Chuka University
info@chuka.ac.ke | 020 231 0512 / 020 231 0518

SARIMA MODELS: REVIEW AND ITS APPLICATION TO KENYAN'S COMMODITY PRICE INDEX OF FOOD AND BEVERAGE.

Teddy, M. Wanjuki*[*1], Adolphus Wagala[1] and Dennis, K. Muriithi[1]
[1]Department of Physical Sciences, Chuka University, P.O. Box 109-60400, Chuka, Kenya
**Corresponding Author Email: wanjukiteddymutugi@gmail.com; tmutugi@chuka.ac.ke; dkariuki@chuka.ac.ke*

How to cite:
Wanjuki, T. M., Wagala, A. and Muriithi, D. K. (2021). Sarima models: review and its application to Kenyan's commodity price index of food and beverage. *In: Isutsa, D. K. (Ed.). Proceedings of the 7th International Research Conference held in Chuka University from 3rd to 4th December 2020, Chuka, Kenya, p. 574-586*

ABSTRACT
Attaining price stability is one of the objectives of monetary policy in any economy to protect both consumers' and producers' interest. Unpredictable food and beverages prices make it difficult for consumers to plan for their expenditure in case of unexpected inflation. On the flip side, low prices may hurt producers as they may not be able to protect their profit margins. It is therefore imperative to develop a precise and accurate model to forecast Kenya's commodity prices. Therefore, the current sought to model the commodities price of food and beverage in Kenya using a Seasonal Autoregressive Integrated Moving Average (SARIMA). SARIMA model takes into account the seasonal periodic fluctuations in a series that usually recur with about the same time interval. Secondary data on monthly food price index was obtained from the KNBS website. The data covered the period from January 1991 to June 2017 with a total of 318 monthly observations. Data analysis was carried out using the R-statistical software. Using the Maximum Likelihood Estimation method, the SARIMA (0,1,2) $(0,1,1)_{12}$ model had better forecasts accuracy than other competing orders based on the Bayesian Information Criterion (BIC=1638.42) criterion with MAE of 2.25 in its forecasting ability. The two-year predictions of food and beverages price index showed an oscillatory behaviour with an increasing trend. The forecasts can help consumers adjust expenditure in preparation for periods of inflation. Policymakers should make priorities to ensure stability of future commodity prices.
Keywords: SARIMA Models, Consumer Price Index, Monetary Policy, Price Stability, Modelling, Forecasting

INTRODUCTION
Inflation is a general increase in price levels over time. Prices of different consumer goods and services are evaluated by the Consumer Price Index (CPI). Thus, the monthly or yearly inflation rate is determined by comparing the CPI for a given month to the CPI of that same month in the previous year. Price stability remains the primary objective of monetary policy formulation and implementation (CBK, 2017). The Central Bank of Kenya (CBK) through its monetary policies is mandated to maintain stable price levels to avoid inflation (Mutwiri, 2017). However, with the fluctuation in oil prices, other commodity prices are bound to fluctuate. Kaufmann (2011) & ECB (2010) attributes the surging oil prices to the role of supply sides contributed by both OPEC and non-OPEC countries. According to OECD report 2011, rise in oil prices results from the increasing oil demand, declining oil production by OPEC coupled with geopolitical and climatic conditions. Price fluctuation in commodity prices may lead to deterioration in the balance of payments and the resultant uncertainty in future prices is likely to affect investment and depress economic growth. Price volatility is a major feature of commodity markets and calls for understand future trends to make informed decisions on investments and government policies concerned with price control (Knaut & Paschmann, 2019).

Several economic studies have employed econometric tools to examine how inflation affects most crucial macroeconomic variables like consumption, investment, and economic growth. For example, using Structural vector autoregressive (SVAR) and Granger causality techniques, Mwangi *et al*. (2013) researched the linkages between product prices and both general inflation (non-food and non-fuel inflation). It was believed that food price and oil prices shocks had a key role in building up of relentless inflationary pressures in the Kenyan context. Food and oil prices were found to have a significant role in predicting inflation. Moreover, the effect of oil prices was more persistent in determining inflation than that of food prices. Ngwen *et al*. (2015) evaluated the linkage between inflation, growth and progress of the economy and government expenditure, in the case of Cameroon. The study outcomes found a long-run association between growth and economic progress, government expenditure and inflation. In another study, Maxwell *et al*., (2000) discovered that in the short run, both the inflation and government disbursements directly impacted on the economic progress. Monitoring these price indices is therefore important as it helps in making evidenced-based policies concerning consumption and investment. Deflation may affect the economy as it reduces the profitability of firms hence act as a disincentive to investment. It is therefore imperative that price levels be stabilized in a manner that drives economic growth. In Kenya, inflation is controlled by the monetary policy committee who provides policies aimed at maintaining price stability (Onderi & Njuru, 2015).

The Box-Jenkins (1960) techniques have gained popularity in the modelling of time series exhibiting seasonality. Specifically, SARIMA models have found their applicability in many fields and areas. It has been adopted mostly in modelling either monthly or quarterly data. Saz (2011) highlighted the comparative advantage of using SARIMA models in terms of parsimony and the efficiency of data modelling especially for a series exhibiting seasonal periodic fluctuation over time for instance Kenya commodity price index. A study which dwelled on consumer price index was done by Nyoni (2019) using annual time series data on CPI in Belgium from 1960 to 2017, he modelled and forecasted CPI using the Box
– Jenkins ARIMA technique. The study presented the ARIMA (0, 2, 1) model for predicting CPI in Belgium. The model used in the study did not account for the seasonal component of the CPI data as it used annual time series data. Research by Kumar *et al.* (2014) fitted an ARIMA (2, 1, 0) and forecasted annual sugarcane production data in India from 1950 to 2012. This model was able to forecast sugarcane production in India and also predict an increase in sugarcane production in 2013 and a sharp decrease in 2014. Mwanga *et al.* (2017) when modelling quarterly sugarcane yields data and forecast future yields based on the past values found out that SARIMA (2,1,2) (2,0,3)4 as the best model which had the lowest AIC and thus a good fit to the quarterly sugarcane yields data from 1973-2014. They conclude that Seasonal ARIMA models are proving to be good candidates of modelling time series with seasonal patterns and can be applied in any sector.

The most recent study was done by Mutwiri (2019) on Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya using the SARIMA model. The selected SARIMA (2, 1, 1) x (1, 0, 1)12 model was used to forecast the mean monthly real tomato prices from January-2011 to December-2011 by using the observed data of the period January- 2003 to December-2016. The study only addressed one particular commodity that is Tomato wholesale prices in a particular region and did not focus on the average CPI for food and beverage prices in Kenya. Therefore, this study will focus on modelling and forecasting of food and beverage prices in Kenya using SARIMA models using monthly series data on CPI from 1991 to 2017 to provide a model that may help track price indices in Kenya, particularly food and beverage prices. The modelling process will be done following the Box-Jenkins (1976) model building approach which entails, identification, specification of the model, model estimation and model diagnostics.

