IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 17, Issue 6 Ser. I (Nov. – Dec. 2021), PP 18-27
www.iosrjournals.org

Modeling of Petroleum Prices in Kenya Using Autoregressive Integrated Moving Average and Vector Autoregressive Models

K.S Fondo[1], A. A Onago[1], L.A Kiti[2], C.W Otulo*[1]
[1](Department of Mathematics and Physics, Technical University of Mombasa, Kenya)
2(Department of Mathematics and Computer Science, Pwani University, Kilifi, Kenya)

Abstract
*The demand for crude oil and petroleum products in Kenya has been increasing very fast over the past twenty years. This is mainly because this particular commodity is used in many sectors of the country's economy. Ever changing prices affect the exchange rates which also affects industrial production of goods in Kenya. The oil production sector has a crucial impact on the other industries. Any change in the price of petroleum products has a great impact on the prices of other goods produced and even the growth of the economy. This is mainly due to the transport cost involved in transporting the goods. The major aim of this research is to model petroleum products prices in Kenya using Autoregressive Integrated Moving Average(ARIMA) and Vector Autoregression (VAR) models; the models are then compared to determine which of them predicts better the prices in Kenya. Modeling of the prices will greatly guide the government and investors in the energy sector so that they can accurately forecast the future prices. The main sources of data in this research were secondary petroleum pump prices data from the Energy and Petroleum Regulatory Authority (EPRA) of Kenya, exchange rates, inflation rates and the crude oil prices in the world market. Petroleum products prices data from January, 2011 to December, 2018 was used for the modeling process. Comparing several ARIMA candidate models using model criterion, ARIMA (1,1,0) emerged the best model was used to check how international oil prices, the exchange rate of Kenyan shilling against the dollar and inflation rate of the Kenyan shilling affect the petroleum prices in Kenya. Comparison of forecasting ability for the ARIMA and VAR models was done using the mean absolute percentage error (MAPE) mean absolute error (MAE) and the root mean squared error (RMSE). The results showed that VAR was better for forecasting the petroleum prices in Kenya as compared to ARIMA (1,1,0).***Keywords:** China insurance industry, Foreign fund, Challenge
Key words: Modeling, Forecasting, Stationarity, ARIMA and VAR models

---

Date of Submission: 12-11-2021 Date of Acceptance: 28-11-2021

I. Introduction

Petroleum is an important product which plays a major role in our daily life. Its high demand is related to many activities related to the country's economy [1]. In a developing country, fuel users need to know a cost effective fuel usage plan to be able to meet their needs.

Kenya heavily depends on petroleum products which are used in most of its sectors of the economy on daily basis [2]. Modeling and forecasting of oil prices are a vital concern in most developing economies [2]. The prices of oil products from the OPEC (Organisation of Petroleum Exporting Countries) countries are unstable and this leads to a great challenge in our country since the high and unstable oil prices is a major obstacle in the growing of Kenya's economy [3]. Oil price volatility has a major negative impact on Kenya's economic growth. The Kenya's oil industry continues to be a very important part of the country's growing economy since oil is used in transportation of goods, generation of water plus electricity and also in the industrial production as an input [2]. This study aimed at finding the best ARIMA and VAR models that can be used to predict the prices of petroleum products in the near future in Kenya based on past prices. ARIMA model is used for analyzing a single variable using its previous values and error terms [2].

In a VAR model, more than one variables are considered. Each variable in the model has an equation which is based on its own lags and those of other variables which are also contained in the model. There is high volatility of the petroleum products prices in Kenya, hence the need to model them so as to guide both the businessmen and customers in the energy sector so that they can predict the future prices accurately and make useful decisions. Investors need to know the future price trends so that their investments can be based on them.[4]

II. Literature Review

In Kenya little has been done to predict the future petroleum products price trends. Though there is substantial amount of some work done on petroleum products prices in Kenya, little attention has been given to predicting prices by comparison of different types of models. This study intended to forecast the prices in the near future by making a clear comparison of the forecast performance of ARIMA and VAR models.

[5] tried to analyze the oil price volatility in Nigeria. They mainly employed Generalised Autoregressive Conditional Heterodasticity (GARCH) model and checked its variants using every day data, monthly based and quarterly based data. Their results showed that the macroeconomic variables considered were all volatile. [2] researched on nonlinear time series modeling of diesel prices in Kenya. Their results finally opted for the use of a non- linear model which implied these prices are not stable at all and instead they keep on fluctuating unexpectedly.

In Nigeria, [6] studied inflation rates using annual time series data from 1950 to 2014, and made use of ARIMA methodology and discovered that ARIMA (3,1,0) model was the best for this study.

