International Journal of Statistics and Applied Mathematics 2020; 5(4): 229-245

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2020; 5(4): 229-245

www.mathsjournal.com
Received: 30-04-2020
Accepted: 02-06-2020

KK Saxena
Department of Statistics, University of Dodoma, Dodoma, Tanzania

Doris Robert Mhohelo
Department of Statistics, University of Dodoma, Dodoma, Tanzania

Corresponding Author:
KK Saxena
Department of Statistics, University of Dodoma, Dodoma, Tanzania

Modelling and forecasting retail prices of maize for three agricultural markets in Tanzania

KK Saxena and Doris Robert Mhohelo

Abstract
This study examined the modelling of maize prices using Autoregressive Integrated Moving Average (ARIMA) technique to determine the most efficient and adequate model for analyzing the maize monthly prices at the Gairo market in Morogoro Region, Manyoni market in Singida Region and Kibaigwa market in Dodoma Region. The results indicate that ARIMA (1, 1, 4) model is the most adequate and efficient model for Gairo market, ARIMA (2, 1, 3) model is the most adequate and efficient mode1 for Manyoni market and ARIMA (2, 2, 3) model is the most adequate and efficient model for Kibaigwa market. This was determined by comparing the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) and Mean Absolute Percentage Error (MAPE). Time-series analysis was done using STATGRAPHICS, EXCEL, R software and SAS JPM. The forecast results suggest that there are expectations of increasing maize prices in Manyoni market from June-2018 to May-2019, the maize prices in Kibaigwa market are also expected to increase with time from January 2016 to December 2016 and the maize prices at Gairo market are expected to keep on increasing with time from June 2018 to May 2019. The results will make better understanding of maize prices situation and future prices will enable producers and consumers to make the right choices concerning buying and selling arrangements of Maize crop in Tanzania.

Keywords: ARIMA Model, Box-Jenkins Methodology, Maiza Price, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), MAPE

Introduction
Review of Literature
Empirical lliterature review outside Tanzania
Several investigators have discussed the use of Univariate time series in modelling and forecasting of various agronomic food crop prices worldwide. The ARIMA technique have been used extensively by a number of researchers to fit model and forecast prices, demands in terms of internal consumption, imports and exports to adopt appropriate solutions. These approaches were employed extensively for forecasting economic time series, inventory and sales modelling also [Ljung and Box (1978) [9], Pindyck and Rubinfeld (1981) [15] and Sohail *et al.* (1994)] [18]. Contreras *et al.* (2003) [4] conducted a study on ARIMA models to predict next-day electricity prices. Rangsan and Nochai (2006) [17] studied oil palm price of Thailand in three categories as farmstead price, general price and wholesome oil palm price by using ARIMA models. Rachana *et al.* (2010) used ARIMA models to forecast pigeon pea production in India. Badmus and Ariyo (2011) [2] forecasted the area of cultivation and production of maize in Nigeria using ARIMA model. Some more studies on modelling and forecasting by ARIMA were conducted by Adejumo and Momo (2013) [1], Pierre *et al.* (2014) [14], and Gertler *et al.* (2016) [6]. Kirimi, (2016) [8] conducted study on modelling the volatility of maize prices using ARIMA models so as to achieve the utmost effective and satisfactory model for investigating the unpredictability of prices of maize in Kenya.. Manoj and Anand (2017) [10] studied the application of time series ARIMA forecasting model for predicting sugarcane production in India and found the best ARIMA model as ARIMA (2,1,0). Darekar & Reddy (2017) [5] studied the prediction of paddy prices for Kharif 2017-2018. Venkatesh *et al.* (2017) [20] studied on the Maize price forecasting by using ARIMA. Nyangarika *et al.* (2019) [13] has done a study on oil prices factors for forecasting by ARIMA model and forecasts using exponential smoothing.

Pradesh *et al.* (2019) studied on the estimation of weekly green gram prices for the Odisha state in India to evaluate the performance by comparing it with ARIMA models with respect to MAPE criteria.

Empirical lliterature review in Tanzania

The studies by Nkonya *et al.* (1998) [11] and Nathaniel *et al.* (1998) [12] were based on the adoption of maize production technologies in northern and south Tanzania, which formed part of a larger study to assess the influence of maize research and extension throughout Tanzania over the past 20 years. The results revealed that the formal credit system needed to be changed to address the credit problems faced by small-scale farmers. A more efficient marketing system for inputs and outputs would benefit farmers by providing higher maize prices and reducing fertilizer costs. Such a system would need supporting policies from the government. The results also revealed that extension should be strengthened to increase the adoption of fertilizer, and farmers should receive more advice about using organic manure to supplement chemical fertilizers. Extension efforts should also be made towards promoting the adoption of improved varieties, weeding, and management practices for controlling diseases and field and storage pests. Nkonya *et al.* (1998) [11] recommended the development of additional hybrids for the Northern Zone and/or village level production of composite seed, improved varieties (including both composites and hybrids), and more research and extension effort directed toward efficient use of fertilizers (manure, chemical fertilizer, and crop residues). The study suggested encouraging measures by banks and policy makers to give more credit to small maize farmers. A study on supply response of maize in Tanzania conducted by Waryoba (2015) [21] used Error Correction Mechanism (ECM) approach due to its ability to calculate both short and long run elasticity of agricultural supply. It was recommended that, the government should indirectly intervene in the market to promote efficiency in price mechanism, make effective use of grain reserve to ensure market for maize output even in periods of bumper harvest so that farmers can be able to buy fertilizers other farm inputs to improve food crop production. Kibona & Mbago, (2017) [7] sought to estimate general maize prices in Tanzania using ARIMA model for the maize data from 2004 to 2017 obtained from the Bank of Tanzania. The study found that ARIMA (3,1,1) as the effective model for predicting maize general prices based on minimum Akaike's Information Criterion (AIC) and the fitted model was brought suitably into being using Ljung-Box test.

Urassa, (2017) conducted a study on factors influencing maize crop production at households levels for the case of Rukwa region Tanzania. The study found that maize crop continues to play an important role in most households' livelihood. Some of the studies did predictions on wholesale of maize prices in Tanzania but no study was done to fit a model for retail prices of maize in various markets of different regions of Tanzania. Hence this study was undertaken to fit model and forecast the retail prices of Maize of various markets for different regions in Tanzania, which is important for traders to decide where to purchase/sell maize at reasonable prices. Therefore, this study was conducted to fill the gaps, which in turn, would be helpful to the government, producers, and consumers.

Research Methodology

Study area

This study was conducted in three agricultural markets which are Gairo agricultural market at Morogoro, Manyoni agricultural market at Singida and Kibaigwa agricultural market at Dodoma in Tanzania. These markets were purposively selected because of the convenience of availability of data.

Time series components analysis

The components of the time series under this study were examined by making time series plots and then seasonal decomposition. Time series plots can detect whether the data are stationary or non-stationary and the plots can help the researcher to recognize the form of essential pattern of the specified data well-arranged over time.

The steps in ARIMA model building

A. Plotting of time series data

This is done through initial plotting of the historical data and observing its graph whether it is stationary or non-stationary. Plotting the graphs of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the historical data helps in identifying if the data existing are stationary or non-stationary.

B. Test for time series stationarity

This test decides whether the series is fixed or not by seeing the ACF graphs. If the ACF graph of past standards either cuts off equally rapidly or passes on down equally rapidly, then the past standards should be used as fixed. If the ACF graph pass on down unhurriedly, then the past values should be used as non-fixed. If the series is not fixed, it can be transformed to a fixed series by differencing. That is, the innovative series is substituted by a series of differences. An ARMA model is formerly stated for the differencing series. Differencing is completed till a plot of the data displays the series fluctuating approximately to a static level, and the graph of ACF is either cut off equally or passes on down equally rapidly.

C. The Augmented Dickey-Fuller Test (ADF)

Stationarity test of a differenced time series develops the Augmented Dickey-Fuller (ADF) technique (Dickey and Fuller (1981), which is a comprehensive auto-regression model. The hypotheses are formulated as H_0: Non-stationary and H_1 Stationarity. ADF Statistics is tested on the basis of critical values to make decision about stationarity.

D. Identification of model by ACF and PACF

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Autocorrelation is calculated as a simple correlation between existing observations (y_t) and the former remark from p periods to the existing period (y_{t-p}). Partial autocorrelations are used to obtain the relationship among (y_t) and (y_{t-p}) when the special effects of additional time lag $1, 2, 3, \ldots, p-1$ are eradicated. Partial autocorrelation function (PACF) and autocorrelation function (ACF) can show whether the series is non-stationary by suggesting the type of the model whether it is AR, MA, and ARMA after checking the cutoff of the lag.