Seasonal Autoregressive Integrated Moving Average Models

According to Shekhar & Williams (2007), SARIMA models are an adaptation of autoregressive integrated moving average (ARIMA) models used to fit seasonal time series data. That is, it takes into account the underlying seasonal nature of the series to be modelled. SARIMA model is suitable when the series contains both seasonal and non- seasonal behaviour (Kumar & Vanajakshi, 2015). Seasonality in a time series refers to a regular pattern of changes that repeats over in time-periods, where h defines the number of time-periods until the pattern repeats (Verbesselt *et al*., 2010). Data observed every month has h being 12. In a seasonal ARIMA model, seasonal AR and MA terms predict using past values and errors at times with lags that are multiples of S (the span of the seasonality).

Seasonality behaviour is common on most time-series data makes the ARIMA model inefficient to be applied to the series. The seasonal ARIMA model incorporates non-seasonal and seasonal factors in a multiplicative model and is denoted as the SARIMA model (Mutwiri, 2019). Non-seasonal ARIMA models are usually denoted as ARIMA (p, d, q) where the parameters p, d, and q are non-negative integers and p is the order of the autoregressive model indicating the number of time lags, d is the degree of differencing, and q is the order of the moving-average model (Wang & Niu, 2009). However, the seasonal ARIMA model incorporates non-seasonal and seasonal factors in a multiplicative model and is denoted as ARIMA (, ,) (, ,)h (Chen & Wang, 2007).

The first part with the lower notation consists of the non-seasonal component and the second part with upper case notation consist of the seasonal part. The h is the number of periods in each season or the period of repeating seasonal pattern. That is, the seasonal component repeats itself after every h observations. Periodicity (h) can be 12, for monthly data or four for quarterly time-series data. Chatfield (2000), states that if the series is seasonal, with s periods per year, then a Seasonal ARIMA (hereby abbreviated SARIMA) model may be obtained as a generalization of an ARIMA.

The SARIMA model with seasonality can be represented as follows:

$$(^{h})() - = ()(). \quad (1)$$

Where; (h) is the seasonal AR operator, ( ) is the non-seasonal AR part, ( ) is the seasonal MA operator, $\Theta(B)e_t$ is the non-seasonal MA operator, is the time series of the observation at time t, B is the backshift operator, e_t is the white noise, h is the number of periods in a year.

Applying the backward shift operator, the operators can be written as:

$$(h)=1- {}_1{}^{h}- {}_2{}^{2h}-\cdots- {}^{h} \quad (2)$$

$$( ) =1- \quad {}_{-1}-\cdots- \quad {}_{-} \quad (3)$$

$$( )=1+ {}_1 + {}_2{}^{2}+\cdots+ \quad (4)$$

$$( )e =1+ {}_1e_{-1}+\cdots+ \; e_{-} \quad (5)$$

Data and Methods

The study employed a time series research design. In this study design, measurements of the same variable are taken at different points in time. SARIMA modelling technique was applied to fit and forecast monthly Kenya's food and beverage prices index data from January 1991 to June 2017. The selected data has 318 data points. The secondary data was extracted from the website of KNBS (knbs.or.ke). The data analysis was done using R -Statistical Software (R Core Team, 2013). The preliminary stage entailed checking for stationarity of the data. SARIMA models are appropriate for stationary time series satisfying the condition of constant mean and variance over time. For a stationary time-series data, the scatter points in the time series plot appear to have a constant mean and variance and oscillates around the mean. Statistically, stationarity of the data was tested using the Augmented Dickey-Fuller (ADF) according to Dickey and Fuller (1979).

The ADF procedure attempts to retain the validity of the test based on white noise errors in the regression model by ensuring that errors are indeed white noise. If the ADF statistic will be greater than the critical value at 5% significance level then the null hypothesis that there is a unit root would be accepted and the series is considered non-stationary. If the series is found to be non-stationary, differencing at different degrees of freedom is done until a stationary series of a given degree of differencing is obtained. The main objective of seasonal differencing will be to remove seasonal trend and seasonal random walks in our data. Chatfield (2000), further elaborates that a seasonal autoregressive term, for example, is one which depends linearly on seasonal lagged values of the series. The SARIMA model with non-seasonal terms of the order (p, d, q) and seasonal terms of the order (P, D, Q) is abbreviated as SARIMA(p, d, q)(P, D, Q) model and may be written as

$$()(1-) = ()() \quad (6)$$

where and denote polynomials in of order P, Q respectively. One model which is particularly useful for seasonal data is the SARIMA model of order (0,1,1)(0,1,1) $_{12}$. Shumway and Stoffer (2006) confirmed that the multiplicative seasonal autoregressive integrated moving average model, or SARIMA model, of Box and Jenkins (1970) is given by;

$$()()\nabla\nabla = + ()() \quad (7)$$

Where: is the Gaussian white noise process. The general model is denoted as ARIMA (p d q) (P, D, Q)$_{12}$. The ordinary autoregressive and moving average components are represented by polynomials ( ) and ( ) of orders p and q, respectively and the seasonal autoregressive and moving average components by ( ) and ( ) of orders P and Q and ordinary and seasonal difference components by

$$\nabla=(1-)\nabla =(1- \;) \quad (8)$$

Since the data of interest in this study is monthly data, s = 12, hence our equation will be written as,

$$(1-)(1- {}^{12}) =(1+ \;) ({}^{12}) \quad (9)$$

under this method, we shall employ monthly CPI data with 12 seasons per year (s= 12), where the first-order AR(1) model will use $_{-12}$ to predict , while the seasonal first order MA(1) will use $_{-12}$ as its predictor.

Modelling Process

The first step of modelling according to Box and Jenkins approach will be identifying the suitable order of SARIMA (p, d, q) (P, D, Q) model. This will entail plotting and examining the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series to check the degree of internal correlation between observations in a time series and at different lags. The number of significant spikes suggests the order of the model.

The non-seasonal and the seasonal AR orders, p and P, will be estimated by observing the non-seasonal and the seasonal cut-off lags on the ACF plot. Similarly, the MA orders, q and Q, are respectively estimated by the cut-off lags of the partial autocorrelation function (PACF). For the order of differencing d and D depends on the order of differencing done to make the series stationary.

ACF AND PACF of a stationary time-series data

Autocovariance Function

Assuming the variance of is finite, the autocovariance function is defined as the second moment product

(,) = (,) = (,) = [(−)(−)], for all s and t. Note that (,) = (,) for all time points s and t. The autocovariance measures the linear dependence between two points on the same series observed at different times. Very smooth series exhibit autocovariance functions that stay large even when the t and s are far apart, whereas choppy series tend to have autocovariance functions that are nearly zero for large separations. Given that the autocovariance (,) of stationary time series depends on s and t only through| − |, the notation can be rewritten a + h, where h represents the time shift.