[7]tried to model and predict the Consumer Price Index in Rwanda using monthly data from February,1995 - December, 2015 and employed the ARIMA approach. Getting a model for forecasting Consumer Price Index was their main aim as this is significant to the energy sector and the Rwanda's Central Bank in formulating of economic issues. They found that ARIMA (4,1,6) model was the best and that it fits the data well. They discovered that the country's CPI was going to rise up with time.

[8] modeled inflation rates in Nigeria using the two models ARIMA and GARCH. According to their findings ARIMA (1,0,2) was better than other models. Their results showed a rising trend to approximately 17% p.a by December 2021 and might be more after a period of six years.

[9] used ARIMA modeling approach in trying to predict the future production of wheat comparing to the growing population in India. The ARIMA (1,1,0) proved to be the better model. It was then used for estimation up to ten years of production of wheat in India using previously collected data. It showed an increasing production of wheat in the next ten years.

[10] used past information to come up with ARIMA models and predict future demand in manufacturing of food in Morocco. They selected ARIMA (1,0,1) and proved it was useful in both modeling plus estimating the demand of food manufacturing in future.

[11], made an attempt to come up with the statistical models for forecasting petroleum pump prices in Kenya. He applied double exponential smoothing tools and ARIMA models. An increasing trend was noted in the data and therefore smoothing of the data was necessary. He discovered that double exponential smoothing technique was more appropriate than the simple and triple exponential smoothing. By the use of the ARIMA (1,1,1) model he estimated the parameters and came up with models for forecasting the petroleum prices.

[12] were able to carry out modeling of the Nigerian crude oil prices where they managed to apply ARIMA Intervention. Their results showed that pre- intervention model was more superior than the ARIMA (1,1,1) and post- intervention models.

[13] found that ARIMA- GARCH hybrid model was suitable for predicting the returns of oil prices over a short term. He managed to construct a hybrid model for forecasting oil prices. [14] applied ARIMA in crude oil price forecasting. They managed to

examine the time series and nonlinear feature of the oil prices. According to their results, ARIMA (0,1,4) was the most appropriate model for prediction of the oil prices. [15] predicted oil prices based on univariate time series models. They compared the performance accuracy of exponential smoothing, Holt Winter's and ARIMA models. They were applied to regular oil prices for WTI crude oil.

It was shown that the Holt Winter model performed better than the exponential smoothing for a 95% confidence interval.

ARIMA (2,1,2) model yielded the best results. Recently [16] proposed an ARIMA-SVR hybrid model which they used to predict or forecast the short-term Electricity Consumption Transmitted by the internet of things. A brief report by [17] focused on analyzing the robustness of ARIMA and Artificial Neural Network (ANN) as a predictive model in the forecasting of the crude oil prices. ARIMA showed less errors hence better model. [18] in their recent study, investigated a high- dimensional VAR model in mortality modeling and forecasting. They proposed an extension of the sparse VAR (SVAR) model fitted on the log-mortality improvements which they named spatially penalized smoothed VAR (SSVAR).

[19], attempted to apply the ARIMA model in forecasting GDP and CPI in Jordan. Based on his study, ARIMA (3,1,1) emerged the best model for the GDP while ARIMA (1,1,0) was the best for forecasting the CPI.

Recently [20] managed to forecast COVID – 19 confirmed cases in the major Indian cities plus their connectedness with mobility and also the weather-related parameters. He was able to use both univariate and multivariate time series forecasting techniques, that is
SARIMA, ARMAX and VAR models to predict the COVID – 19 cases in New Delhi, Mumbai and Bengaluru.

III. Methodology

This study was carried out on petroleum products prices provided by the Energy and Petroleum Regulatory Authority (EPRA) of Kenya. It was based on the prices for super, and diesel from January, 2011 to December, 2018. The study design applied in this study was time series design. This study was based on prices of petroleum products in Kenya which are available from the EPRA website.
It considered the prices of these products in
Mombasa, Nairobi and Kisumu which are the three major cities in Kenya. This was used to represent the country. The subjects in this study were the prices of two petroleum products in Kenya. After
collecting the prices of the two petroleum
products in the three cities, the data was then
analyzed using the R software. It was modeled using ARIMA and VAR models.

