06 per unit increase respectively.

Without lagged real interest rate:

Table 4.9: Stepwise regression estimates for consumer price inflation in Uganda, excluding lagged real interest rate

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-27.687	6.445	-4.294	0.000391 ***
NE.CON.PRVT.PC.KD.ZG	-0.413	0.304	-1.36	0.189774
NE.IMP.GNFS.KD.ZG	0.306	0.100	3.086	0.006086 **
FR.INR.DPST	1.929	0.181	10.658	1.87e-09 ***
SL.UEM.TOTL.MA.NE.ZS	5.980	2.327	2.57	0.018744 *

Signif.codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 4.138 on 19 degrees of freedom
Multiple R-squared: 0.8685, Adjusted R-squared: 0.8408
F-statistic: 31.37 on 4 and 19 DF, p-value: 3.947e-08

At 5% level of significance, significant features are Imports of goods and services (annual % growth), Deposit interest rate and Male unemployment (% of male labor force) (national estimate) with a marginal effect of 0.305, 1.92 and 5.9 per unit increase respectively.

Burundi:

Using the threshold of 5, the features selected for Burundi using VIF are: NE.CON.PRVT.PC.KD.ZG, FR.INR.RINR, NE.CON.GOVT.KD.ZG, CPINFUGA, NE.IMP.GNFS.KD.ZG, SL.UEM.TOTL.MA.ZS, BN.KLT.DINV.CD, CPINFRWA, BM.GSR.GNFS.CD, SL.UEM.TOTL.FE.ZS, CPINFKEN, TM.VAL.AGRI.ZS.UN, NE.CON.GOVT.ZS and FP.CPI.TOTL.ZG.

Stepwise Regression on VIF Selected Features:

With lagged real interest at k=1:

Table 4. 10: Stepwise regression estimates for consumer price inflation in Burundi

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.67E+00	4.53E+00	-1.251	0.22684
TM.VAL.AGRI.ZS.UN	3.48E+00	2.31E+00	1.508	0.14889
BN.KLT.DINV.CD	-1.38E-06	4.81E-07	-2.859	0.01043 *
CPINFKEN	-4.40E-01	1.66E-01	-2.642	0.01658 *

CPINFTZA	4.60E-01	1.80E-01	2.555	0.01990 *
CPINFRWA	1.39E+00	3.70E-01	3.757	0.00144 **

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 5.858 on 18 degrees of freedom
Multiple R-squared: 0.6162, Adjusted R-squared: 0.5096
F-statistic: 5.779 on 5 and 18 DF, p-value: 0.002364

At 5% level of significance, significant features are Foreign direct investment, net (BoP, current US$), Kenya CPINF, Tanzania CPINF and Rwanda CPINF with a marginal effect of -1.37e- 06, -0.436, 0.459 and 1.39 per unit increase respectively.

Since lagged interest rate was not chosen in the selection we did not do a without lagged interest version. The low accuracy is Burundi CPINF regression is associated with a lot of missing values found in the dataset used for Burundi.

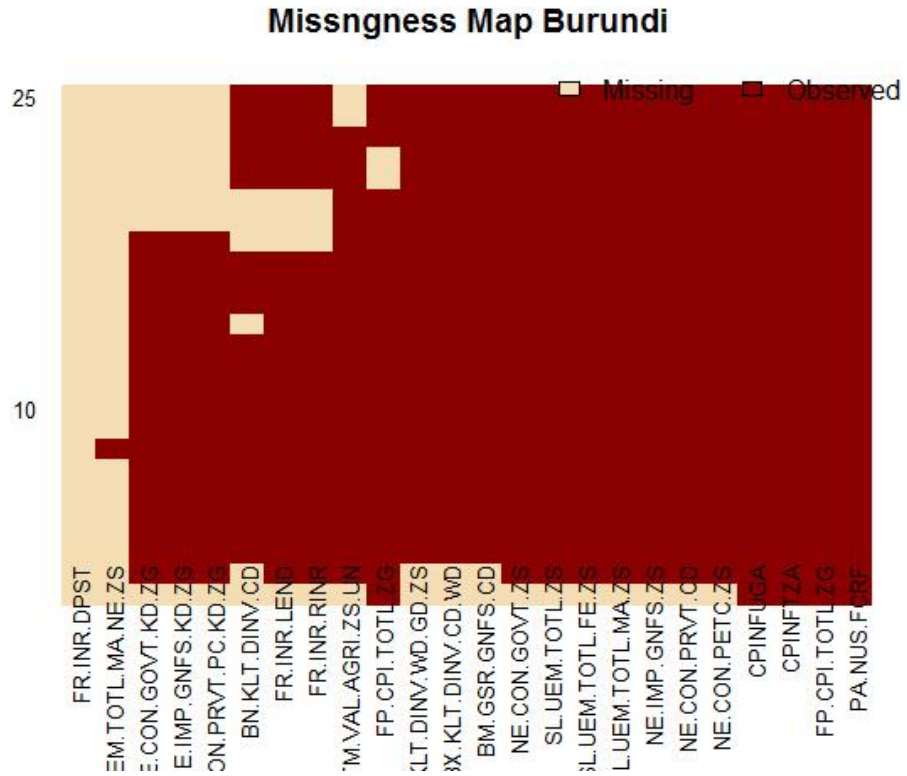

Figure 4.7: Visualization of Missing Data in Burundi

Almost the whole data about deposit interest rate was missing; other variables had poor data as well such as Unemployment, male (% of male labor force) (national estimate).