Model for non-seasonal series are known as Autoregressive Integrated Moving Average model signified by ARIMA (p, d, q). At this point, p specifies the order of autoregressive part, d identifies the amount of differencing and q means the order of the moving average part. If the innovative series is fixed d=0 then ARIMA models diminish to ARMA models. The change linear operator (Δ) is demarcated by $\Delta y_t = y_t - y_{t-1} = y_t - \beta y_t = (1-\beta) y_t$. A fixed series ($w_t$) attained as the dth difference (Δ^d) of (y_t) where $w_t = \Delta^d y_t = (1-\beta)^d y_t$. ARIMA (p, d, q) has the common form of $\phi_p(\beta)(1-\beta)^d y_t = \mu + \theta(\beta)\varepsilon_t$ or $\phi_p(\beta)w_t = \mu + \theta(\beta)\varepsilon_t$ .When the fixed series has been achieved, then classify the formula of the model to be used by means of the graph of autocorrelation function (ACF) and the sample partial autocorrelation function (PACF).

Model Estimation

The procedures are unified by statistical software used for the estimation of the parameters in the model. After reviewing the ACF and PACF, the stationary ARMA, ARIMA and SARIMA can be predicted. In this study, the researcher applied the maximum likelihood technique to estimate the parameters by using the MINTAB and SAS software.

Model checking

The model adequacy test is used before using forecasting step and the Ljung –Box test can be applied to the residuals. A model is said to be adequate if the residuals left over after the model fit seem to be white noise. This means that the residuals should not be correlated with constant variance. The pattern of ACF and the PACF are used to detect misspecifications, which lead to the identification of a different model. The best model can be obtained by considering the following diagnostics:

Akaike Information Criteria (AIC)

Akaike's Information Criterion (AIC) is a way of choosing the best fit model from a set of appropriate models. The (AIC) deals with a relative estimation of the data missing when a particular model is used to signify the procedure that produces the data. The selected model is the one that diminishes the Kullback - Leibler distance concerning the model and the truth. It is built on information theory and it is a measure that searches for a model, which has a good fit but contains a small number of parameters. The finest model has the smallest AIC value.

Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) projected by Schwarz (1978) is an alternative measure, which tries to precise the AIC's propensity to over-fit. This measure is specified as weakness for the added parameters. As per all measures, the chosen model is the one which has a least BIC.

a) ACF and PACF plots of residuals

As soon as the proper ARIMA model is fitted, the goodness of fit is ascertained by means of plotting the ACF and PACF plots of residuals of the fitted model. If the best sample autocorrelation coefficients of the residuals are inside the limits $\pm 1.96/N$ where N is the number of observations upon which the model is built, then the residuals are white noise specifying that the model is a good fit.

b) Analysis of autocorrelations of residuals by Box-Pierce or Ljung-Box test

The Ljung – Box test, is useful to the residuals once an ARIMA model has been fixed to check for the uncertainty in the residuals. The Ljung – Box test is built on the autocorrelation plot. However, as a substitute of checking for uncertainty to each distinct lags, it checks the "general" uncertainty based on number of lags.

Forecasting

Forecasting speaks of the estimation of upcoming values of a variable from the historical and present values of that variable or other interrelated variables. This includes the use of the fitted model for estimating the upcoming standards, which can be short-term, medium term and long term predicting. The estimated values are described with confidence intervals listed with the level of significance for out of sample predictions.In prediction, the targeted point is to predict the upcoming values of historical details y_{n+m} , where n and m are both constant $n = m = 1, 2, 3, \dots$ based on the data given. Therefore, if the model satisfies all the diagnostic checks, it is well-thought-out for predictions.

Validity

According to Gachengo, (2015) validity helps to increase accuracy, which is useful to the findings by removing as many confounding variables as possible. Finally, analysis was made using Statgraphics, Excel, R software and SAS JPM to ensure both reliability and validity of the outcomes.

Results and Discussion

Model Fitting on the Maize Prices for Gairo Market in Morogoro

Model Identification for Gairo Market

In modelling, the maize prices time series, the data set of the last twelve months (June 2018 to May 2019) were used for comparing forecast and the modelling was done using the monthly maize prices data from January 2009 to May 2018.

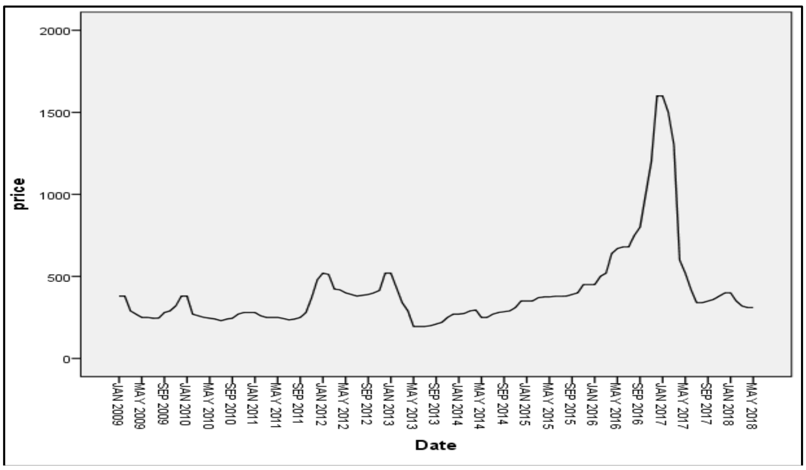

Fig 4.1.1: A time series plot of monthly maize prices data (from Jan 2009 to May 2018) at Gairo Market

Figure 4.1.1 above specifies a time series plot of the monthly maize prices data of the Morogoro region specifically at Gairo market from January 2009 to May 2018. After the time series plot in maize prices, it is obviously detected that there were variations in prices with increase in time which in turn indicated non-stationary of the series in variance. The best idea is to start with differencing with the lowest order (d=1) and test the data for unit root problems. So, we obtained a time series of first order differencing and figure 4.1.2 below is the plot of the first order differenced maize data.

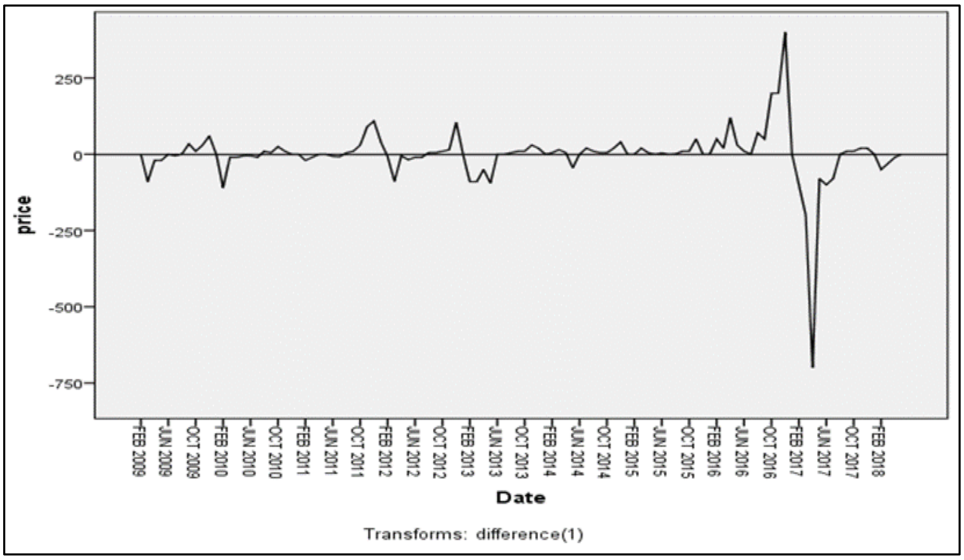

Fig 4.1.2: Plot of the first differenced maize prices data (d=1) at Gairo Market

From figure 4.1.2 above, it is concluded that the time series appears to be stationary in both mean and variance. But before moving into another step, Augmented Dickey-Fuller test was applied to the differenced time series data also for testing stationary.