(+ ℎ,)= [(+ℎ − μ)(− μ)]= (+ℎ,)

(+,) = (ℎ, 0) = (ℎ).

Autocorrelation Function (ACF)

The autocorrelation function is defined as: (10)

( )=

√ (,)

(11)

According to Cauchy-Schwarz inequality:

|(,)|² ≤ (,) (,) (12)

it can be shown that −1 ≤ (,) ≤ 1.

The ACF measures the linear predictability of using only . Given a stationary time series data; the ACF can be expressed as:

(ℎ) = (+ℎ,) / √(+ℎ, +ℎ)(,) = (ℎ)/(0)

(13)

Partial Autocorrelation Function

Another important measure is called partial autocorrelation, which is the correlation between and with the linear effect of "everything in the middle" removed. For a stationary process , the PACF (denoted as ℎℎ), for ℎ = 1, 2, … is defined as:

11 = corr(Xt+1, Xt) = 1

ℎℎ = corr(Xt+h − X̂t+h , Xt − X̂t), ℎ ≥ 2 = 1 (14)

Where; X̂t+h and X̂t is defined as:

X̂t+h = 1 t+h-1 + ⋯ + ℎ−1 +1 (15)

X̂t = 1 +1 + ⋯ + ℎ−1 +ℎ−1 (16)

If is Gaussian, then ℎℎ is actually conditional correlation

ℎℎ = (, +ℎ| +1 , +2 , ⋯ , +ℎ−1) (17)

Competing orders of the models were compared based on minimization of the Akaike Information Criterion (Akaike, 1977), Schwarz Criterion (Schwarz, 1978) and Hannan-Quinn Criterion (Hannan & Quinn, 1979). Model evaluation criterion in time series modelling and forecasting can be done using various parameters such as the Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC). Several competing models may be ranked according to their AIC, or BIC with the one having the lowest information criterion value being the best. These information criteria judge a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. The criterion attempts to find the model that best explains the data with a minimum of free parameters but also includes a penalty that is an increasing function of the number of estimated parameters. This study employed Akaike's Information Criteria and the BIC to select the best model which is parsimonious. A model which has a lower value of these information criteria will be preferred.

Akaike' s Information Criterion

According to Akaike (1974) and Schwarz (1978), the AIC function is denoted by:

= ((18)

where; =Number of observations, =Number of parameters, and =Error sum of square

According to (Zhang, 2013), the AIC statistic can also be obtained by:

$= -2\log(\ ) + 2\log(\ +\ )$, (19) where indicates the likelihood of the data with a certain model, and indicates the lagged orders of AR and MA term respectively.

Bayesian Information Criterion

According to Gideon E. Schwarz (1978), BIC can be obtained as;

$= -2\log(\ ) + \log(\ )$ (20)

Or $= \log(\ ^2)$ ( ) . (21)

where; k: is the number of parameters in the statistical model, (p+q+P+Q+1), L: is the maximized value of the likelihood function for the estimated model, n: is the sample size, and

different reasons. While AIC tries to approximate the model towards the reality of the situation, BIC tries to find the perfect. The BIC approach is criticized as no perfect model can be obtained however it is useful since it penalizes the model more heavily for having more parameters as compared to AIC (Hyndman *et al.,* 2015).

Model Diagnostics

The significance level of individual coefficients was measured by Box-Pierce Q statistics and jointly together by Ljung-Box LB statistics. The model is considered adequate if and only if the p-value associated with the Ljung-Box Q statistic is higher than the critical level of significance. Moreover, the correlogram of the residuals can be used to check residuals' autocorrelation. If there is no serial correlation, the autocorrelations and partial autocorrelations at all lags should be nearly zero, and all Q-Statistics should be insignificant with large probability values.

Model Accuracy Evaluation

The accuracy of the model selected for forecasting must be higher than that of all the competing models. Hyndman & Koehler (2006) studied and compared all measures accuracy and settled on Mean Absolute Scaled Error (MASE) as the best measure of accuracy on forecasting. Thus, the accuracy of the model in this study employed all the three methods namely; MAE, MAPE and MASE for testing the model accuracy. The model with the minimum of these values is considered to be the best for forecasting. In a perfect forecast, the MAE = MASE = MAPE = 0. The smaller the value the better the prediction and the big the value the poorer the predictive ability of the model.

Root Mean Square Error (RMSE)

The RMSE is defined as:

$= \sqrt{\ }$ (25)

RMSE measures the "mean prediction error". For a "Perfect" fit the RMSE will be zero. It is used to measure the difference/residuals between values predicted by the model and the observed values.

Where: is the error sum of squares

$= \Sigma$ (26)

which implies that

$= \sqrt{\ } \quad = \sqrt{\ }$ (27)

Mean Absolute Error (MAE)
The MAE is defined as:

[illegible] , (28)

Where:is the actual observation value at time t, Ŷ is the forecasted observation value at time t, and is the total number of observations.

RESULTS AND DISCUSSION

Descriptive Statistics

This study aimed at analyzing the consumer price index data covering the period from January 1991 until June 2017, a total of 318 monthly observations. Table 1 shows that the series the least Food and Beverage Price Index (FBPI) to be recorded was 75.83 and the highest was 196. The series is more varying since the standard deviation is large (33.36). This is also a suggestion that the data is non-stationary. However, the measures of variation indicate that the data is approximately normally distrusted since the skewness statistic close to zero.

Table 83: Descriptive Statistics

Statistic	Min	Mean	Variance	SD	Max	Skewness	Kurtosis
Price Index	75.83	119.88	1113.21	33.36	196	0.624	0.906

Decomposing the Food and Beverage Price Index Data

A seasonal time series consists of a trend component, a seasonal component and an irregular component. Decomposing the time-series means separating the time series into these three components. Seasonal decomposition, as used by Doulah (2018) while examining temperature data in Bangladesh, was employed to determine whether the series exhibits the three components. As expected in a time series data like food and beverage index data, figure 1 shows that the original data has a changing variance over time with its associated three components and random error. The trend analysis shows an upward trend. Since the series has a seasonal component, modelling it using SARIMA is suitable. (Figure 1).