ARIMA Model

ARIMA (p,d,q) model is one of the models commonly used in forecasting time series. The three integers p, d and q usually refer to the Auto regression (AR), Integrated (I) and Moving Average (MA) parts of a given data set respectively. When an ARMA model is differenced, it becomes an ARIMA model. The ARIMA model considers previous data of a given variable and decomposes it into the autoregressive process, an integrated process which checks stationarity and moving average of the forecast errors. Autoregression process expresses a dependent variable in terms of its own previous values. Its applied to a data whose mean, variance and the autocorrelation function are constant over time i.e its stationary. An autoregressive *pth*order process can be expressed as given below

$$yt = \alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

y_t represents dependent variable to be predicted at that period t, $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ are the previous values of the variable at previous lags $t-1, t-2, \ldots, t-p$, respectively and α is the constant of the process, $\varphi_1, \varphi_2, \ldots, \varphi_p$ are the parameters that should be estimated and ε_t is the term that contains error and its normal distributed i.e$\varepsilon_t \approx N(0, \sigma^2)$. A q^{th} order process of a moving average is defined as yt$= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$
where the number of lags in the moving average is represented by q and $\theta_1, \theta_2 \ldots \theta_q$ are parameters that should be estimated.
To create an ARMA model, we start with an econometric equation with no independent variables i.e$y_t = \beta_0 + \varepsilon_t$ then autoregressive and moving average processes are added to it.

$$y_t = \alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_{t-q} \ldots\ldots\ldots\ldots (3)$$

$whereby \;\; y_t = \alpha + \varphi_1 y_{t-1} + \varphi_2 \ldots\ldots\ldots\ldots\ldots + \varphi_p y_{t-p} + \varepsilon_t$is the AR($_p$) and
$\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q$is the MA($_q$) process.

The φ_s and $\theta_s$are the respective parameters of the auto regression and moving average processes.

To estimate the parameters φ, β and θ, selected models are run as guided by the log likelihood, standard error and the obtained values of the AIC (Akaike Information Criteria). Each element of the model will be estimated and this will be shown in the results. The parameters that will have the lowest standard error in RMSE, RMSPE and MAE are the ones that will be most suitable.
The estimated model should be checked for adequacy by finding out whether the residuals form a model that have normal distribution and its random. Box-Ljung Q statistic is used to check the adequacy. The test statistic Q is given by,

$$Q_m = n(n+2)\sum_{k=1}^{m} \frac{r^2 k(e)}{n-k} \sim \chi_{m-r}$$

Where r^2k (e) = residual autocorrelation, n = no. of residuals, m = no. of time lags.
If p value associated with Q statistic is small ($p < \alpha$), the model is inadequate. Then analysis has to be done on new models until a satisfactory one is determined.

VAR Model

The Vector Auto Regression (VAR) model is used in analyzing of multivariate time series where more than one variables are considered.

Let $Y_t = (y_{1t}, y_{2t...}, y_{nt})$ denote an (n x 1) vector of time series variables.
A Vector Auto regression model that contains p number of lags will be given by:

$$Y_t = m + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t, \quad t = 1, 2, \ldots. T. \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (5)$$

A_i is an (n x n) co-efficient matrix, ε_t is an (n x 1) white noise vector process with a mean of zero and m is an (n x 1) vector containing the constants.
A simple VAR with n = 2 and p = 1 is given by

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} \mathcal{E}_{1t} \\ \mathcal{E}_{2t} \end{bmatrix} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

Each of the variables is given as a linear combination of its lagged values plus those of the others in the group. The methods commonly used are Ordinary Least Square Estimator (OLSE) and the Maximum Likelihood Estimator (MLE).
The log likelihood function is expressed as,

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{T}{2} log - \frac{1}{2\delta^2} \sum_{t=1}^{T} \varepsilon \ t^2 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (7)$$

T represents value for time t = 1, 2 …….. T of the given data, ε_t is the term for error and δ^2 the constant variance.

Comparison of models
The performance of the models is also compared so as to determine which one best forecasts the petroleum prices in Kenya. The methods applied here are finding the percentage errors and mean absolute percentage errors (MAPE) for the two models.

$$MAPE = \left[\frac{1}{n} \sum_{i=1}^{n} abs \ \frac{(\hat{y}_t - y_t)}{y_t} \right] \times 100 \ \% \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

whereby $\hat{y}_t$ represents the forecasted value while y_t the exact value, and n are points that have been fitted into the model. The analysis was done by means of R statistical software. The data was found to be non-stationary and hence differencing was necessary. It became stationary after first order difference. After differencing, several ARIMA model combinations were suggested. Log likelihood and AIC was applied to choose the best ARIMA candidate model among the models.

The VAR model was used to check the effect of three independent variables on the petroleum prices. These are international oil prices, the exchange rate and the inflation rate. An analysis of these products was done and a suitable VAR model obtained after parameter estimation. Prediction was made for the next twelve months.

Statistical analysis
Augumented Dickey Fuller test is applied to check for stationarity in the ARIMA modeling. This is through the existence or the absence of unit-root. When the test statistic is greater than ADF critical value, then unit root exists and this implies non- stationarity of the given data set.