Test for stationary ADF Test

Using ADF test, there is null hypothesis (H_o) and alternative hypothesis (H_1) where by H_o represents non-stationary time series data while H_1 represents the stationary time series data. The hypothesis is tested by carrying out appropriate differencing of the data in the dth order and applying the ADF test to the differenced time series data. The first order differencing of the data (d=1) means the table of differenced data of current and previous one $(X_t = X_t - X_{t-1})$ is created. The ADF test results are presented in Table 4.1.1.

Table 4.1.1: Test Results

Lags	P-Value	Significance level
0	0.001	<0.05
5	0.001	<0.05
10	0.0019363	<0.05
15	0.0011464	<0.05

For maize prices at Gairo market, the ADF test statistic in Table 4.1.1 is less than 0.05 p-value for lag order 0 to 15 showing that the series is stationary. Therefore, it is not necessary to add another differencing and ARIMA (p, d, q) is adopted where d=1.

This step is essential since it is helpful in further steps for ARIMA (p, 1, q) that is, in finding suitable p in AR and q in MA process. Now, the next step is to examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Figure 4.1.3 below represents the plot of ACF for lag 1 to 20 of the first order differenced time series of the maize prices at Gairo market.

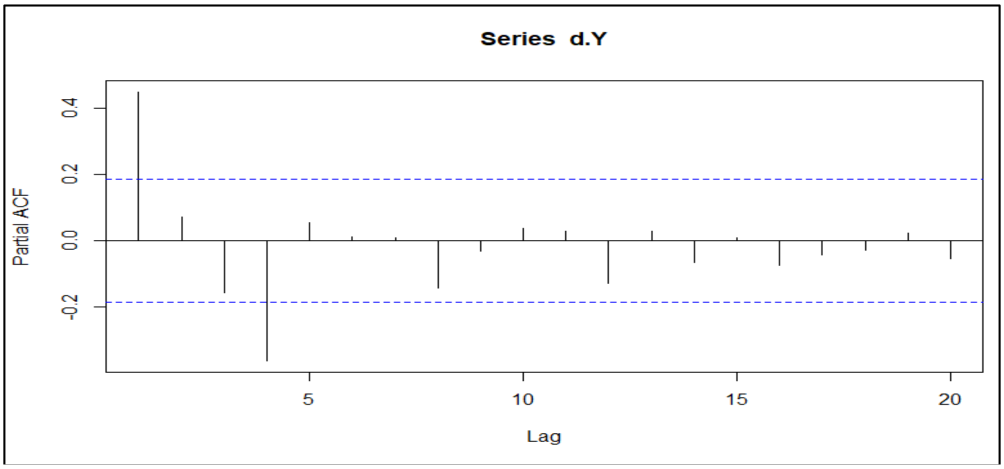

Fig 4.1.3: Estimated ACF of Maize Prices Data at Gairo Market

The above ACF infers that the autocorrelation at lag 1 and lag 2 exceed the significance limits and autocorrelation tails off to zero after lag 6. The autocorrelation at lag 3, lag 7 up to lag 20 does not exceed the significant limits. Therefore, it is concluded that those outside the significant limits are assumed as errors that happened by chance.

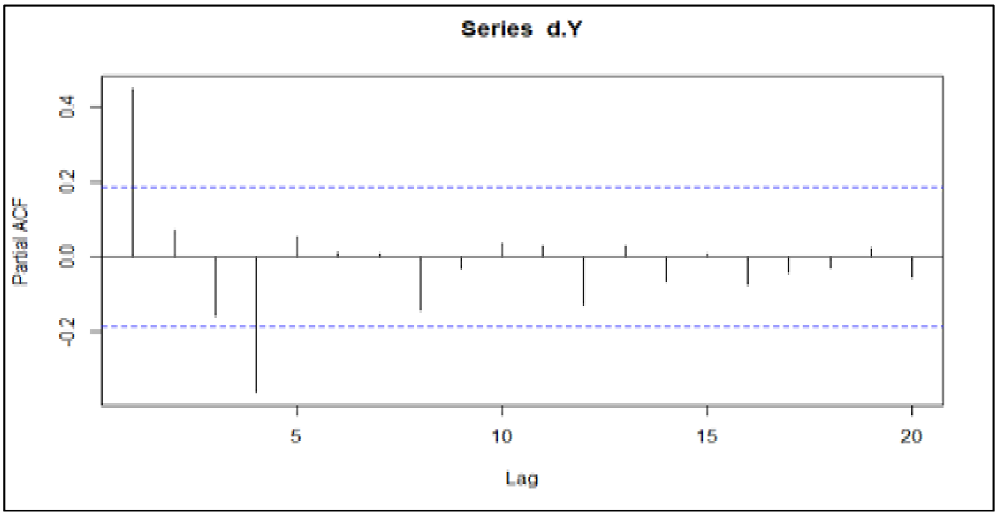

Fig 4.1.4: Estimated PACF of the Maize Prices Data at Gairo Market

Figure 4.1.4 above represents the partial autocorrelation function (PACF) for the first order differenced time series from lag 1 to lag 20. It concludes that, PACF tails off to zero after lag 3,which can be assumed as an error that happened by chance because all PACFs from lag 4 to 20 are within the significant limits. The ACF tailing off to zero after lag 6 and the PACF tailing off to zero after lag 3 can define the following possible ARIMA models for the first differenced time series data of maize prices at Gairo market.

Suggested Models for the First Differenced Maize Prices Data

1. ARIMA (1,1,4)
2. ARIMA (2,1,4)
3. ARIMA (4,1,0)
4. ARIMA (3,1,4)

Model Selection Criteria

To select the best model from the four models above, ARIMA model with the lowest Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values were selected. The following table 4.1.2 summarizes the output of each fitted ARIMA model in the time series maize prices data at Gairo market.

Table 4.1.1: AIC and BIC values of the fitted model ARIMA (1, 1, 4)

MODEL	RMSE	MAPE	MAE	BIC	AIC
ARIMA (1,1,4)	**73.9632**	**9.1343**	**39.9286**	**8.81631**	**8.69563**
ARIMA (2,1,4)	74.2918	9.2109	40.1760	8.86701	8.7222
ARIMA (4,1,0)	75.9269	9.59546	42.0841	8.82688	8.73034
ARIMA (3,1,4)	76.3969	8.87019	39.3151	8.83923	8.74268

Based on AIC and BIC values, the best model was found to be ARIMA (1, 1, 4). Hence, this is the best model for predicting future values of maize prices at Gairo market.

Table 4.1.2: Estimation Summary for the ARIMA (1, 1, 4) Model

Models	Coefficient	Standard error
AR (1) $\hat{\varphi}$	0.7651	0.0891
MA (1) $\hat{\theta}_1$	-0.4097	0.1106
MA (2) $\hat{\theta}_2$	0.0133	0.0998
MA (3) $\hat{\theta}_3$	-0.0479	0.1023
MA (4) $\hat{\theta}_4$	-0.5051	0.0995

From Table 4.1.1 above, we fit the maize prices ARIMA time series model using the multiplicative form of

$$(1-\phi_1 B)(1-B)X_t = (1-\theta_1 B-\theta_2 B^2-\theta_3 B^3-\theta_4 B^4)\varepsilon_t,$$

where the estimates of parameters are found as $\hat{\varphi} = 0.7651, \hat{\theta}_1 = -0.4097, \hat{\theta}_2 = 0.0133, \hat{\theta}_3 = -0.0479$ *and* $\hat{\theta}_4 = -0.5051$. Hence the estimated equation for ARIMA (1, 1, 4) is obtained as

$$(1-0.7651B)(1-B)X_t = (1+0.4097B-0.0133B^2+0.0479B^3+0.5051B^4)\varepsilon_t$$

From this study, the fitted model for maize prices time series data at Gairo market is ARIMA (1, 1, 4). It consists of both AR and MA processes and it is free from seasonal component, which means that seasonality in maize prices is non-significant.

The model adequacy is further tested to draw empirical conclusions regarding the model as good fit for forecasting time series. Ljung-box test was done in addition to ACF and PACF residuals plots.

Diagnostic Checking

The diagnostics of the residuals by ACF values in figure 4.1.5 show that the ACF values are all within the 95% confidence limit indicating that there is no correlation among residuals.