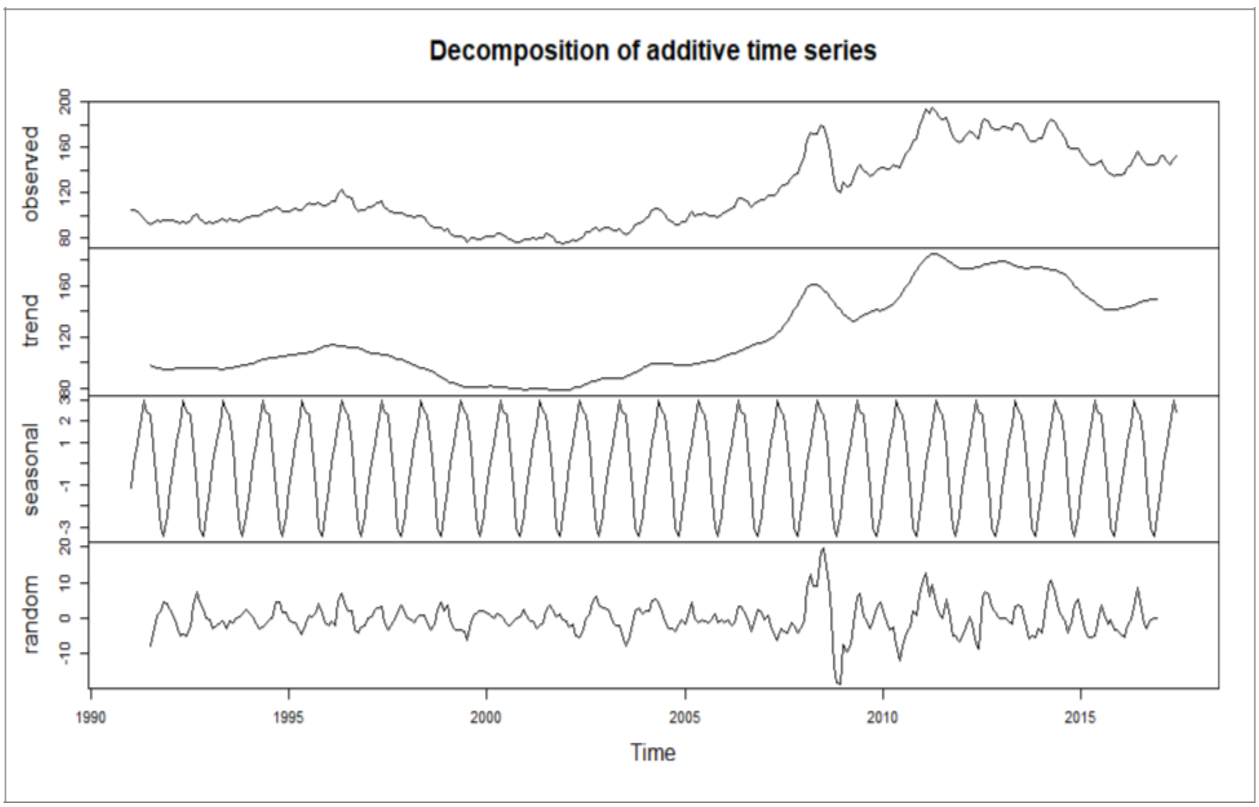

Figure 20: Decomposition of the Price Index Series

Testing for Non- stationarity

The plot in figure 1 shows the seasonal fluctuation in the monthly food and beverage price index data. The data appears to be not stationary. The non-stationary data has mean, variance, and correlation that changes over time. Generally, there is an increasing trend with no clear evidence of systematic variation about the mean on the time series plot. This makes the SARIMA model suitable since it captures seasonal variation a given series.

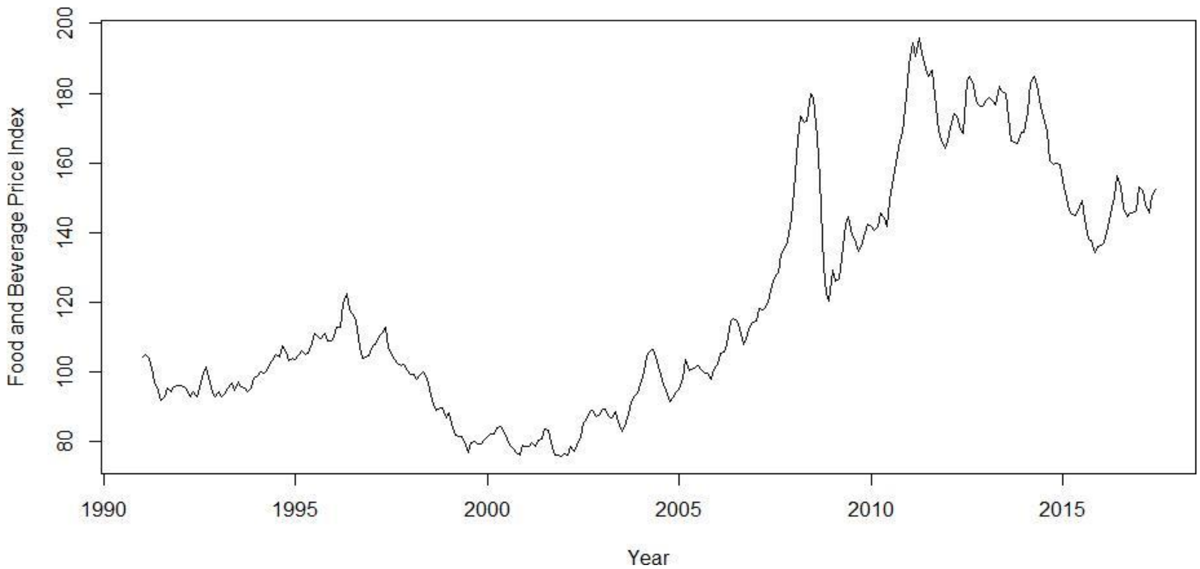

Figure 21: Time Series Plot of Food and Beverage Price Index

Moreover, with not stationary data the behaviour of the ACF and PACF plot may not aid in the identification of the order of both seasonal and non-seasonal AR and MA parts. As shown in figure 3, the ACF plot of the series does not tail of, even at lag 25. The coefficients on the autocorrelation function have a slow fall affirming that the series is non-stationary.