IV. Result

Unit Root Test for super and diesel.
To test for unity, ADF test is applied to all the three products.
Table 1: shows the outcomes of the ADF tests for stationarity for super and diesel.

Table no.1: ADF tests for stationarity.
H_0: super prices data has unit root vs H_1: super prices data doesn't have unit root.

Super	Before differencing		First order differencing
	t- Statistics	Probability	t-Statistics Probability
The ADF test	-1.2587	0.8825	-4.3004 0.01

Critical values for the test 1% 5% 10%	-4.04 -3.45 -3.15	

H_0: diesel prices data has unit root vs H_1: diesel prices data doesn't have unit root.

Diesel	Before differencing		First order differencing	
	t-statistics	Probability	t-statistics	Probability
The ADF test	-0.67969	0.9693	-4.2419	0.01
Critical values for the test 1% 5% 10%	-4.04 -3.45 -3.15			

From table no. 1 above, all the two tests produced ADF statistics which are more than their respective critical values. Since the p- values exceed 0.05, we fail to reject the null hypotheses and conclude that the sets of data aren't stationary.

After carrying out first order differencing, the p-values for the ADF statistics were less at $\alpha = 0.01$, 0.05 and 0.1 hence the null hypotheses are now rejected and conclude that the petroleum prices attain stationarity after first order differencing. This implies d=1 in the ARIMA (p,d,q) model. ACF graphs suggested several Autoregressive Integrated Moving Average models which are given as ARIMA(1,1,0) ARIMA(1,1,1),ARIMA (1,1,2) and the ARIMA(1,1,3) models. After using log likelihood and AIC criterion, ARIMA (1,1,0) emerged better for predicting petroleum prices. It had the smallest values of AIC as compared to the other models. Table no.2shows results of the four ARIMA combinations. ARIMA (1,1,0) model emerged the most appropriate in predicting the prices.

Table no. 2 : Results of ARIMA combinations

Diesel	**ARIMA(1,1,0)**	**ARIMA(1,1,1)**	**ARIMA(1,1,2)**	**ARIMA(1,1,3)**
Standard error	11.80	11.73	11.52	11.50
Log likelihood	-249.38	-249.11	-248.32	-248.24
AIC	504	506.22	506.64	508.48

Super	**ARIMA(1,1,0)**	**ARIMA(1,1,1)**	**ARIMA(1,1,2)**	**ARIMA(1,1,3)**
Standard error	13.81	13.69	13.69	13.69
Log likelihood	-256.79	-256.38	-256.38	-256.38
AIC	519.58	520.76	522.76	524.76

Table no.3 below shows the results of estimation for super and diesel from ARIMA (1,1,0) model.
It shows the estimation of the parameter coefficients, standard error, log likelihood and the AIC.
Let $\mathbf{X_t}$ be first order difference for super time series, then the model for super oil prices will be given by $\mathbf{X_t = 0.1512 - 0.0424X_{t-1} + \varepsilon_t}$, where $\square_t$ is the error term following mean zero and constant variance.Let $\mathbf{X_t}$ be first order difference for diesel time series, then the model for diesel oil prices will be given by $\mathbf{X_t = 0.1432 + 0.001X_{t-1} + \varepsilon_t}$

Table no.3: Estimation results for the super and diesel

Variable	**Intercept**	**Co-efficient**	**Std error**	**Std error**	**Log likelihood**	**AIC**	**Variance**	**MAE**	**MAPE**
		AR1	AR1	SAR1					
Super	0.1512	-0.0424	0.1033	0.1235	-256.97	519.58	13.81	2.563	0.0802
Diesel	0.1432	0.0010	0.1023	0.1223	-249.38	504.76	11.80	2.326	0.1309

Diagnostic checking

This is done to check if the data is fitted well by the selected model. The residuals are expected to be randomly independent and identically distributed following a normal distribution. Normal Q-Q plots and histograms were used for checking normality. Box- Ljung test and ACF were used to check whether the residuals are correlated or not. The two figures below show the normal Q-Q plots for residuals of the fitted ARIMA (1,1,0) model.

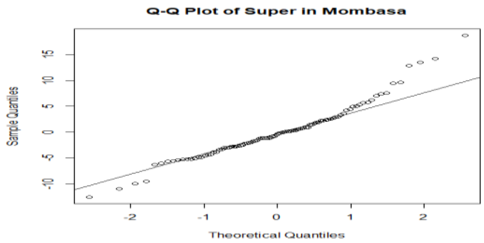

Fig: Normal Q-Q plot for super residuals

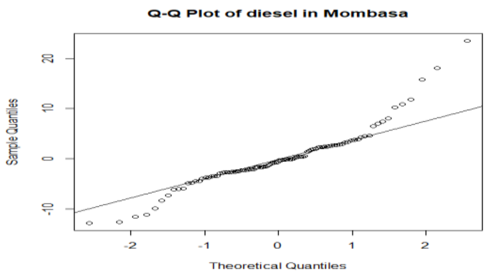

Fig.2:Normal Q-Q plot for diesel residuals

The normal Q-Q plots for all the two products residuals are normal. There are few outliers at the end points. Most of the data points lie on the normal line and those not on the line deviate to the similar extent below and above the normal line. It can be concluded that with respect to mean and variance, the standard errors are roughly constant. This confirms that the residuals are normal distributed.