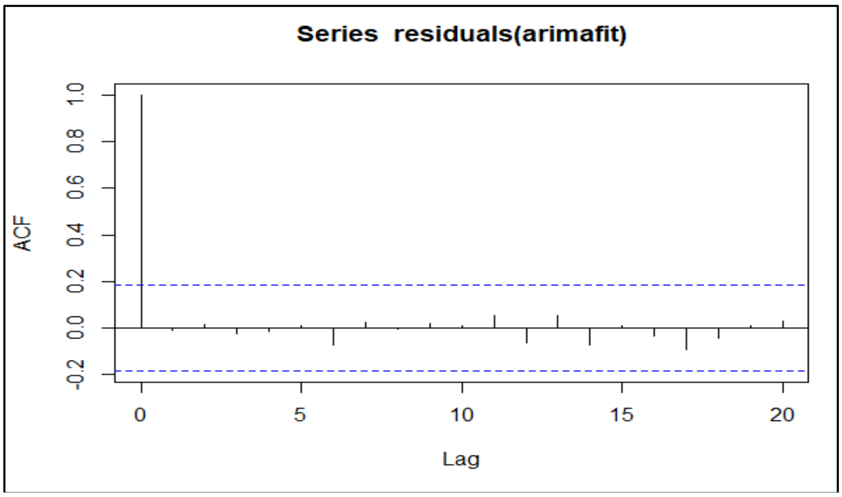

Fig 4.1.5: ACF Residuals for Maize Prices at Gairo Market

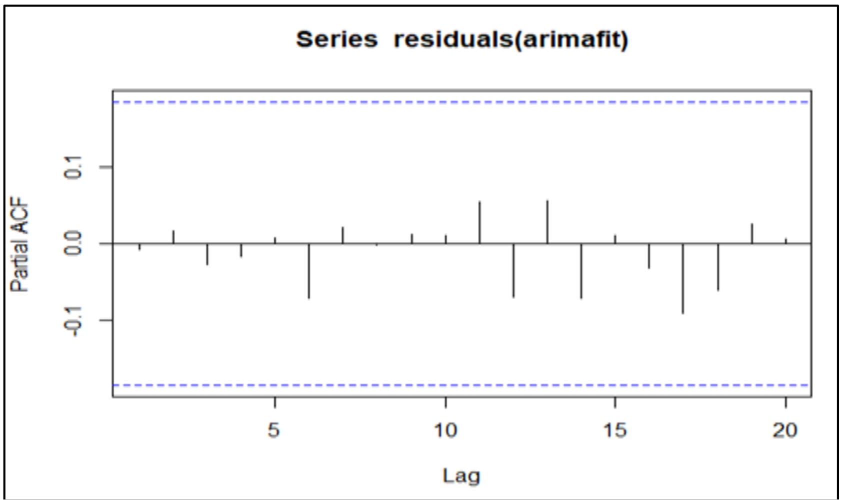

Fig 4.1.6: PACF Residuals for Maize prices at Gairo Market

The diagnostics of the residuals by PACF values in figure 4.1.6 also show that the PACF values are all within the 95% confidence limit indicating that there is no correlation among residuals. The plot of the fitted ARIMA model (1,1,4) is presented in Figure 4.1.7.

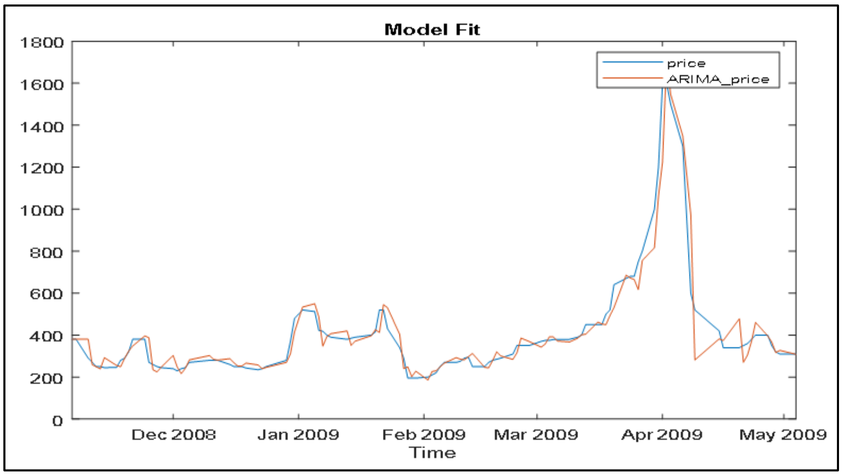

Fig 4.1.7: The Plot of fitted ARIMA (1,1,4) Model

Forecasting with the Fitted Model at Gairo Market

After diagnostic checking, the model fit can be used to predict the upcoming values of the variable of interest if it is adequate. But, before forecasting, we need to measure the accuracy of its predictions and it is completed by one-step-ahead of forecasting. As a result, the fitted model ARIMA (1, 1, 4) was used to predict the maize prices for twelve months (June 2018-May 2019). The observed and predicted values are shown in Table 4.1.4.

Table 4.1.3: One-step-ahead Forecast of Maize prices at Gairo Market

Period	Observed value	Forecast	forecast error	absolute error	squared error	Absolute (%) error
Jun-18	300	318.02	-18.02	18.02	324.7204	-6.007
Jul-18	300	322.53	-22.53	22.53	507.6009	-7.51
Aug-18	460	336.16	123.84	123.84	15,336.35	26.922
Sep-18	480	345.312	134.688	134.688	18,140.86	28.056
Oct-18	530	352.344	177.656	177.656	31561.65434	33.52
Nov-18	630	357.746	272.254	272.254	74122.24052	43.215
Dec-18	720	361.896	358.104	358.104	128238.4748	49.737
Jan-19	720	365.085	354.915	354.915	125964.6572	49.294
Feb-19	630	367.535	262.465	262.465	68887.87622	41.661
Mar-19	621	369.417	251.583	251.583	62294.00589	40.513
Apr-19	611	370.863	240.137	240.137	57665.77877	39.302
May-19	611	371.974	239.026	239.026	57133.42868	39.12
Total	6613	4238.882	2374.118	2455.218	640177.6407	377.823

Forecasting Accuracy of the Fitted Model

To assess the model predicting capability, the standard measures of forecast accuracy were obtained as shown in Table 4.1.5. The values of these measures were obtained using the formulae stated in methodological part.

Table 4.1.4: Measures of Forecasting Accuracy

Variables	Maize Prices
Mean Squared error (MSE)	53348.14
Mean Absolute error (MAE)	204.60
Mean percentage error (MPE)	31.49
Mean absolute percentage error (MAPE)	33.74

Analysis of Forecasting Errors

The analysis of forecasting errors is very important because this aspect is used in evaluating the accuracy of future forecasts of the fitted model. To assess the model forecasting capability, we consider the standard measures of forecast accuracy in Table 4.5. These measures of forecasting accuracy are evaluated as part of validation of the fitted models. The mean forecast error values for maize prices are not close to zero indicating that the forecasts produced by the fit ARIMA (1,1,4) are not unbiased. Both MAE and MSE indicate that there is variability in forecasting errors with the fitted model for maize prices. The relative (or percent) forecast error shows MAPE value for maize prices is 33.74%.

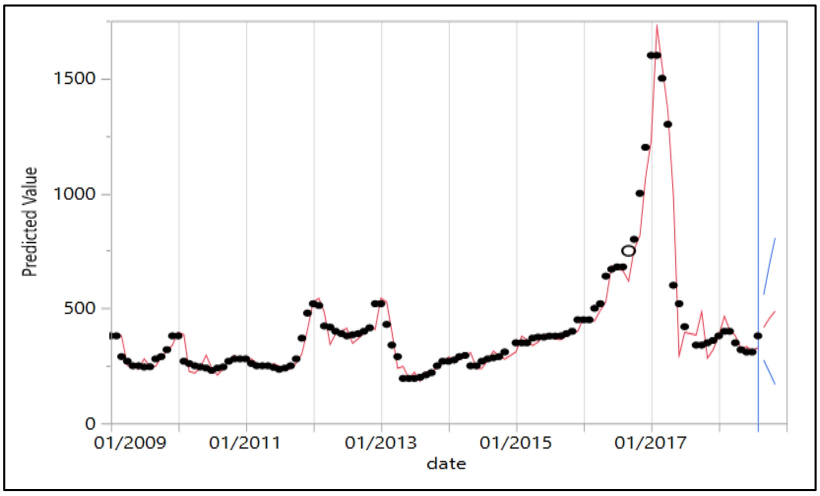

Fig 4.1.3: Time Plot for forecast Maize prices at Gairo Market for ARIMA (1, 1, 4)

In figure 4.1.8 above, the two blue lines of forecast represent the 95% (lower and upper side) projection of the forecasting intervals. The maize prices were expected to keep on increasing with time from June 2018 to May 2019.