Tanzania:

Using the threshold of 5, the features selected for Tanzania using VIF are: NE.IMP.GNFS.KD.ZG, SL.UEM.TOTL.MA.ZS, CPINFKEN, PA.NUS.FCRF, SL.UEM.TOTL.MA.NE.ZS, CPINFTZA, FR.INR.DPST, NE.CON.GOVT.KD.ZG, CPINFUGA, FR.INR.RINR and BX.KLT.DINV.WD.GD.ZS.

Stepwise Regression on VIF Selected Features:

With lagged real interest at k=1:

Table 4.11: Stepwise regression estimates for consumer price inflation in Tanzania

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.52192	6.214822	6.037	8.29e-06 ***
NE.IMP.GNFS.KD.ZG	-0.14291	0.067464	-2.118	0.04756 *
PA.NUS.FCRF	-0.0083	0.002639	-3.144	0.00535 **
FR.INR.DPST	0.580305	0.215165	2.697	0.01428 *
SL.UEM.TOTL.MA.NE.ZS	-4.09166	1.297509	-3.153	0.00523 **
BX.KLT.DINV.WD.GD.ZS	-2.05273	0.67124	-3.058	0.00647 **

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.004 on 19 degrees of freedom
Multiple R-squared: 0.8351, Adjusted R-squared: 0.7917
F-statistic: 19.25 on 5 and 19 DF, p-value: 7.605e-07

Significant features are Imports of goods and services (annual % growth),Official exchange rate (LCU per US$, period average), Deposit interest rate, male unemployment (% of male labor force) (national estimate) and Foreign direct investment, net inflows (% of GDP)with a marginal effect of -0.14, -0.0082, 0.58, -4.09 and -2.052 per unit increase respectively.

For Tanzania, the lagged values of the interest rate are not influential in predicting the CPINF, hence there is no need to do for "without real interest" case.

Kenya:

Using the threshold of 5, the features selected for Kenya using VIF are: NE.CON.PRVT.PC.KD.ZG, NE.IMP.GNFS.ZS, NE.CON.GOVT.ZS, CPINFRWA, NE.IMP.GNFS.KD.ZG, FR.INR.DPST, NE.CON.GOVT.KD.ZG, CPINFBDI, BM.GSR.GNFS.CD, FR.INR.RINR, BX.KLT.DINV.WD.GD.ZS and CPINFUGA

Stepwise Regression on VIF Selected Features:

With lagged real interest lagged values at k=1:

Table 4.3 : Stepwise regression estimates for consumer price inflation in Kenya

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	26.37134	16.03187	1.645	0.120772
NE.CON.PRVT.PC.KD.ZG	-1.70111	0.26412	-6.441	1.11e-05 ***
NE.IMP.GNFS.KD.ZG	0.49326	0.09703	5.083	0.000135 ***
FR.INR.RINR	0.1475	0.13057	1.13	0.276369
NE.CON.GOVT.ZS	-1.57529	0.98235	-1.604	0.129648
NE.CON.GOVT.KD.ZG	0.20618	0.13424	1.536	0.145388
BX.KLT.DINV.WD.GD.ZS	1.89213	1.5862	1.193	0.251449
CPINFRWA	0.70695	0.27054	2.613	0.019583 *
CPINFBDI	-0.18004	0.11077	-1.625	0.124919
CPINFUGA	0.2015	0.1042	1.934	0.072227 .

Signif.codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.646 on 15 degrees of freedom
Multiple R-squared: 0.919, Adjusted R-squared: 0.8704
F-statistic: 18.91 on 9 and 15 DF, p-value: 1.228e-06

At 5% level of significance, significant features are Household final consumption expenditure per capita growth (annual %), Imports of goods and services (annual % growth), and Rwanda CPINF rate with a marginal effect of –1.701, 0.49, and 0.706 per unit increase respectively.

Without lagged interest rate:

Table 4.13: Stepwise regression estimates for consumer price inflation in Kenya, excluding lagged real interest rate

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.05558	12.64439	2.693	0.015388 *
NE.CON.PRVT.PC.KD.ZG	-1.8285	0.23771	-7.692	6.2e-07 ***
NE.IMP.GNFS.KD.ZG	0.48414	0.0976	4.961	0.000119 ***
NE.CON.GOVT.ZS	-1.8193	0.7835	-2.322	0.032906 *
NE.CON.GOVT.KD.ZG	0.17132	0.13303	1.288	0.215066
CPINFRWA	0.66457	0.23474	2.831	0.011524 *
CPINFBDI	-0.20469	0.10645	-1.923	0.071408 .