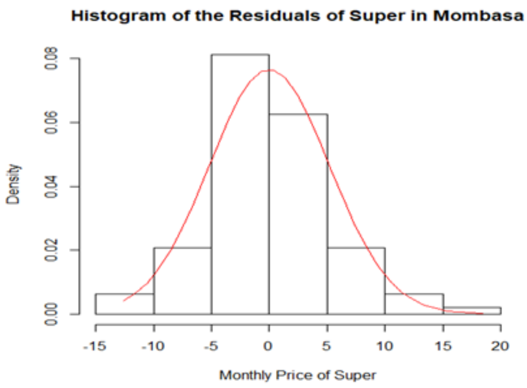

Fig.3: Histogram of super residuals

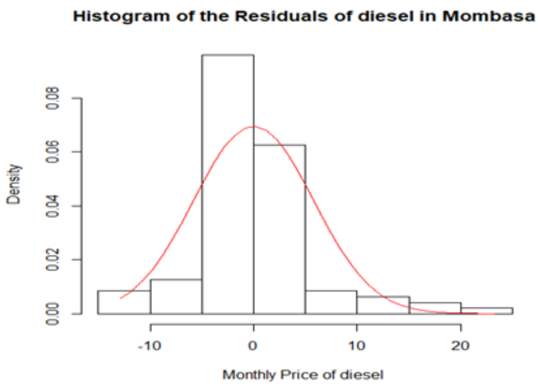

Fig.4: Histogram of diesel residuals

The histograms above show that the residuals appear to be normally distributed since normal curves fitted in all the two products and the mean of the distribution seems to be zero.This confirms the residuals for ARIMA (1,1,0) are normally distributed. Box-Ljung tests were also conducted to test the autocorrelations in the ACF lags and the results shown in the figures below.

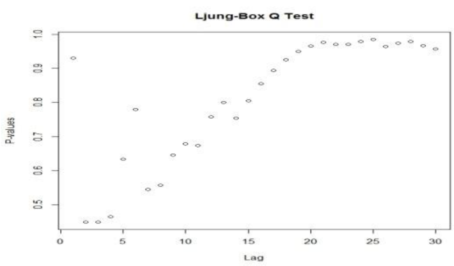

Fig. 5: Box- Ljung test for super

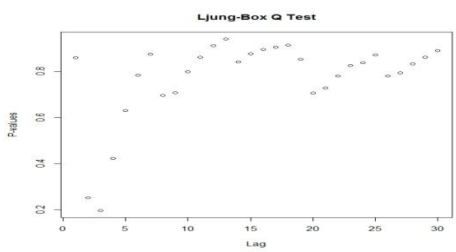

Fig.6: Box-Ljung test for diesel

From the two figures above, the p values for the Box- Ljung test are all over 0.05 which indicates non significance. The residuals appear to be uncorrelated at 1% significance level. Hence the model is a good fit and we fail to reject the null hypotheses.