Model Fitting on the Maize Prices for Manyoni Market in SingidaModel Identification for Manyoni Market

In modelling, the maize prices time series, the data set of the last twelve months (June 2018 to May 2019) were used for comparing forecast and the modelling was done using the monthly maize prices data from January 2009 to May 2018.

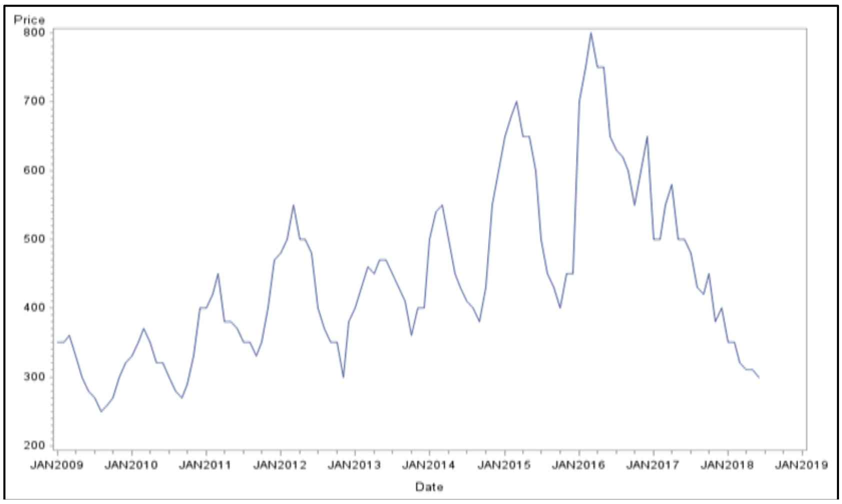

Fig 4.2.1: A Time Series Plot of Monthly Maize Prices Data (from Jan 2009 to May 2018) at Manyoni Market

Figure 4.2.1 above specifies a time series plot of the monthly maize prices data of the Singida region specifically at Manyoni Market from January 2009 to May 2018. After the time series plot in maize prices, it is obviously noticed that there were variations in prices with increase in time which in turn specified the non-stationary of the series in variance. The best idea is to start with differencing with the lowest order (d=1) and test the data for unit root problems. So, we obtained a time series of first order differencing and figure 4.2.2 below is the plot of the first order differenced maize data.

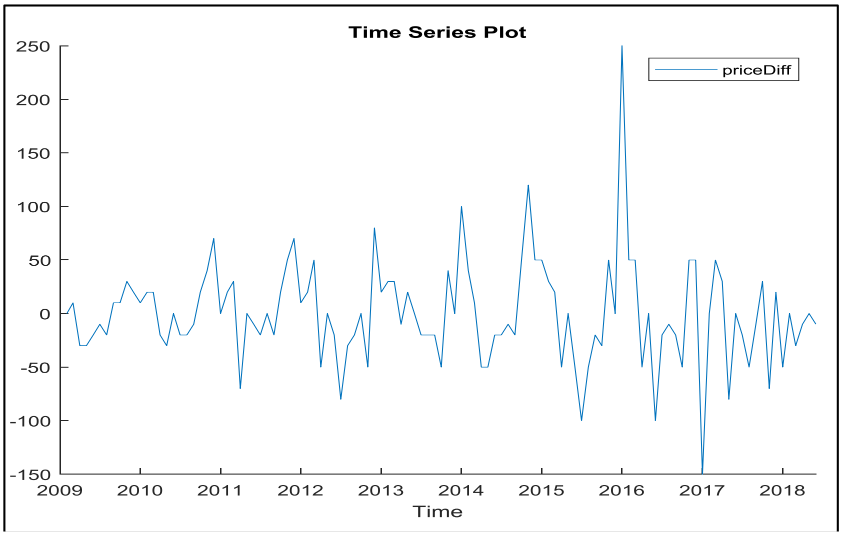

Fig 4.2.2: Plot of the First Differenced Maize Prices Data (d=1) at Manyoni Market

From the plot above, it is concluded that the time series appears to be stationary in both mean and variance. However, before moving to another step, Augmented Dickey-Fuller test was applied to the differenced time series data also for testing stationary.

Test for Stationary ADF Test

Using ADF test, there is null hypothesis (H_o) and alternative hypothesis (H_1) where by H_o represents non-stationary time series data while H_1 represents the stationary time series data. The hypothesis is then tested by carrying out appropriate differencing of the data in the d[th] order and applying the ADF test to the differenced time series data. The first order differencing of the data (d=1) means that the table of differenced data of current and previous one $(X_t = X_t - X_{t-1})$ is created. The ADF test result is shown in Table 4.2.1 below

Table 4.2.1: Test Results

Lags	P-Value	Significance level
0	0.001	<0.05
5	0.001	<0.05
10	0.001	<0.05

For maize prices at Manyoni District, the ADF test statistic in Table 4.2.1 above is less than 0.05 p-value for lag order 0 to 10 showing that the series is stationary, therefore H_o is rejected and the conclusion is that the alternative hypothesis is true which means, the series is stationary in both mean and variance. Therefore, it is not necessary to add another differencing and ARIMA (p, d, q) is adopted where d=1. This step is essential since it is helpful in further steps for ARIMA (p, 1, q) in finding suitable p in AR and q in MA process.

Now, the next step is to examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Figure 4.2.3 below represents the plot of ACF for lag 1 to 20 of the first order differenced time series of the maize prices at Manyoni market.

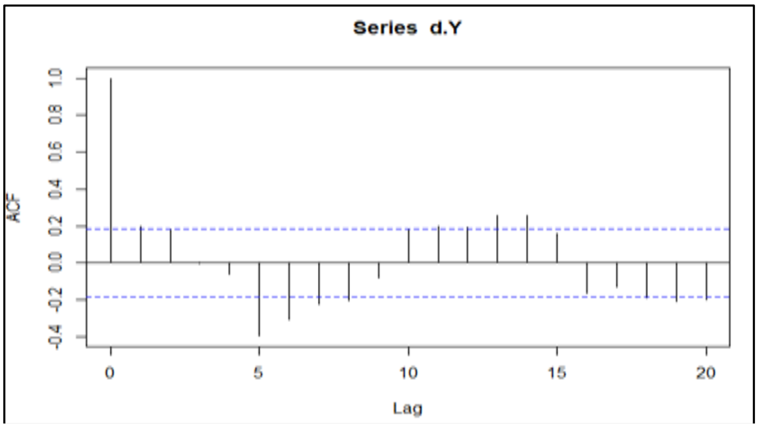

Fig 4.2.3a: Estimated ACF of Maize Prices Data in Manyoni District

The above ACF concludes that the autocorrelation at lag 5 exceeds the significance limits and autocorrelation tails off to zero after lag 8. The autocorrelation at lag 13 and lag 14 shows that there is a very low seasonal characteristic that cannot be detected by SPSS. Therefore, it is concluded that those outside the significant limits can be assumed as errors that happened by chance.

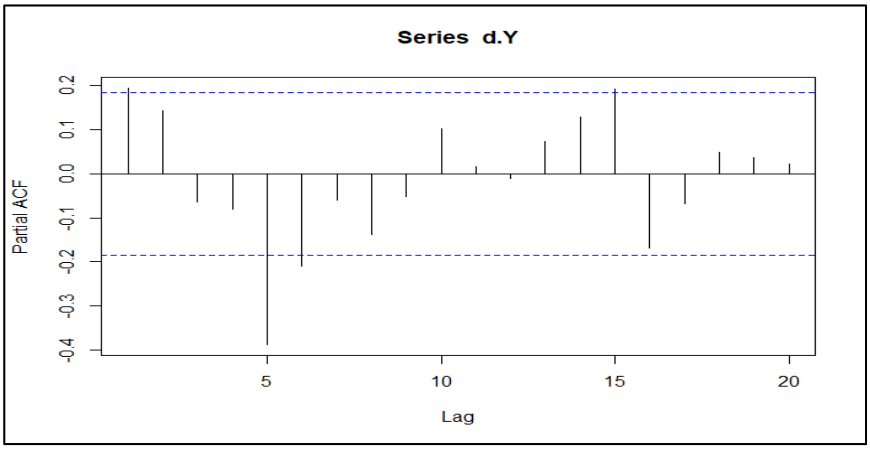

Fig 4.2.4: Estimated PACF of the Maize Prices Data at Manyoni Market

Figure 4.2.4 above represents the partial autocorrelation function (PACF) for the first order differenced time series from lag 1 to lag 20. It concludes that PACF exceed significant limits at lag 1, lag 5 and lag 6, then the PACF of all other lags tails off to zero although there is an outlier at lag 15 which can be assumed as an error happened by chance. The ACF and PACF tailing off to zero can define the following possible ARIMA models for the first differenced time series data of maize prices at Manyoni market.