CPINFUGA	0.14762	0.09179	1.608	0.126181

Signif.codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 3.676 on 17 degrees of freedom
Multiple R-squared: 0.9067, Adjusted R-squared: 0.8683
F-statistic: 23.6 on 7 and 17 DF, p-value: 1.422e-07

At 5% level of significance, significant features are Household final consumption expenditure per capita growth (annual %), Imports of goods and services (annual % growth),General government final consumption expenditure (% of GDP), and Rwanda CPINF rate with a marginal effect of –1.8, 0.48, 0.17 and 0.66 per unit increase respectively.

4.3.2 Variation of the Impact of Most Common Selected Features

We start by showing the table of marginal effect of the selected significant features in predicting CPINF for EAC countries, and then a table with most commonly selected features, to have a look at the variation of their impacts in EAC.

Table 4.4 Significant Variables and their Coefficients in Predicting CPINF in EAC

Country	NE.IMP.GNFS.ZS	FR.INR.RINR	FR.INR.LEND	BX.KLT.DINV.WD.GD.ZS	CPINFBDI	FR.INR.DPST	NE.IMP.GNFS.KD.ZG	PA.NUS.FCRF	SL.UEM.TOTL.MA.NE.ZS	NE.CON.PRVT.PC.KD.ZG	CPINFRWA	BN.KLT.DINV.CD	CPINFKEN	CPINFTZA
Rwanda	-1.21962	-0.2472	-2.97915		0.47701									
Uganda						1.928	0.313							
Burundi											1.39	-0.000001376	-0.4396	0.4596
Tanzania				-2.05273		0.5803	-0.1429	-0.0083	-4.081					
Kenya							0.49326			-1.701	0.70695			

Table 4.5: Significant Variable and their Coefficients Excluding Lagged Real Interest Rate in the Modeling Process

	NE.IMP.GNFS.ZS	NE.CON.GOVT.ZS	BX.KLT.DINV.WD.GD.ZS	CPINFBDI	FR.INR.DPST	NE.IMP.GNFS.KD.ZG	PA.NUS.FCRF	SL.UEM.TOTL.MA.NE.ZS	NE.CON.PRVT.PC.KD.ZG	CPINFRWA	BN.KLT.DINV.CD	CPINFKEN	CPINFTZA
Rwanda	-1.18167		2.48782	0.23187		0.103							
Uganda					1.987	0.313							
Burundi											-0.000001375	-0.4396	0.4595
Tanzania			-2.05273		0.5803	-0.1429	-0.0083	-4.081					
Kenya		-1.8163				0.48414			-1.82	0.66			

The without real interest rate version, shows more about common pattern of influential features between country than the other version, but to analyze the variation of the impact of influential features between countries, we have used common influential features appearing in both of the versions, but the coefficients are for the without lagged real interest rate version. Besides, the tables above support further the interdependence of EAC countries CPINF with statically

significant Rwandan CPINF and Burundian CPINF in predicting Kenya and Rwanda respectively. Below we present the table of most common influential features.

Table 4.16: Most Common Influential Features and their Effect Dispersion

	FR.INR.DPST	NE.IMP.GNFS.KD.ZG	Variance of St. Dev
Rwanda		0.103	
Uganda	1.987	0.313	
Burundi			
Tanzania	0.5803	-0.143	
Kenya		0.484	
Mean	1.284	0.189	
Std.Dev	0.995	0.271	0.262

The most common selected features are NE.IMP.GNFS.KD.ZG and FR.INR.DPST with a mean of 1.284 and 0.189 and standard deviation of 0.27 and 0.99 respectively across EAC. The variance of the standard deviation of the marginal effect of the most common influential features on the list of our independent variables is 0.26189. This shows that the effect of change in each of these variables on CPINF is much likely to be the same in EAC.

However, we are more interested in a by country impact difference rather than a by feature impact difference, below is the ANOVA test for a by country impact difference.

H_o: There is no significant difference of overall impact between countries

H_a: There is a significant difference of overall impact between countries

```
Analysis of Variance Table

Response: varb
Df Sum Sq Mean Sq F value Pr(>F)
Labels     4 2.5999 0.64997  0.5866 0.6873
Residuals  5 5.5404 1.10807
```

Given, the results obtained we fail to reject the null hypothesis and conclude that there is no significant difference of overall impact of selected features between EAC countries.

This shows that the overall economic reaction to these features in terms of CPINF is much likely to be the same in EAC.

4.3 Prediction Performance of OLS, SVM, KNN, PLM and ARIMA on EAC CPINF for each Country Member

In this section we assess the prediction performance of the stated models by comparing mean absolute error (MAE) and mean root square error (MRSE) and the coefficient of determination (R_squared) of every model. For OLS, SVM, KNN and PLM we will use the features selected using step wise regression in section 4.3. ARIMA model will use lagged values of CPINF.