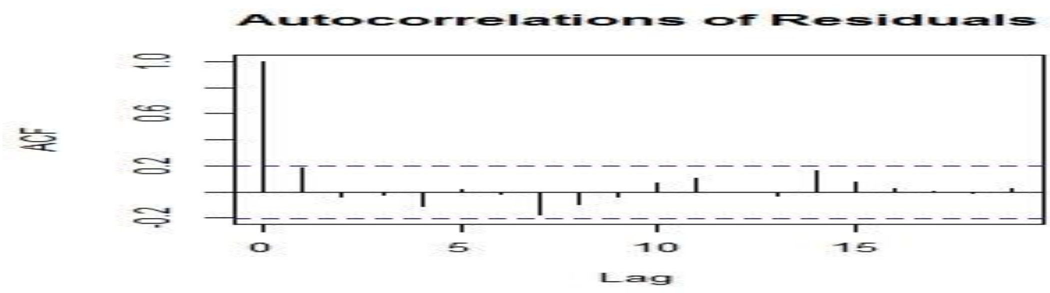

Fig. 7: ACF plot for super residual

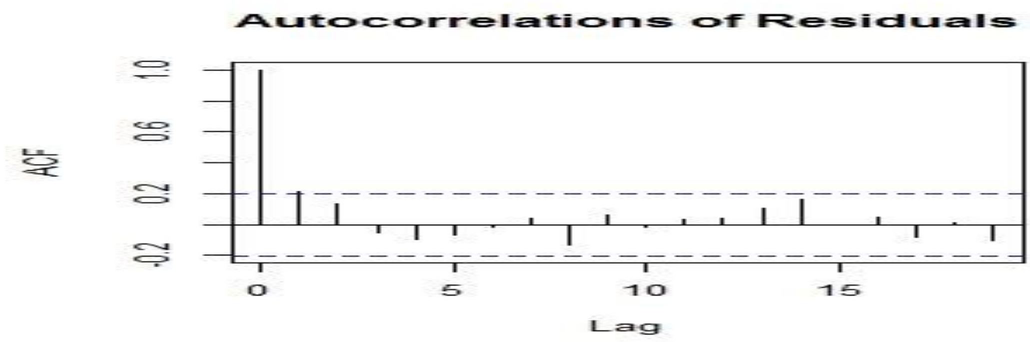

Fig.8: ACF plot for diesel residuals

The ACF plot residuals above also support the view that the residuals are uncorrelated at 1% level of significance. There are no significant spikes. The spikes at lag 1 in all the products are as a result of randomness in the petroleum products prices. The ARIMA(1,1,0) model is therefore good fit for the two products. Forecasting was done for the next twelve months.

After predicting using the ARIMA(1,1,0) model, it can be observed that the prices of the petroleum products are not stable. They keep on fluctuating over time, hence the high volatility.

Table no.4: General trend of forecasted prices for the two products

Month	**Predicted Super price**	**Standard error**	**Predicted Diesel price**	**Standard error**
January,2019	109.9522	5.6664	109.5759	5.2167
February,2019	109.9547	7.1514	109.5752	6.5719
March,2019	109.9820	8.3838	109.5750	7.6971
April,2019	109.9655	9.4579	109.5757	8.6782
May,2019	109.8689	10.4219	109.5804	9.5594
June,2019	109.7528	11.3041	109.5801	10.3658
July,2019	109.0924	12.1222	109.5915	11.1139
August,2019	109.5304	12.8886	109.5854	11.8142
September,2019	109.5749	13.6118	109.5870	12.4763
October,2019	109.4732	14.2986	109.5900	13.1045
November,2019	109.6681	14.9538	109.5894	13.7039
December,2019	110.1250	13.7159	109.5850	13.4347

Vector Autoregressive (VAR) Modeling

Dependent variable in the vector autoregressive modeling was the petroleum prices. It has thee independent variables, the international oil prices, exchange rate and inflation rate. These are the main factors affecting petroleum prices in Kenya. Table no. 5 below shows a display of the summary for the four variables. Let A denote the international oil prices, B denote the exchange rate and C denote the inflation rate.

Table no. 5: Descriptive statistics for the dependent and independent variables

Statistics	*Mombasa price*	*Int'l oil prices A*	*Exchange rate B*	*Inflation rate C*
Mean	*102.26*	*78.01*	*94.00*	*7.73*
SD	*10.44*	*26.53*	*7.92*	*3.87*
Median	*105.06*	*76.05*	*93.05*	*6.57*
Trimmed	*102.90*	*78.64*	*94.15*	*7.05*
Mad	*10.96*	*39.33*	*12.00*	*1.78*
Minimum	*77.43*	*29.78*	*80.74*	*3.20*
Maximum	*124.49*	*117.79*	*105.38*	*19.72*
Range	*47.06*	*88.01*	*24.64*	*16.52*
Skew	*-0.46*	*-0.12*	*-0.04*	*1.58*
Kurtosis	*-0.83*	*--1.63*	*-1.73*	*1.73*
SE	*1.07*	*2.71*	*0.81*	*0.40*
Observations	*96*	*96*	*96*	*96*

Table 5 shows the descriptive statistics for the variables. The large margins between the maximum and minimum values of all the series indicate evidence of significant variations of the trend of the series. The Mombasa prices, international prices and exchange rate show evidence of negative skewness implying extreme left tail. The inflation rate indicate positive skewness which denotes the right tail is extreme.

Stationarity testing

The three independent variables were not stationary. Differencing was therefore necessary to make them stationary.

Unit Root Test

To test for unity, ADF test was applied to the three independent variables.

All the tests produced ADF statistical values which are more than the respective critical values at the various levels. The null hypotheses are therefore rejected and conclude that the independent variables are non stationary. They become stationary after first order differencing.