Suggested Models for the First Differenced Maize Prices Data

1. ARIMA (2,1,2)
2. ARIMA (2,1,3)

Model Selection Criteria

To select the best model from the two models above, ARIMA model with the lowest Akaike Information Criterion (AIC) values were selected. The following table 4.2.2 summarizes the output of each fitted ARIMA model in the time series maize prices data at Manyoni market.

Table 4.2.2: AIC and BIC Values of the Fitted Model ARIMA (2, 1, 3)

Model	AIC	BIC
ARIMA (2,1,2) Model	1188.0578	1201.6054
ARIMA (2,1,3) Model	1186.5837	1202.8409

Based on AIC values, the best model was found to be ARIMA (2, 1, 3). Hence, this is the best model for predicting future values of maize prices at Manyoni market.

Table 4.2.3: Estimation Summary for the ARIMA (2, 1, 3) Model

Models	Coefficient	Standard error
AR (1) $\hat{\Phi}_1$	1.6856	0.0463
AR (2) $\hat{\Phi}_2$	-0.9328	0.0545
MA (1) $\hat{\theta}_1$	-1.7044	0.1056
MA (2) $\hat{\theta}_2$	1.0651	0.0076
MA (3) $\hat{\theta}_3$	-0.0973	0.0973
MA (4) $\hat{\theta}_4$	-0.5051	0.0995

From table 4.2.3 above, we fit the maize prices ARIMA time series model using the multiplicative form of $(1-\phi_1 B-\phi_2 B^2)(1-B)X_t=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)\varepsilon_t$, where the estimates of parameters are found as $\hat{\varphi}_1 = 1.6856$, $\hat{\varphi}_2$=-0.9328, $\hat{\theta}_1 = -1.7044, \hat{\theta}_2 = 1.0651, \hat{\theta}_3 = -0.0973, and\ \hat{\theta}_4$=-0.5051.

Hence the estimated equation for ARIMA (2, 1, 3) is obtained as $(1-1.6856B+0.9328B^2)(1-B)X_t=(1+1.7044B-1.0651B^2+0.0973B^3)\varepsilon_t$

From this study, the fitted model for maize prices time series data at Manyoni Market is ARIMA (2, 1, 3). It consists of both AR and MA processes and it is free from seasonal component, which means that seasonality in maize prices is non-significant. The model adequacy is further tested to draw empirical conclusions regarding the model as good fit for forecasting time series. Ljung-box test was used in addition to ACF and PACF residuals plots.

Diagnostic Checking

The diagnostics of the residuals by ACF values in figure 4.2.5 show that the ACF values are all within the 95% confidence limit indicating that there is no correlation among residuals.

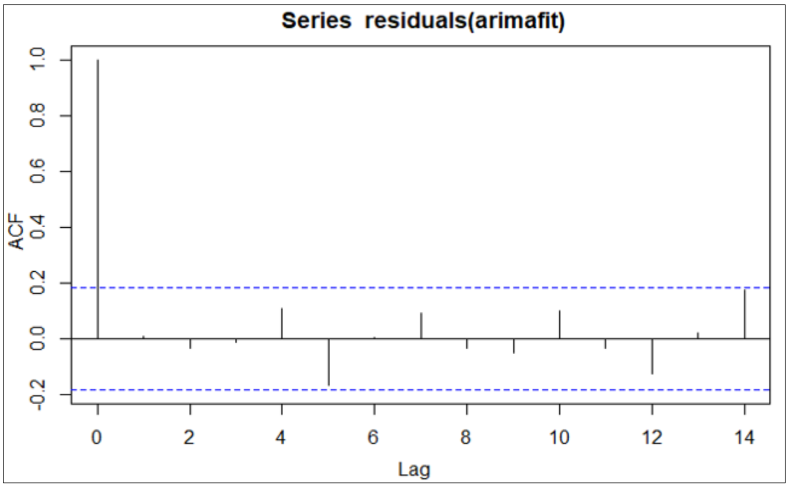

Fig 4.2.5: ACF Residuals for Maize Prices at Manyoni Market

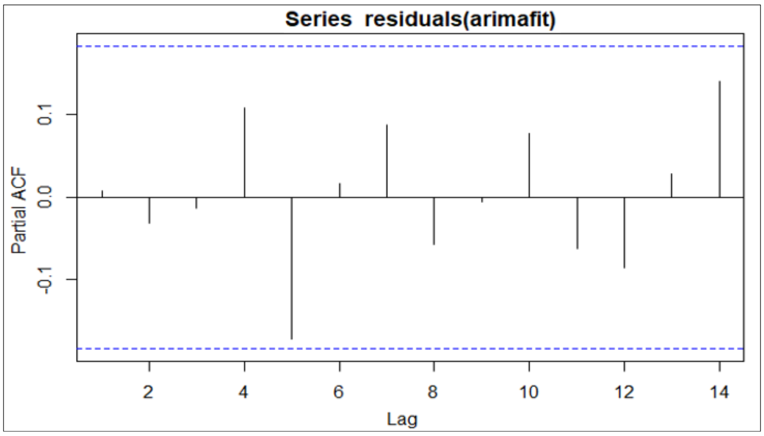

Fig 4.2.6: PACF Residuals for Maize Prices at Manyoni Market

The diagnostics of the residuals by PACF values in figure 4.2.6 show that the PACF values are all within the 95% confidence limit indicating that there is no correlation among residuals. The plot of the fitted ARIMA model (2,1,3) is presented in Figure 4.2.7.

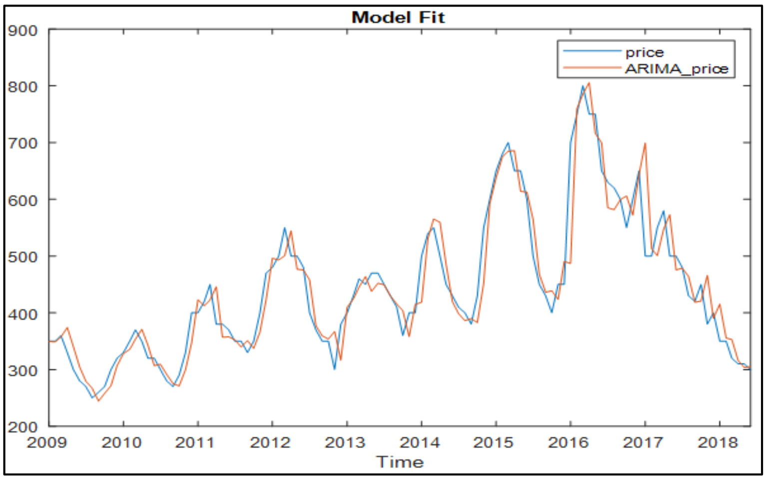

Fig 4.2.4: The Plot of Fitted ARIMA (2, 1, 3) Model

Forecasting with the Fitted Model at Manyoni Market

After diagnostic checking, the model fit is used to predict the upcoming values of the variable of interest if it is adequate. But, before forecasting we need to measure the accuracy of its predictions and it is completed by one-step-ahead forecasting. As a result, the fitted model ARIMA (2, 1, 3) was used to predict the maize prices for twelve months (June 2018-May 2019). The observed and predicted values are shown in Table 4.2.4.

Table 4.2.4: One-Step-ahead Forecast of Maize Prices at Manyoni Market

Period	Observed	Forecast	Forecast error	Absolute error	Squared error	Absolute % error
Jul-18	300	306.8857	-6.88574	6.88574	47.41342	2.295247
Aug-18	350	312.6242	37.37576	37.37576	1396.947	10.67879
Sep-18	350	318.5782	31.42177	31.42177	987.3276	8.977648
Oct-18	370	323.262	46.73803	46.73803	2184.444	12.6319
Nov-18	400	325.6035	74.39649	74.39649	5534.838	18.59912
Dec-18	450	325.1817	124.8183	124.8183	15579.6	27.73739
Jan-19	400	322.2867	77.7133	77.7133	6039.357	19.42833
Feb-19	400	317.8001	82.19993	82.19993	6756.828	20.54998
Mar-19	500	312.9375	187.0625	187.0625	34992.36	37.41249
Apr-19	600	308.9259	291.0741	291.0741	84724.12	48.51235
May-19	600	306.6993	293.3007	293.3007	86025.31	48.88345
Jun-19	500	306.6878	193.3122	193.3122	37369.6	38.66244
Total	5220	3787.473	1432.527	1446.299	281638.1	294.3691

Forecasting Accuracy of the Fitted Model

To assess the model predicting capability, the standard measures of forecast accuracy were obtained as shown in Table 4.2.5. The values of these measures were obtained using the formulae stated in methodological part and the forecast errors existing in Table 4.2.5.