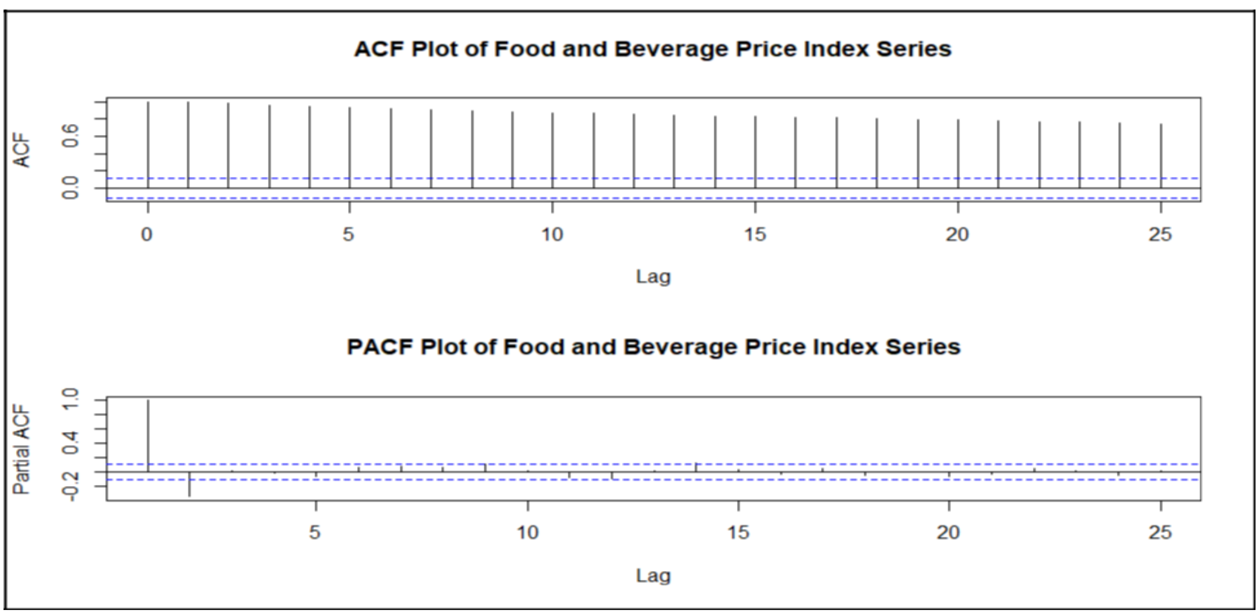

Figure 22: ACF and PACF Plots of Food and Beverage Price Index Series

Therefore, we get the seasonal differences of the series at lag 12 and then examine stationarity. Figure 4 shows the present monthly FBPI data and the associated ACF and PACF plots. An examination of the seasonally differenced series shows the presence of a trend. The associated ACF plot shows strong seasonal wave patterns that decline moderately. The non-seasonal lags decay rapidly confirming the presence of seasonal behaviour. Therefore, the series is non-stationary.

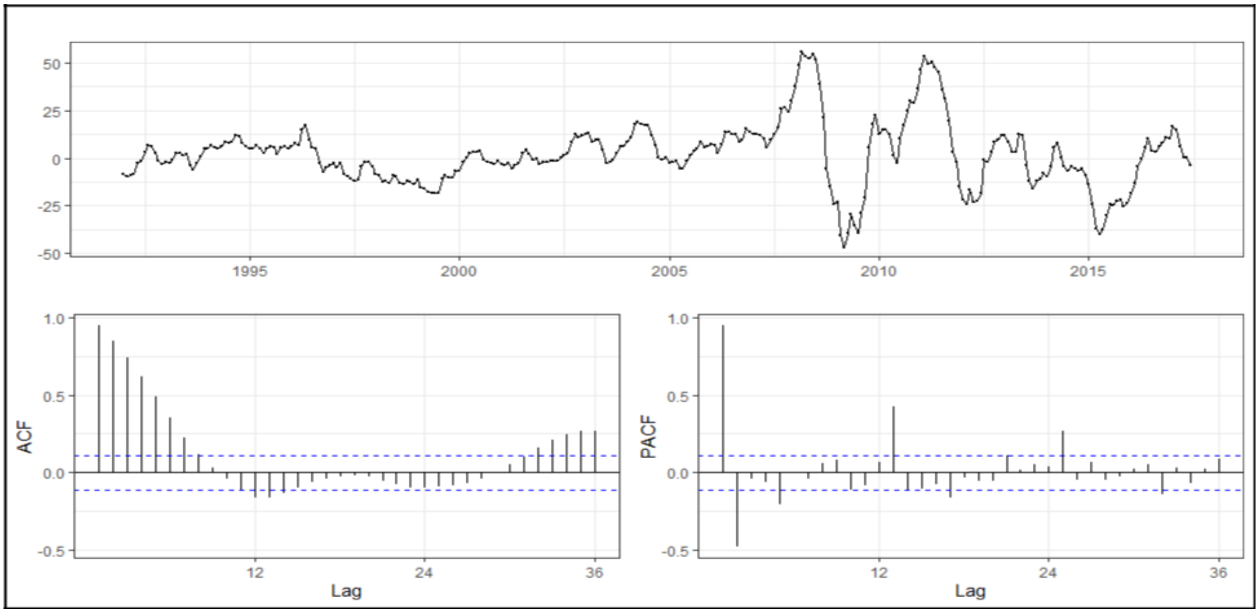

Figure 23: Time series plot of seasonally differenced series at lag 12 and Associated ACF

Since the series is not entirely stationary, thus, the first difference of the seasonally differenced series is applied on the resultant series. Based on this visualization in figure 5, differenced data has slightly more stationary properties and no obvious trend or seasonality makes it suitable for use in estimating the parameters of the model.

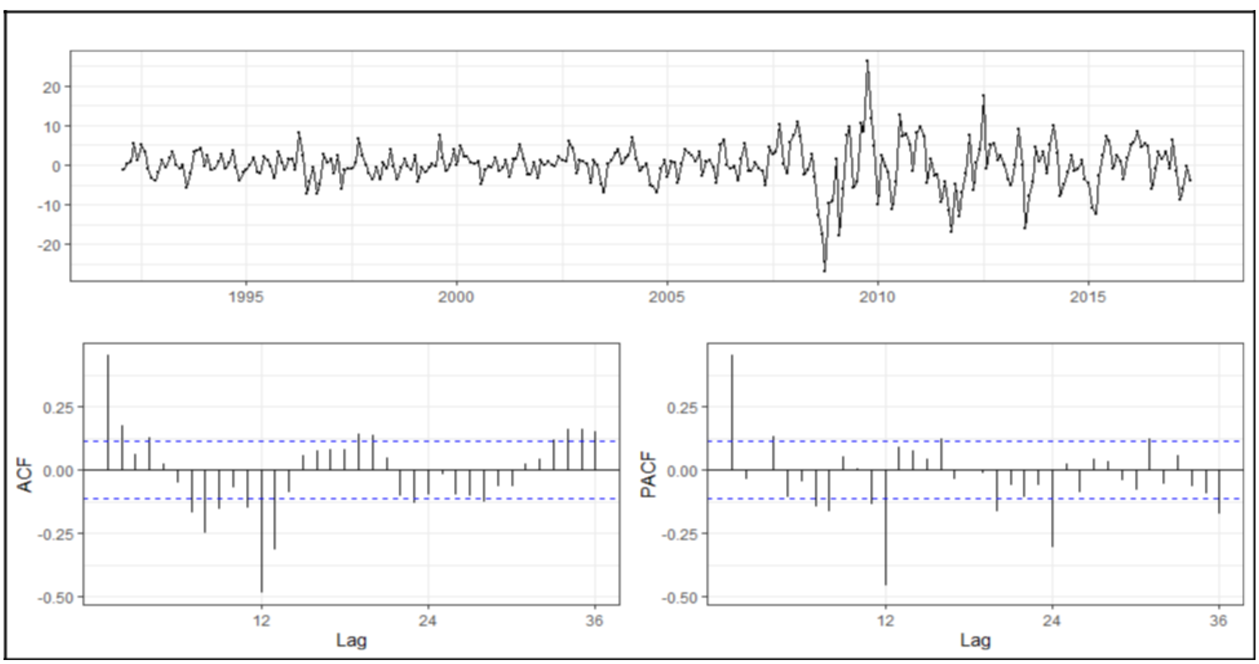

Figure 24: First difference of the seasonally differenced series at lag 12

To affirm the test, we applied the statistical test for the series stationarity using ADF unit roots tests. The results of the Augmented Dickey-Fuller (ADF) test on the FBPI series are represented in Table 2. The results indicate that series is stationary in the First difference and at the first difference of the seasonally differenced series at lag 12 ($p<0.05$). Thus, the model ARIMA (p, d, q) we will have the value d=1whereas the model ARIMA (P, D, Q) we will have the value D=0.