Table no. 6: ADF test results for checking the stationarity of the three independent variables

H_0: International prices have unit root vs H_1: International prices don't have unit root.

International prices	Before differencing		First order differencing	
	t- Statistics	Probability	t-Statistics	Probability
The ADF Test	-1.7393	0.6842	-4.3719	0.01
Critical values for the Test 1% 5% 10%	-4.04 -3.45 -3.15			

H_o: Exchange rate have unit root H_1: Exchange rate don't have unit root

Exchange rate	Before differencing		First order difference	
	t- statistics	Probability	t-Statistics	Probability
The ADF Test	-2.2812	0.4603	-5.4120	0.01
Critical values for the Test 1% 5% 10%	-4.04 -3.45 -3.15			

H_o:

Ho:Inflation rate rate have unit root vs H_1: Inflation rate doesn't have unit root.

Exchange rate	Before differencing		First order difference	
	t- statistics	Probability	t-Statistics	Probability
The ADF Test	-3.5543	0.0412	-6.2151	0.01
Critical values for the Test	-4.04			

1% 5% 10%	-3.45 -3.15	

Test for cointegration

This was for checking if there is meaningful relationship between dependent and independent variables. Dependent variable is petroleum price in Mombasa while the international price (A), exchange rate (B) and inflation rate (C) are the independent variables. The null hypothesis for running the test is no co-integration. After using Portmanteau test, the p – value was 0.007063. Since p value is less than 0.05 significance level, the null hypothesis is rejected. Therefore, a relationship between the dependent and independent variables exists

Estimation Results

Maximum Likelihood Estimation method was used in parameter estimation. Dependent variable is the petroleum price in Mombasa while the dependent variables are international oil price (A), the exchange rate (B) and the inflation rate (C). The estimated VAR model is;

$\mathbf{Y_{(t)} = -52.91 + 0.61Y_{(t-1)} + 0.35A_{(t-1)} + 0.69B_{(t-1)} - 0.27C_{(t-1)}}$. The international oil prices and the exchange rate play a very significant role in determining the petroleum product prices in Kenya.

Table no.7: Estimate results for independent and dependent variables using VAR

Variable	Estimate	Std error	t value	Pr (>(t))
Petrol price $Y_{(t)}$	0.60698	0.05186	11.704	$< 2e^{-16}$
Int'l price (A)	0.34681	0.04083	8.493	$3.89e^{-13}$
Exchange rate (B)	0.68513	0.04083	7.727	$1.49e^{-11}$
Inflation rate (C)	-0.27373	0.07710	-3.551	0.000614
Constant	-52.90637	8.45250	-6.2591	$1.29e^{-0.8}$

Residual std error = 2.623, Adjusted R-squared = 0.9545, Multiple R-squared = 0.9565
$F_{(4,90)}$ statistic = 494.3 d.f = 90, p-value < 2.2e-16

Diagnostic Checking

After performing a residual diagnostic serial correlation for the residuals, the value of R squared was 95.65% which is greater than 5%. This implies nonexistence of serial correlation and this is a good sign. Since the kurtosis value is between -2 and +2, the model residuals are normally distributed.

Normality testing using Jarque- Bera statistic showed that p value was greater than 5% which indicates that our VAR model is normally distributed

Forecasting

Since residuals were normally distributed, the model was used to forecast for the next twelve months. Table no. 8 show the VAR forecasted values of the dependent variable. The values indicate that the variable is not constant at all. The values keep on fluctuating hence the volatility of the petroleum products prices. It can be observed that there is little variation of prices within the year. The forecasted Mombasa oil prices were decreasing for the next twelve months.

Table no 8: VAR forecasted Mombasa oil prices

Month	Forecasted	Lower	Upper	SE
January,2019	104.4538	99.1600	109.7476	5.2938
February,2019	100.4685	93.2409	107.6960	7.2275
March,2019	98.0572	88.7631	107.3510	9.2941
April,2019	96.6357	85.5004	107.7710	11.1352
May,2019	95.8233	83.2207	108.4259	12.6025
June,2019	95.3756	81.6614	109.0899	13.7142
July,2019	95.1381	80.5966	109.6797	14.5415
August,2019	95.0152	79.8561	110.1743	15.1591
September,2019	94.9493	79.3206	110.5779	15.6286
October,2019	94.9068	78.9106	110.9031	15.9962
November,2019	94.8696	78.5757	111.1635	16.2939
December,2019	94.8284	78.2849	111.3718	16.5434

V. Discussion

The **ob**jective of the study was to model petroleum prices in Kenya using ARIMA and VAR model. Once constructed, these models are used to predict petroleum products prices in the near future. Prediction was made for the next twelve months. The study showed that these prices are non-stationary which implies the non

-stability of the prices.
Table 9 below shows the comparison of ARIMA and VAR models.
It compares the prediction accuracy of the two models using root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).
The VAR model had the least error of 0.9332% mean absolute percentage error.