Rwanda:

Table4.17: Prediction Accuracy of Models Considered, Rwanda

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	1.269	0.784	3.492	2.580	2.770	2.730	3.333
MRSE	1.693	1.159	4.170	3.102	3.301	3.368	4.132
R_Squared	0.842	0.926	0.248	0.469	0.398	0.348	0.235

The results show that SVM outperforms all other models used for CPINF prediction in Rwanda.

Uganda:

Table 4.186 : Prediction Accuracy of Models Considered, Uganda

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	2.726	1.653	5.936	2.669	2.689	3.539	5.704
MRSE	3.297	1.790	10.425	3.480	3.359	5.381	10.151
R_Squared	0.895	0.969	-0.011	0.882	0.890	0.719	0

The results show that SVM outperforms all other models used for predicting CPINF in Uganda.

Burundi:

Table 4.197: Prediction Accuracy of Models Considered, Burundi

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	4.130	1.606	8.982	6.542	6.542	6.542	6.359
RMSE	5.073	1.621	5.830	8.189	8.189	8.189	8.042
R_Squared	0.616	0.961	0.232	1.412e-16	1.412e-16	1.413e-16	0

The results show that SVM outperforms all other models used for predicting CPINF in Burundi.

Kenya:

Table 4.20: Prediction Accuracy of Models Considered, Kenya

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	2.228	1.57	5.552	2.749	2.635	2.990	5.647
MRSE	2.824	1.678	9.263	3.516	3.359	3.934	8.159
R_Squared	0.919	0.972	0.2575166	0.874	0.885	0.843	0.324

The results show that SVM outperforms all other models used for predicting CPINF in Kenya.

Tanzania:

Table 4.81 : Prediction Accuracy of Models Considered, Tanzania

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	2.486	1.827	3.288	2.713	3.154	3.306	3.313
MRSE	3.490	2.209	4.363	3.918	4.339	4.411	4.302
R_Squared	0.835	0.936	0.743	0.792	0.745	0.737	0.761

The results show that SVM outperforms all other models used for predicting CPINF prediction in Tanzania.

SVM outperformed other models in EAC countries in predicting CPINF. SVM used for each country was tuned using hyper-parameter optimization.

4.4 Prediction Performance of OLS, SVM, KNN, PLM and ARIMA on EAC CPINF Weighted Average

Using the threshold of 5, the features selected for Kenya using VIF are: NE.CON.PRVT.PC.KD.ZG, NE.IMP.GNFS.ZS, NE.CON.GOVT.ZS,BX.KLT.DINV.WD.GD.ZS, NE.IMP.GNFS.KD.ZG, FR.INR.DPST, NE.CON.GOVT.KD.ZG, TM.VAL.AGRI.ZS.UN, FR.INR.RINR and BN.KLT.DINV.CD.

Stepwise Regression on VIF Selected Features:

With real interest lagged values at k=1:

Table 4. 92 : Stepwise regression estimates for weighted consumer price inflation average in EAC

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.40E+00	1.66E+02	-0.027	0.979206
NE.CON.PRVT.PC.KD.ZG	-4.15E-01	9.75E-01	-0.426	0.676047
NE.IMP.GNFS.KD.ZG	6.37E-01	3.29E-01	1.938	0.071733 .
TM.VAL.AGRI.ZS.UN	-2.65E+00	1.32E+00	-2.012	0.062577 .
NE.IMP.GNFS.ZS	8.53E-02	9.00E-02	0.948	0.358259
FR.INR.DPST	4.28E-01	4.76E-01	0.898	0.383407
NE.CON.GOVT.ZS	1.57E-01	2.06E-01	0.76	0.459129
NE.CON.GOVT.KD.ZG	-2.83E-01	5.50E-01	-0.514	0.614846
BN.KLT.DINV.CD	2.38E-09	6.45E-09	0.369	0.717277
BX.KLT.DINV.WD.GD.ZS	-4.75E+00	1.14E+00	-4.169	0.00082 ***

Signif.codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Residual standard error: 21.9 on 15 degrees of freedom
Multiple R-squared: 0.7609, Adjusted R-squared: 0.6175
F-statistic: 5.305 on 9 and 15 DF, p-value: 0.002339

At 5% level of significance, only Foreign Direct Investment, Net Inflows (% of GDP) is significant with a marginal effect of –4.75 per unit increase. The expected sign on FDI was positive, however the negative sign associated with it makes sense in the way that we will explain after the without lagged real interest rate version.