Table 84: Unit Root Test

	Level	First Difference	Seasonal Difference at lag 12 (SDL12)	First Difference of the SDL12
ADF	-2.0813	-7.3809	-5.4164	-7.0196
P-value	0.5422	0.01**	0.01**	0.01**

Model building for the monthly Food and Beverage Price Index Series

According to Shumway & Stoffer (2017), the process of model fitting entails data visualization, data transformation where necessary, model identification of dependence order, estimation of the parameter, diagnostic analysis and choosing an appropriate model. Thus, this section presents the second stage of the analysis of determining the best suitable SARIMA model.

Model Identification

After the detection of stationarity of the series, the order of both seasonal and seasonal part AR's and MA's process suiting the model is then determined by examining the behaviour of both ACF and PACF. Figure 4 shows ACF and PACF plots, taking seasonal lags up to 50 lags.

Non-seasonal behaviour:

We examine the early lags to determine the non-seasonal terms. Spikes in the ACF, at low lags, indicate non- seasonal MA terms. Whereas, spikes in the PACF, at low lags, indicate possible non-seasonal AR terms. The early lags in the PACF tails at lag 2. Thus, we can estimate the non-seasonal MA (2) process. The PACF shows a highly significant spike at lag 1 which then levels until lag 12, 24, and so on. However, the spikes dwindle with time. The sequence is accompanied by a tapering pattern in the lags of the ACF plot. Thus, a non-seasonal AR (1) can suit part of the model.

Seasonal behaviour:

To determine the order of this component we examine the patterns across lags that are multiples of S. For monthly data, as in this case, the behaviours of the spikes around lags 12, 24, 36, and so on, is examined. The ACF plot shows a cluster of (negative) spikes around lag 12 and lags 24, and 48. However, positive clusters of spikes are seen at lag 36. The PACF is tapering in multiples of S=12; That is the PACF has significant lags at 12, 24, and 36 akin to seasonal MA (1) process. Based on the ACF and PACF of the 12th differences, we can estimate an ARIMA (1,1,0) $(0,1,1)_{12}$.

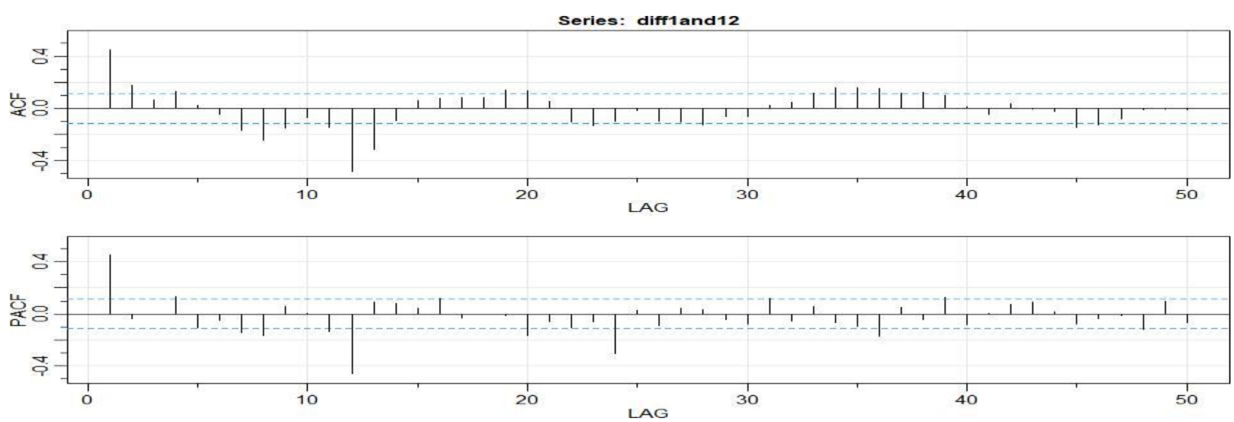

Table 85: AFC and PACF Plots of the First Difference of the Seasonally Differenced series at lag 12

It is a common practice to examine a combination of orders in a bid to determine a model that minimizes either of the chosen information criteria. Table 4 shows the comparison of models using AIC and BIC.

Table 86: Comparison of SARIMA Models Using AIC and BIC

SARIMA MODEL(Order)	S	AIC	BIC
(0,1,2) (1,0,0)	12	1673.422	1688.457
(0,1,2) (0,0,1)	12	1673.413	1688.449
(0,1,2) (0,1,1)	12	1638.415	1653.296
(0,1,3) (1,0,1)	12	1673.473	1696.026
(1,1,0) (1,0,0)	12	1674.78	1686.056
(1,0,0) (0,1,1)	12	1674.78	1686.056
(1,1,0) (0,1,1)	12	1639.008	1650.169
(1,1,1) (1,0,0)	12	1675.383	1690.418
(1,1,1) (0,0,1)	12	1675.374	1690.409
(1,1,2) (1,0,0)	12	1673.856	1692.651
(2,1,0) (1,0,0)	12	1675.108	1690.143
(2,1,1) (1,0,0)	12	1674.154	1692.949
(3,1,0) (1,0,0)	12	1676.813	1695.608
(3,1,1) (1,0,0)	12	1677.179	1699.732
(3,1,1) (0,0,1)	12	1677.173	1699.726
(0,1,3) (1,0,0)	12	1674.732	1693.527

The results from Table four indicate that according to Akaike (AIC) and Schwartz (SIC) criteria, SARIMA ((0,1,2) $(0,1,1)_{12}$ model is the most suitable. The parameters given are the non-seasonal specification of AR (0), differencing (1), and MA (2), followed by the seasonal specification of seasonal AR (0), seasonal differencing (1), seasonal MA (1), and period or span for the seasonality (12).

Estimation of the Model

The maximum likelihood estimation method is employed in the estimation of the coefficients of the model. An iteration procedure is required when estimating the parameters of SARMA models (Box & Jenkins, 1976). The results of the SARIMA ((0,1,2) $(0,1,1)_{12}$ model show that all the estimated SARIMA coefficients are statistically significant (Table 5).