Table 4.2.5: Measures of Forecasting Accuracy

Variables	Maize prices
Mean Squared error (MSE)	23469.85
Mean Absolute error (MAE)	120.52
Mean percentage error (MPE)	24.15
Mean absolute percentage error (MAPE)	24.53

Analysis of Forecasting Errors

The analysis of forecast errors is very important because this aspect is used in evaluating the accuracy of future forecasts of the fitted model. To assess the model forecasting capability, we consider the standard measures of forecast accuracy in Table 4.2.5. These measures of forecast accuracy are evaluated as part of validation of the fitted models. The mean forecast error values for maize prices are not close to zero indicating that the forecasts produced by the fit ARIMA (2,1,3) are not unbiased. Both MAE and MSE indicate that there is variability in forecasting errors with the fitted model for maize prices. The relative (or percent) forecast error shows that MAPE value for maize prices is 24.53%.

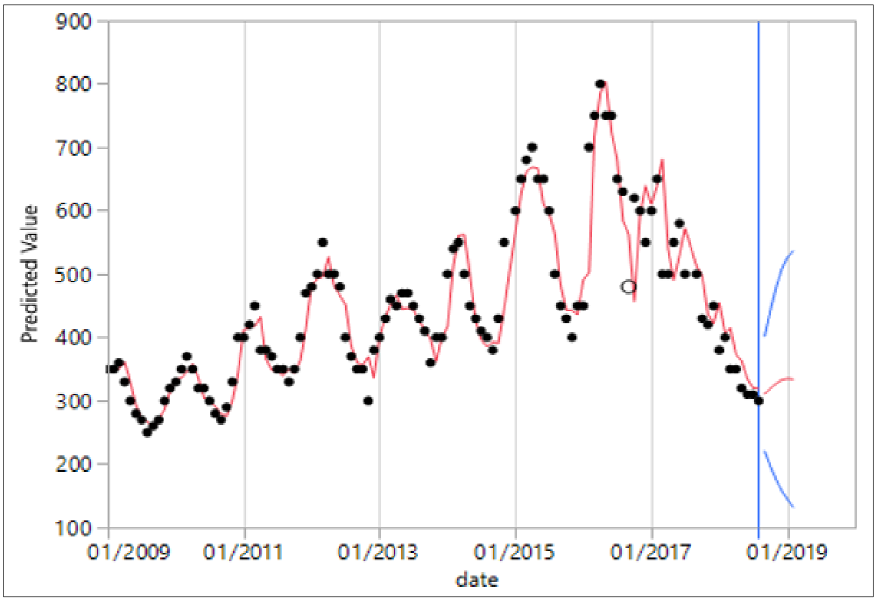

Fig 4.2.8: Time Plot for forecast Maize Prices at Manyoni Market for ARIMA (2, 1, 3)

In 4.2.8 figure above, the two blue lines of forecast represent 95% (lower and upper side) projection of the forecasting intervals. The maize prices were expected to increase with time from June 2018 to March 2019.

Model Fitting on the Maize Prices for Kibaigwa Market in Dodoma Model Identification for Kibaigwa Market

In modelling the maize prices time series, the data set of the last twelve months (Jan 2016 to December 2016) were used for comparing forecast and the modelling was done using the monthly maize prices data from January 2005 to December 2015.

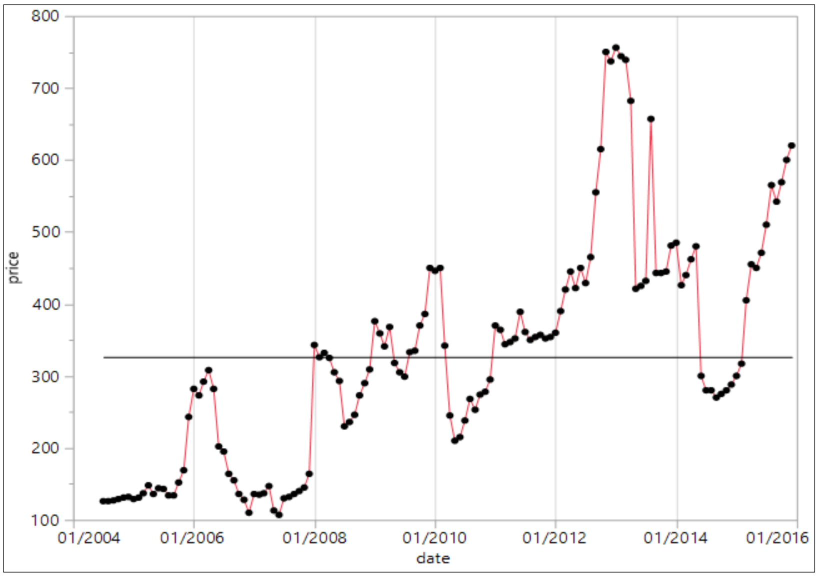

Fig 4.3.1: A Time Series Plot of Monthly Maize Prices Data (from Jan 2005 to December 2015) at Kibaigwa Market

Figure 4.3.1 above specifies a time series plot of the monthly maize prices data of the Dodoma region specifically at Kibaigwa market from January 2005 to 2015. After the time series plot in maize prices, it is obviously detected that there were variations in prices with increase in time which in turn indicated the non-stationary of the series in variance. The best idea was to start with differencing with the lowest order (d=1) and test the data for unit root problems. So we obtained a time series of second order differencing and figure 4.3.2 below is the plot of the second order differenced maize data at Kibaigwa market.

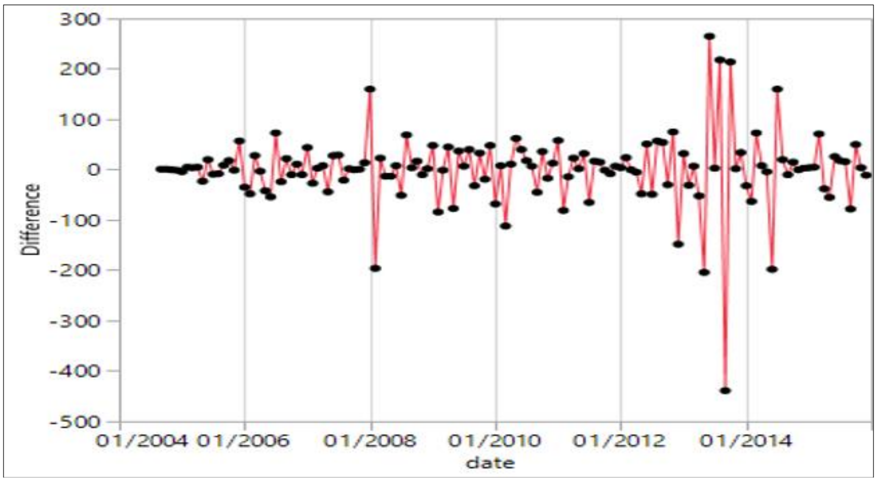

Fig 4.3.2: Plot of the First Differenced Maize Prices Data (d=1) at Kibaigwa Market

From figure 4.3.2 above, it can easily be concluded that the time series appears to be stationary in both mean and variance. But before moving into another step, Augmented Dickey-Fuller test was applied to the differenced time series data also for testing stationary.

Test for Stationary ADF Test

Using ADF test, there is null hypothesis (H_o) and alternative hypothesis (H_1) where by H_o represents non-stationary time series data while H_1 represents the stationary time series data. The hypothesis is then tested by carrying out appropriate differencing of the data in the dth order and applying the ADF test to the differenced time series data. The first order differencing of the data (d=1) means the table of differenced data of current and previous one $(X_t = X_t - X_{t-1})$ is created. The ADF test result is shown in Table 4.3.1 below.