Without lagged real interest rate:

Table 4. 103: Stepwise regression estimates for weighted consumer price inflation average in EAC, excluding lagged real interest rate

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	173.9833	24.9778	6.966	7.03e-07 ***
NE.IMP.GNFS.KD.ZG	0.6424	0.2294	2.8	0.01072 *
TM.VAL.AGRI.ZS.UN	-2.9491	0.7898	-3.734	0.00122 **
BX.KLT.DINV.WD.GD.ZS	-4.5667	0.6815	-6.701	1.25e-06 ***

Signif.codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 20.3 on 21 degrees of freedom
Multiple R-squared: 0.7124, Adjusted R-squared: 0.6713
F-statistic: 17.34 on 3 and 21 DF, p-value: 6.745e-06

At 5% level of significance, significant features are Imports of goods and services (annual % growth) and Agricultural raw materials imports (% of merchandise imports)with a marginal effect of 0.6424, -2.9491and -4.5667 per unit increase respectively. The negative sign associated with FDI is due to reversed causality between FDI and CPINF which we will not tackle in our study. In details, FDI in EAC tend to increase when CPINF is low, hence regressing CPINF on FDI will give FDI's negative marginal impact.

This can be shown by linearly regressing FDI on inflation and the rest of VIF chosen features after stepwise regression on the above case.

Table 4.114; OLS regression estimates for FDI on inflation and the rest of VIF chosen features

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.03E+01	1.33E+01	-0.774	0.44865
NE.IMP.GNFS.KD.ZG	6.71E-02	4.05E-02	1.658	0.11367
NE.IMP.GNFS.ZS	1.84E-02	1.11E-02	1.664	0.11244
NE.CON.GOVT.ZS	2.70E-02	2.08E-02	1.3	0.209

BN.KLT.DINV.CD	2.45E-09	6.33E-10	3.867	0.00104 **
y	-1.10E-01	2.21E-02	-4.975	8.4e-05 ***

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.222 on 19 degrees of freedom
Multiple R-squared: 0.8157, Adjusted R-squared: 0.7672
F-statistic: 16.82 on 5 and 19 DF, p-value: 2.126e-06

As we can see, there is a negative relationship between y (CPINF) and FDI by regressing FDI on CPINF.

Table 4.125: Prediction Accuracy of Models Considered, EAC CPINF Weighted Average

Model	OLS	SVM	KNN	LASSO	Elastic Net	Ridge	ARIMA
MAE	14.892	5.743	22.551	16.419	17.641	18.230	16.083
MRSE	18.605	6.192	29.867	20.528	21.990	22.755	20.673
R_Squared	0.712	0.968	0.308	0.650	0.598	0.570	0.659

The results in table 4.20 show that SVM outperforms all other considered model in predicting weighted average of CPINF EAC countries. The series were weighted by the ratio of mean and standard deviation.

SVMs used for all cases were tuned using hyper-parameter optimization.

CHAPTER V

5. Conclusion, Recommendation and Future research

5.1 Conclusion

The results of the study show that there is interesting relationship of CPINF between EAC countries, and on the other hand the levels of inflation between members of this community are the same statistically. Using Pearson correlation coefficients, at 5% level of significance, all observed significant correlations are positively related.

Tanzania CPINF is significantly correlated with each country's CPINF in the community apart from Burundian's; the country has its highest correlation with Rwanda CPINF. Tanzanian CPINF is the most correlated with other countries' in the community. This might be associated with different factors such as the geo-location of the country and its trade among the country members of the community.

Kenya and Rwanda second Tanzania in having a significant correlation with many countries' CPINF in EAC. Kenya and Rwanda CPINF are more correlated to each other than any other country any of both countries are significantly related with.

Uganda and Burundi CPINF are the least correlated with other countries', with Uganda having a significant correlation with Tanzania only, and Burundi having no significant correlation with any country in the community. Pearson coefficients for mathematical results strongly support the observed reality for the case of Burundi. The country has been suffering from political instability for a long time which might have contributed to the unique behavior of its consumer price inflation with respect to the rest of the community. Moreover, it's the only country which uses French as the primary official language after the shift of Rwanda government to English. This might have an impact as well in terms of who they are likely to trade with outside of the community.

Nevertheless, bivariate normality test shows that Burundi is jointly normally distributed with Uganda, and Uganda turns out to be the one that is jointly distributed with most of countries in

the community. These findings about CPINF relationship between EAC countries is accented by granger causality test which shows that each country's CPINF in the community granger-causes at least one country's CPINF in the same community. Moreover OLS regression with VIF and Stepwise selection on our independent variables, show that CPINF of country member states of EAC are statistically significant in explaining variations of CPINF among them.

Burundi and Rwanda CPINF are statistically significant in predicting the variation in Rwanda CPINF and Kenya CPINF respectively, with a positive marginal effect. On the other hand CPINF of Kenya, Rwanda, and Tanzania are statistically significant in explaining Burundi's CPINF variation. Besides, the variance homogeneity and means equality test also show no significant difference between the levels of CPINF in the 5 EAC countries considered.