Table 87: Table of Coefficients

	Estimate	SE	t. value	p. value
ma1	0.4625	0.0587	7.8743	0.0000
ma2	0.1501	0.0548	2.7386	0.0065
sma1	-1.0000	0.0796	-12.5645	0.0000

Diagnostics Checks of the Model

Normality Test

The standardized residuals estimated from the models should behave as an independently and identically distributed sequence with zero mean and constant variance. A normal probability plot or a Q-Q plot can aid in the identification of the departure of the distribution of residuals from normality. The residual plot shows that the standardized sample residual plots depict a constant variance. The normal Q-Q plots show that the residuals are approximately normally distributed since they lie along the 45^0 line (Figure 6).

Autocorrelation

There is no correlation between each lag of its values since the residuals don't have any significant autocorrelation as shown by the ACF plot (Figure 6). That is, the plots of the ACF of the residuals lack enough evidence of significant spikes an indication that the residuals are white noise. The results also showed that the residuals are non- significant with the Ljung – Box-Pierce statistics up to lag 7. That is the autocorrelation and partial autocorrelation coefficients are non-statistically significant in up to lag 7 since the associated Q-statistics have large probability- values. The results imply that that residuals are not autocorrelated hence do not call for the usage of ARCH models. The fitted SARIMA (0,1,2) $(0,1,1)_{12}$ model can therefore be used for forecasting.

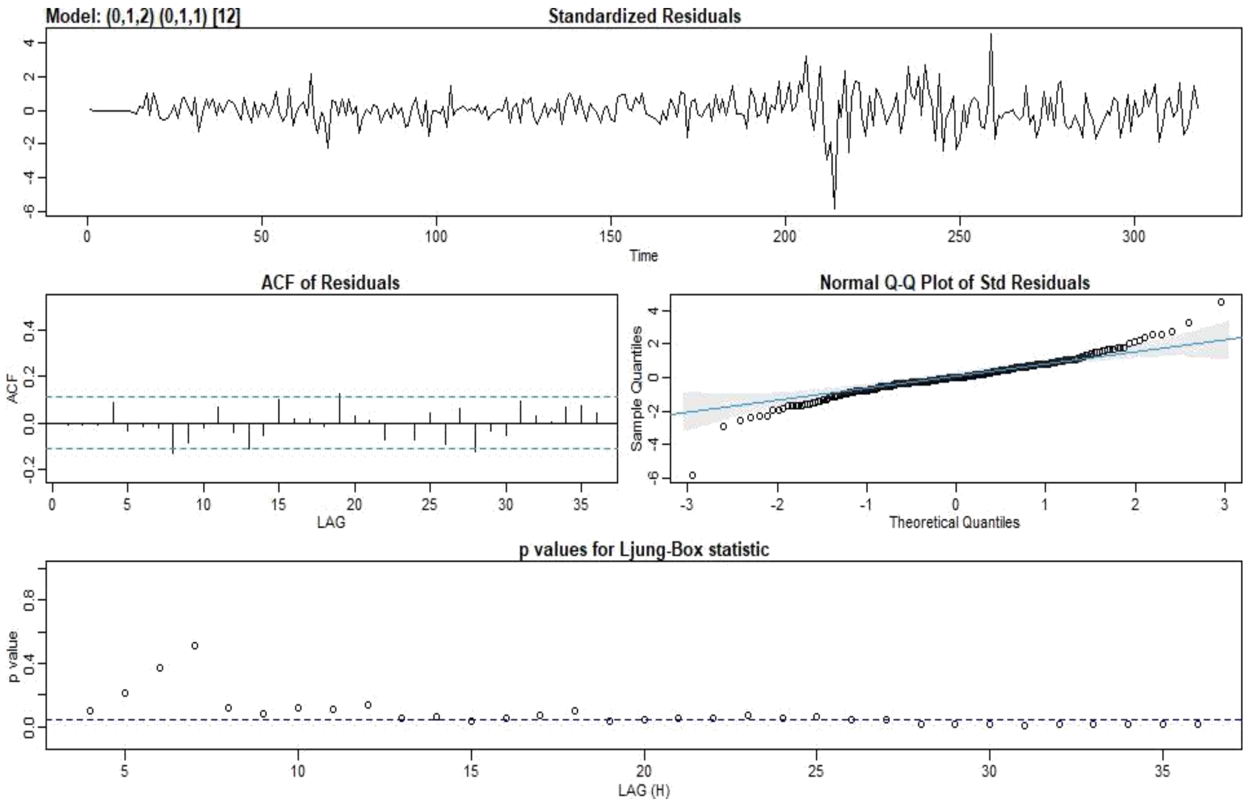

Figure 25: Diagnostic Checks

Forecasting

Forecasting helps in planning and decision-making process since it gives an insight into the future uncertainty using the past and current behaviour of given observations (Doulah 2018). Therefore, the resultant model was used to fit h-step ahead forecasts. The forecast for the next 24 months from July 2017 to July 2020 with the 95% confidence interval as shown in figure 7. The forecast results depicted a decreasing pattern of FBPI in the last quarter of 2017 followed by an increasing trend in the third quarter of 2018. Thereafter the price index increases steadily up to the third quarter of 2020. The accuracy of the model in making h-step ahead prediction was done using the accuracy-test summarized in table 6.

Table 88: Model Accuracy Test

Parameter	ME	RMSE	MAE	MPE	MAPE	MASE
Value	0.09650335	3.216398	2.252747	0.09048207	1.818399	0.8673462

CONCLUSION

Following the Box-Jenkins approach, this study assessed the ability of the SARIMA models to fit and forecast food and beverage and price index series of Kenya from January 1991 to June 2017. Based on minimum AIC, and BIC values, the findings showed that best order among competing orders of SARIMA was SARIMA (0,1,2) $(0,1,1)_{12}$ model and can forecast the FBPI for Kenya with MASE of 0.867. The forecasts covered the 24 months from July 2017 to July 2020. The forecast results depicted a decreasing pattern of an increasing trend from the third quarter of 2018 till the third quarter of 2020. Based on these forecasts, prices seem to be highly volatile. Thus, policymakers should take appropriate monetary policy options to stabilize the prices. Since price fluctuations have been attributed to surging world oil prices, there is a need to create a buffer stock to even out oil demand within the country more so as Kenya recently began oil exploration and production. The oil production and distribution should be managed to avoid intentional hoarding. Give that investment and consumption decisions continually rely on prospects, future on a comparison of this model with other competing models that capture seasonality is recommended.