Table 9: Comparison of ARIMA (1,1,0) and VAR models

Variable	ARIMA (1,1,0)	VAR
Root Mean Square Error	3.696349	2.07
Mean Absolute Error	2.562906	0.936
Mean Absolute Percentage Error	2.54307	0.9332

VI. Conclusion

VAR is a better model for predicting petroleum prices in Kenya compared to the ARIMA (1,1,0) model. The international oil prices and the exchange rate have a great effect on the prices of petroleum products in Kenya.

VII. Recommendations

The general recommendation is that further studies can be undertaken to examine the effectiveness of other models like Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and simple exponential smoothing in modeling and forecasting of petroleum prices in Kenya.

These can then be compared with ARIMA and VAR models and determine which one best analyses the volatility of the petroleum prices. Research can also be done to establish other factors that influence petroleum prices in Kenya.

References

[1]. Kotut C, Menjo K, Jepkwony J. An Econometric Approach to the Economic Impact ofPetroleum oil prices Fluctuation in Kenya. Journal of Economics and SustainableDevelopment 2012; 3:7.

[2]. Nyongesa D, Wagala A. Non- Linear Time Series Modeling of the Diesel Prices in Kenya. International Journal of Academic Research in Economics and Management Sciences. 2016.

[3]. Mecheo J and Omiti, J., Petroleum markets structure and pricing following irregulation. Institute of Policy Analysis and Research. 2003.

[4]. Munyua JN, Ragui M. Drivers of Instability in prices of petroleum products in Kenya. Prime Journal of Business Administration and Management (BAM), 2013; 3(3): 919-926.

[5]. Al hassan A and Klishi A.Analyzing Oil Price-Macroeconomic Volatility in Nigeria. Journal of Applied Statistics 2016; **7**(1)

[6]. Kuhe D and Egemba R.Modeling and Forecasting CPI inflation in Nigeria: Application of Autoregression Integrated Moving Average Homoscedastic Model, Journal of Scientificand Engineering Research.2016; 3 (2): 57 – 66.

[7]. Habimana and others. Modeling and Forecasting of the CPI (Case of Rwanda). AmericanJournal of Theoretical and Applied statistics. 2016; **5**(3): 101 – 107.

[8]. NyoniT.Demystifying Inflation dynamics in Rwanda, Munich University Library Munich Personal Re (MPRA) paper No. 93982, (2019).

[9]. Nath B. Dhakre D and Bhattacharya, D., Forecasting Wheat Production in India: An ARIMA modeling approach. Journal ofpharmacognosy and photochemistry. (2018).

[10]. Fattah J and others.Forecasting of demand using ARIMA model. International Journal of Engineering BusinessManagement 2018;10: 1 – 9.

[11]. Bichanga LA. Statistical Models for Forecasting Petroleum Pump Prices in Kenya. Open Science Journal of statistics and Application. 2018; 5(3): 58 – 62.

[12]. Wiri L and Lebari, T. Modeling the Nigeria crude Oil Prices Using ARIMA, Pre – Intervention and Post – Intervention Model, Asian Journal of Probability and Statistics 2019; 3(1): 1 – 12

[13]. Dritsaki C. The Performance of hybrid ARIMA – GARCH Modeling and Forecasting OilPrice. International Journal of Energy EconomicPolicy.2018; 8(3):14–21.

[14]. Shah PA, Kiruthiga G.Crude Oil Price Forecasting Using ARIMA model. International journal of Engineering and Technology. 2020; **7**(3).

[15]. GurudeoAT ,Tareq S.Oil Price Forecasting Based on Various Univariate Time Series Models, American Journal of Operations Research. 2016; 6(3).

[16]. Guo N, Wei C, Manli W, Zijan T, Haoyue J. Applying an Improved Method Based on ARIMA Model to Predict the Short – Term Electricity Consumption Transmitted by the Internet of Things (IoT). Wireless Communications and Mobile Computing. 2021

[17]. Sharma S, Yadav, M. Analyzing the Robustness of ARIMA and the Neural Network as a Predictive Model of Crude Oil Prices. Theoretical and Applied Economics (2020);27: 289 – 300.

[18]. Chang L, Shi, Y.Mortality Forecasting with Spatially Penalized Smoothed VAR Model. AstinBulletin. (2020); 51: 161 – 189.

[19]. Ghazo A. Applying the ARIMA model to the process of Forecasting GDP and CPI in the Jordanian Economy. International Journal of Financial Research. 2021; 12(3),