The positive interdependence of CPINF and the same levels of CPINF in EAC country member states are an advantage to these countries; it signals a possibility of common policies that can jointly be used to control CPINF in all the countries once there is an implementation of one currency in the area. Since their CPINF is positively related to each other and within the same range, the impact of a common policy on this variable can easily propagate and level off within the area.

The most common influential variables on EAC CPINF on the list of independent variable considered are Imports of Goods and Services (annual % growth) and Deposit Interest Rate with a marginal effect mean of 1.284 and 0.189 and a standard deviation of 0.27 and 0.99 respectively across EAC. The ANOVA tests show that there is no difference of the overall impact of these most common influential features between EAC countries. This shows that the overall economic reaction to these features in terms of CPINF is much likely to be the same across EAC countries.

For prediction performance, the assessment shows that SVM outperforms all other models considered for each country CPINF in a within sample prediction with an accuracy of 92%, 96%, 96%, 97% and 93% on Rwanda, Uganda, Burundi, Kenya and Tanzania respectively in a within sample prediction accuracy. It is followed by OLS with an accuracy of 84%, 89%, 10%, 91% and 83% on Rwanda, Uganda, Burundi, Kenya and Tanzania respectively. For a weighted average CPINF for all countries, SVM performs better still with an accuracy of 96% followed by OLS with 71%. The model which turned out to have the lowest performance is KNN regression.

5.2 Recommendation and Future Research

This research shows that there is a possibility of handling together CPINF in EAC countries since they are dependent of each other and are likely to behave the same in all EAC countries.

Given the above, it would be better for EAC countries to look at how they can manage their CPINF together using common policies by targeting mostly the most common influential variables in the area that can directly influence CPINF. In this case, it's depositing interest rate and imports of goods and services. For example for imports of goods and services, it would be better to promote local production within the area and free taxes between countries trading of certified local products.

On the research angle, the findings for this research is a motivation of more other researches that can assess interdependence of other macro-economic variables between EAC countries, since this also explains how each country economic decision might influence others in one way or another. On the prediction side, the study also shows interesting results, that each country CPINF in EAC granger-causes at least one country's CPINF in the community. From this perspective it would be good for a future research which would assess the prediction performance of vector autoregressive model on EAC CPINF using the results of the conducted Granger-Causality test. Moreover, the accuracy has been assessed on a within sample level, an out of sample analysis can also be a good option to further assess the performance of the models considered.

References:

Abrahamsen, T.J. and Hansen, L.K., 2011. A cure for variance inflation in high dimensional kernel principal component analysis.*The Journal of Machine Learning Research*, 12, pp.2027-2044.

Ajayi, S.I. (1974). The Demand for Money in the Nigerian Economy: Comments and Extensions. *Nigerian Journal of Economic and Social Studies*, Vol. 16(1): 165-173.

Alnaa, S.E. and Ahiakpor, F., 2011.ARIMA (autoregressive integrated moving average) approach to predicting inflation in Ghana.*Journal of economics and international finance*, 3(5), pp.328-336.

Baghestani, H. and AbuAl-Foul, B., 2010.Factors influencing Federal Reserve forecasts of inflation.*Journal of Economic Studies*, 37(2), pp.196-207.

Bokil, M. and Schimmelpfennig, A., 2005.Three Attempts at Inflation Forecasting in Pakistan.

Boser, B.E., Guyon, I.M. and Vapnik, V.N., 1992, July.A training algorithm for optimal margin classifiers.*In Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).ACM.

Chen, C.H., Pau, L.F. and Wang, P.S.P. eds., 2010.*Handbook of pattern recognition and computer vision* (Vol. 27).Imperial College Press.

Colannino, J. and Toussaint, G., 2005. Faster algorithms for computing distances between one-dimensional point sets.*Proceedings of the XI Encuentros de GeometriaComputacional*, pp.189-198.

Downes, A., Holder, C. and Leon, H., 1991. A Co integration Approach of Modeling Inflation in a Small Open Economy. *Journal of Economic Development*, 16(1).

Duisenberg, W.F., 1999. Economic and Monetary Union in Europe the Challenges Ahead.New Challenges for Monetary Policy, pp.26-28.

Etuk, E.H., 2014. Modelling of daily Nigerian Naira-British pound exchange rates using Sarima methods.*British Journal of Applied Science & Technology*, 4(1), p.222.

Fama, E.F., 1975. Short-term interest rates as predictors of inflation.*The American Economic Review*, 65(3), pp.269-282.

Friedman, I.S., 1973. Inflation a World-wide Disaster.

Fyfe, C., 2010. Intelligent Data Engineering and Automated Learning--IDEAL 2010: 11th International Conference, Paisley, UK, September 1-3, 2010, Proceedings (Vol. 6283). *Springer Science & Business Media.*

Gneiting, T. and Thorarinsdottir, T.L., 2010.Predicting inflation: Professional experts versus no-change forecasts.arXiv preprint arXiv:1010.2318.

Gunn, S.R., 1998. Support vector machines for classification and regression. *ISIS technical report*, 14.