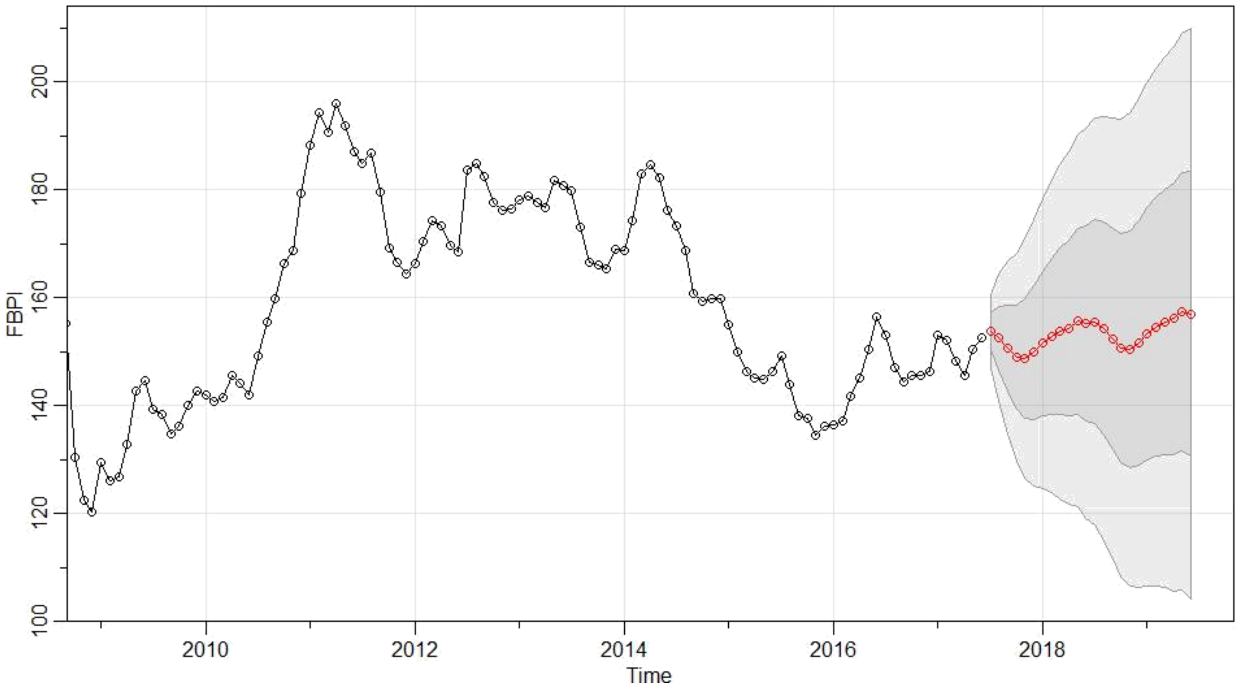

Figure 26: Actual Time Series Plot of FBPI and the h=24 Step Period Forecasts

Acknowledgements

My sincere gratitude goes to my supervisors: Dr. Adolphus Wagala, and Dr. Dennis K. Muriithi (Faculty of Science, Engineering and Technology, Chuka University) for their inspiration and guidance throughout this study.

REFERENCES

Akaike, H. (1977). On entropy maximization principle, Applications of Statistics. Krishnaiah PR.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control San Francisco. *Calif: Holden- Day*.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). Time Series Analysis, Forecasting and Control, 3rd ed.,

Prentice Hall, Englewood Clifs, Brockwell, Peter J., 1994.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). Time series analysis, forecasting and control. Englewood Clifs.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Chatfield, C. (2000). Time-series forecasting. CRC press.

Chen, K. Y., & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications, 32(1), 254-264.

Dickey, D., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, 74 (366), 427-431.

Doulah, M. S. U. (2018). Forecasting Temperatures in Bangladesh: An Application of SARIMA Models. *International Journal of Statistics and Mathematics*, *5*(1), 108-118.

Ding, F., Shi, Y., & Chen, T. (2006). Performance analysis of estimation algorithms of nonstationary ARMA processes. IEEE Transactions on Signal Processing, 54(3), 1041-1053.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological), 41(2), 190-195.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International journal of forecasting, 22(4), 679-688.

Hyndman, R. J., & Khandakar, Y. (2007). Automatic time series for forecasting: the forecast package for R (No. 6/07). Clayton VIC, Australia: Monash University, Department of Econometrics and Business Statistics.

Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review, 7(3), 21.

Kumar, M., & Anand, M. (2014). An application of time series ARIMA forecasting model for predicting sugarcane production in India. Studies in Business and Economics, 9(1), 81-94.

Knaut, A., & Paschmann, M. (2019). Price volatility in commodity markets with restricted participation. Energy Economics, 81, 37-51.

Ljung, G. M., & Box, G. E. P. (1978). "On a Measure of a Lack of Fit in Time Series Models", Biometrika, 65 (2), 297-303.

Mutwiri, N. M. (2017). Monetary Policy Tools and Inflation in Kenya. International Journal of Academic Research in Accounting, Finance and Management Sciences, 7(1), 86-97.

Maxwell Fry, D. J., Lavan Mahadeva, S. R., & Sterne, G. (2000). Key issues in the choice of monetary policy framework. Monetary policy frameworks in a global context, 2.

Muriithi. D.K, Ngeretha.A.N and Muriungi.R.G(2014). Analysis of The Fluctuation of Gross Domestic Product in Kenya Using Autoregressive Integrated Moving Average Model. *Journal of Statistics: Advances in Theory and Applications,* Vol. 11(1) pages 31-43

Mutwiri, R. M. (2019). Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya: Time Series Analysis Using Sarima Model. International Journal of Statistical Distributions and Applications, 5(3), 46.

Nyoni, T. (2019). Time series modeling and forecasting of the consumer price index in Belgium.

Ngwen, N., AMBA OYON, C. M., & Mbratana, T. (2015). Government expense, Consumer Price Index and Economic Growth in Cameroon.

Onderi, V. N., & Njuru, S. G. (2015). The impact of monetary policy on economic growth and price stability in Kenya: 1992–2013. IOSR Journal of Economics & Finance, 6(6), 95-109.

Shumway, R. H., & Stoffer, D. S. (2006). "Time Series Analysis and Its Applications with R Examples Second Edition" Springer Science Business Media, LLC Smoothing.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Saz, G. (2011). The efficacy of SARIMA models for forecasting inflation rates in developing countries: The case for Turkey. International Research Journal of Finance and Economics, 62, 111-142.

Shekhar, S., & Williams, B. M. (2007). Adaptive seasonal time series models for forecasting short-term traffic flow. Transportation Research Record, 2024(1), 116-125.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.
Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. Remote sensing of Environment, 114(1), 106-115.

Wang, W., & Niu, Z. (2009, September). Time series analysis of NASDAQ composite based on seasonal ARIMA model. In 2009 International Conference on Management and Service Science (pp. 1-4). IEEE.

Zhang, H., & Rudholm, N. (2013). Modeling and forecasting regional GDP in Sweden using autoregressive models. Dalama University.

****