Hafer, R.W. and Hein, S.E., 1990. Forecasting inflation using interest-rate and time-series models: Some international evidence. *Journal of Business*, pp.1-17.

Hassani, H., Heravi, S. and Zhigljavsky, A., 2013. Forecasting UK industrial production with multivariate singular spectrum analysis.*Journal of Forecasting*, 32(5), pp.395-408.

Hinshaw, R.W. ed., 1972. Inflation as a global problem. Baltimore: *Johns Hopkins University Press*.

Humphrey, T.M., 1974. The quantity theory of money: its historical evolution and role in policy debates. *FRB Richmond Economic Review*, 60, pp.2-19.

Jochum, C. and Kirchgassner, G., 1999. Hat die ZinsstrukturAussagekraft fur die zukunftige Inflation in Deutschland. Kredit und Kapital, 4(99), pp.493-519.

Johnson, H.G., 1972. Inflation and the monetarist controversy.*North-Holland Pub. Co.*.

Levi, M.D. and Makin, J.H., 1980. Inflation uncertainty and the Phillips curve: some empirical evidence. *The American Economic Review*, pp.1022-1027.

Marwala, T., 2013.Economic modeling using artificial intelligence methods.*Heidelberg: Springer*.

Mishkin, F.S., 1990. What does the term structure tell us about future inflation?.*Journal of monetary economics*, 25(1), pp.77-95.

Önder, A.Ö., 2004. Forecasting inflation in emerging markets by using the Phillips curve and alternative time series models.Emerging Markets Finance and Trade, 40(2), pp.71-82.

Pourgerami, A. and Maskus, K.E., 1987. The effects of inflation on the predictability of price changes in Latin America: some estimates and policy implications. *World Development*, 15(2), pp.287-290.

Scheide, J. and Trabandt, M., 2000.Predicting inflation in Euroland: the Pstar approach (No. 1019).*Kiel Working Papers.*

Shi, Y. and van Albada, G.D., 2007. Computational Science-ICCS 2007: 7th International Conference, Beijing China, May 27-30, 2007, Proceedings (Vol. 3). *Springer Science & Business Media*.

Siliverstovs, B. and Bilan, O., 2005. Modeling inflation dynamics in transition economies: The case of Ukraine.Eastern European Economics, 43(6), pp.66-81.

Singh, S., Singh, M., Apte, C. and Perner, P. eds., 2005. Pattern Recognition and Data Mining: Third International Conference on Advances in Pattern Recognition, ICAR 2005, Bath, UK, August 22-25, 2005 (Vol. 3686). Springer.

Stock, J.H. and Watson, M.W., 2001. Forecasting output and inflation: the role of asset prices (No. w8180). *National Bureau of Economic Research*.

Stock, J.H. and Watson, M.W., 2010.Modeling inflation after the crisis (No. w16488).*National Bureau of Economic Research*.

Teriba, A.0. (1974). The Demand for Money in the Nigerian Economy: Some Methodological Issues and Further Evidence. *Nigerian Journal ofEconomic and Social Studies*, 16(1) 153-164.

Thomakos, D.D. and Bhattacharya, P.S., 2005. Forecasting inflation, industrial output and exchange rates: a template study for India. *Indian economic review*, pp.145-165.

Tzavalis, E. and Wickens, M.R., 1996. Forecasting inflation from the term structure.*Journal of Empirical Finance*, 3(1), pp.103-122.Vancouver

Vu, K.M., 2007. The ARIMA and VARIMA Time Series: Their Modelings, Analyses and Applications. *AuLac Technologies Inc*..

Welfe, A., 2000. Modeling inflation in Poland.Economic Modelling, 17(3), pp.375-385.

Wright, J.H., 2008. Bayesian model averaging and exchange rate forecasts.*Journal of Econometrics*, 146(2), pp.329-341.

Yeung, D.S., Liu, Z.Q., Wang, X.Z. and Yan, H. eds., 2006. Advances in Machine Learning and Cybernetics: 4th International Conference, ICMLC 2005, Guangzhou, China, August 18-21, 2005, Revised Selected Papers (Vol. 3930). *Springer Science & Business Med*

Appendix

Rwanda Consumer Price Inflation since 1990 as retrieved from Word Bank:

Time	Consumer Price Inflation Rate(%)
1990 [YR1990]	4.185763726
1991 [YR1991]	19.63716581
1992 [YR1992]	9.560411876
1993 [YR1993]	12.35438876
1994 [YR1994]	NA
1995 [YR1995]	NA
1996 [YR1996]	7.411371735
1997 [YR1997]	12.01542252
1998 [YR1998]	6.210067095
1999 [YR1999]	-2.405932097
2000 [YR2000]	3.899529803
2001 [YR2001]	3.342855067
2002 [YR2002]	1.992585425
2003 [YR2003]	7.44970014
2004 [YR2004]	12.25071029
2005 [YR2005]	9.014089181
2006 [YR2006]	8.882826548
2007 [YR2007]	9.080722059
2008 [YR2008]	15.44493118
2009 [YR2009]	10.3941857
2010 [YR2010]	2.309146191
2011 [YR2011]	5.670682731
2012 [YR2012]	6.27090301
2013 [YR2013]	4.234780151
2014 [YR2014]	1.784100412
2015 [YR2015]	2.518087704

Uganda Consumer Price Inflation since 1990 as retrieved from Word Bank:

Time	Consumer Price Inflation Rate(%)
1990 [YR1990]	33.11865
1991 [YR1991]	28.06814
1992 [YR1992]	52.44227
1993 [YR1993]	1.163983
1994 [YR1994]	10.03676
1995 [YR1995]	6.55014

1996 [YR1996]	7.191647
1997 [YR1997]	8.169021
1998 [YR1998]	0.068804
1999 [YR1999]	5.777369
2000 [YR2000]	3.392022
2001 [YR2001]	1.865125
2002 [YR2002]	-0.28751
2003 [YR2003]	8.680477
2004 [YR2004]	3.721287
2005 [YR2005]	8.448726
2006 [YR2006]	7.310676
2007 [YR2007]	6.138511
2008 [YR2008]	12.05086
2009 [YR2009]	13.01726
2010 [YR2010]	3.976553
2011 [YR2011]	18.6929
2012 [YR2012]	14.01606
2013 [YR2013]	5.464402
2014 [YR2014]	4.288209
2015 [YR2015]	5.225427

Burundi Consumer Price Inflation since 1990 as retrieved from Word Bank:

Time	Consumer Price Inflation Rate(%)
1990 [YR1990]	7.002169567
1991 [YR1991]	8.996938735
1992 [YR1992]	1.823333333
1993 [YR1993]	9.679346581
1994 [YR1994]	14.85281498
1995 [YR1995]	19.26325364
1996 [YR1996]	26.43678161
1997 [YR1997]	31.11158983
1998 [YR1998]	12.50041077
1999 [YR1999]	3.385424273
2000 [YR2000]	24.31768096
2001 [YR2001]	9.242971751
2002 [YR2002]	-1.370974453
2003 [YR2003]	10.76166973
2004 [YR2004]	7.851688218
2005 [YR2005]	13.52367836
2006 [YR2006]	2.809014978

:

2007 [YR2007]	8.342032406
2008 [YR2008]	24.10735481
2009 [YR2009]	10.98146877
2010 [YR2010]	6.401249024
2011 [YR2011]	9.735019199
2012 [YR2012]	18.01281606
2013 [YR2013]	7.950674328
2014 [YR2014]	4.379840041
2015 [YR2015]	5.554204168

Tanzania Consumer Price Inflation since 1990 as retrieved from Word Bank:

Time	Consumer Price Inflation Rate (%)
1990 [YR1990]	35.82677
1991 [YR1991]	28.69565
1992 [YR1992]	21.84685
1993 [YR1993]	25.27726
1994 [YR1994]	34.08336
1995 [YR1995]	27.42779
1996 [YR1996]	20.97726
1997 [YR1997]	16.09065
1998 [YR1998]	12.79975
1999 [YR1999]	7.890433
2000 [YR2000]	5.923961
2001 [YR2001]	5.147468
2002 [YR2002]	5.317834
2003 [YR2003]	5.303566
2004 [YR2004]	4.735801
2005 [YR2005]	5.03457
2006 [YR2006]	7.250973
2007 [YR2007]	7.025514
2008 [YR2008]	10.27839
2009 [YR2009]	12.14223
2010 [YR2010]	6.200156
2011 [YR2011]	12.69097
2012 [YR2012]	16.00109
2013 [YR2013]	7.870724
2014 [YR2014]	6.131614
2015 [YR2015]	5.587837

Kenya Consumer Price Inflation since 1990 as retrieved from Word Bank:

Time	Consumer Price Inflation Rate (%)
1990 [YR1990]	17.78181
1991 [YR1991]	20.0845
1992 [YR1992]	27.33236
1993 [YR1993]	45.97888
1994 [YR1994]	28.81439
1995 [YR1995]	1.554328
1996 [YR1996]	8.864087
1997 [YR1997]	11.36185
1998 [YR1998]	6.722437
1999 [YR1999]	5.742001
2000 [YR2000]	9.980025
2001 [YR2001]	5.738598
2002 [YR2002]	1.961308
2003 [YR2003]	9.815691
2004 [YR2004]	11.62404
2005 [YR2005]	10.31278
2006 [YR2006]	14.45373
2007 [YR2007]	9.75888
2008 [YR2008]	26.23982
2009 [YR2009]	9.234126
2010 [YR2010]	3.961389
2011 [YR2011]	14.02155
2012 [YR2012]	9.378396
2013 [YR2013]	5.718274
2014 [YR2014]	6.877498
2015 [YR2015]	6.582411