Table 4.3.1: Test Results

Lags	P-Value	Significance level
0	0.001	<0.05
5	0.001	<0.05
10	0.001	<0.05
15	0.0043	<0.05
15	0.0043	<0.05

For maize prices at Kibaigwa market, the ADF test statistic in table 4.3.1 above is less than 0.05 p-value for lag order 0 to 15 showing that the series is stationary, therefore, H_o is rejected and the conclusion is that the alternative hypothesis is true which means, the series is stationary in both mean and variance. Therefore, it is not necessary to add another differencing and ARIMA (p, d, q) is adopted where d=1.

This step is essential since it is helpful in further steps for ARIMA (p, 1, q) in finding suitable p in AR and q in MA process. Now, the next step is to examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Figure 4.3.2a below represents the plot of ACF for lag 1 to 25 of the second order differenced time series of the maize prices in Kongwa District.

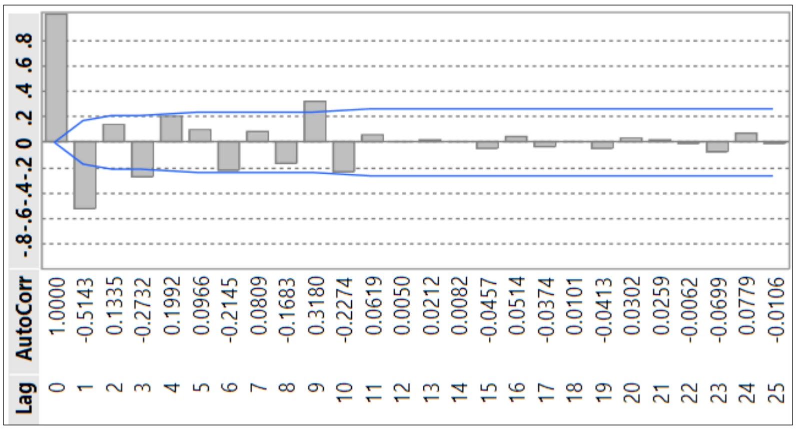

Fig 4.3.3: Estimated ACF of maize prices data at Kibaigwa Market

The above ACF infers that the autocorrelation at lag 1and lag 3 exceeds the significance limits and autocorrelation tails off to zero thereafter except at lag 9. The autocorrelation, which exceeds the significant limits can be concluded as errors that happened by chance.

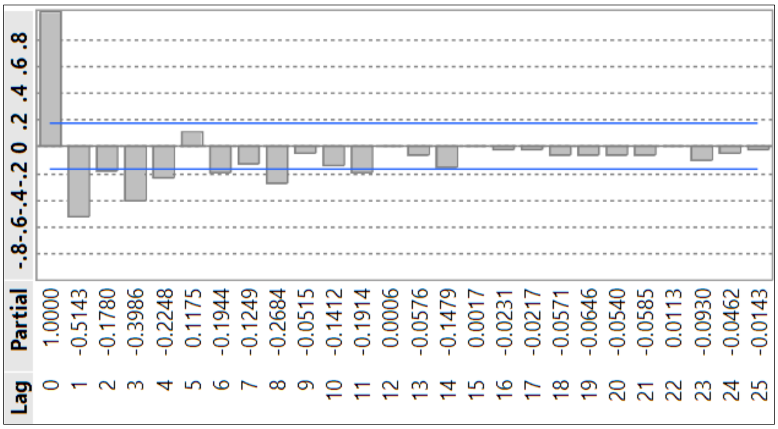

Fig 4.3.4: Estimated PACF of the Maize Prices Data at Kibaigwa Market

Figure 4.3.4 above represents the partial autocorrelation function (PACF) for the first order differenced time series from lag 1 to lag 25. It is concluded that PACF exceeds significant limits at lag 1 and lag 3, after lag 8 PACF tails off to zero. Those outside the significant limit can be assumed as errors that happened by chance because all PACFs from lag 9 to 25 are within the significant limits. The ACF tailing off to zero after lag 9 and the PACF tailing off to zero after lag 8 can define the following possible ARIMA models for the first differenced time series data of maize prices at Kibaigwa market.

Suggested Models for the First Differenced Maize Prices Data

1. ARIMA (2,2,3)
2. ARIMA (1,2,1)
3. ARIMA (3,2,2)
4. ARIMA (2,2,2)
5. ARIMA (1,2,2)

Model Selection Criteria

To select the best model from the five models above, ARIMA model with the lowest Akaike Information Criterion (AIC) values was selected. The following 4.3.2 Table summarizes the output of each fitted ARIMA model in the time series maize prices data at Kibaigwa market.

Table 4.3.2: AIC and BIC Values of the Fitted Model ARIMA (2, 2, 3)

MODEL	Variance	AIC	BIC
ARIMA (2, 2, 3)	2578.711	1464.278	1478.841
ARIMA (1, 2, 1)	2735.134	1469.075	1474.901
ARIMA (3, 2, 2)	2679.271	1469.438	1484.001
ARIMA (2, 2, 2)	2727.043	1470.493	1482.144
ARIMA (1, 2, 2)	2750.404	1470.937	1479.675

Based on AIC values, the best model was found to be ARIMA (2, 2, 3). Hence, this is the best model for predicting future values of maize prices at Kibaigwa market.

Table 4.3.3: Estimation Summary for the ARIMA (2, 2, 3) Model

Models	Coefficient	Std Error
AR1 ϕ_1	-1.21824	0.189413
AR2 ϕ_2	-0.57828	0.158855
MA1 θ_1	-0.29448	0.145219
MA2 θ_2	0.516	0.103701
MA3 θ_3	0.778478	0.113854

From table 4.3.3, we fit the maize prices ARIMA time series model using a form of $(1-\phi_1 B-\phi_2 B^2)(1-B)^2 X_t=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)\varepsilon_t$ where the estimates of parameters are found as $\hat{\varphi}_1=-1.2182$ $\hat{\varphi}_2=-0.5783, \hat{\theta}_1=-0.2945, \hat{\theta}_2=0.5160, \hat{\theta}_3=0.7785.$ Hence the estimated equation for ARIMA (2, 2, 3) is obtained as $(1+1.2182B+0.5783B^2)(1-B)^2 X_t=(1+0.2945B-0.516B^2-0.7785B^3)\varepsilon_t$

From this study, the fitted model for maize prices time series data at Kibaigwa market is ARIMA (2, 2, 3). It consists of both AR and MA processes and it is free from seasonal component, which means that seasonality in maize prices is non-significant. The model adequacy is further tested to draw empirical conclusions regarding the model as good fit for forecasting time series. These tests are performed using the Ljung-box test in addition to ACF and PACF residuals plots.

Diagnostic Checking

The diagnostics of the residuals by ACF and PACF values in figure 4.3.4 show that the ACF and PACF values are all within the 95% confidence limit indicating that there is no correlation among residuals. The plot of the fitted ARIMA model (2,2,3) is presented in Figure 4.3.7.

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.0108	0.0163	0.8985	1	0.0108
2	-0.0420	0.2632	0.8767	2	-0.0421
3	-0.0404	0.4932	0.9204	3	-0.0395
4	0.0240	0.5750	0.9658	4	0.0231
5	0.0837	1.5794	0.9037	5	0.0802
6	-0.2034	7.5550	0.2726	6	-0.2068
7	-0.1635	11.4421	0.1205	7	-0.1565
8	-0.0329	11.6005	0.1699	8	-0.0417
9	0.1482	14.8470	0.0952	9	0.1284
10	-0.1012	16.3736	0.0894	10	-0.1249
11	0.0067	16.3804	0.1276	11	0.0521
12	0.0434	16.6656	0.1626	12	0.0361
13	0.0290	16.7936	0.2089	13	-0.0406
14	0.0001	16.7936	0.2673	14	-0.0670
15	-0.0441	17.0946	0.3132	15	0.0258
16	-0.0133	17.1224	0.3777	16	-0.0241
17	-0.0672	17.8336	0.3994	17	-0.1035
18	-0.0621	18.4462	0.4266	18	-0.0723
19	-0.0615	19.0521	0.4535	19	-0.0144
20	-0.0033	19.0538	0.5183	20	-0.0481
21	0.0228	19.1384	0.5763	21	-0.0050
22	-0.0387	19.3845	0.6215	22	-0.0348
23	-0.0387	19.6337	0.6639	23	-0.0703
24	0.0463	19.9922	0.6972	24	-0.0128
25	0.0230	20.0816	0.7425	25	-0.0225

Fig 4.3.4: ACF and PACF Residuals for Maize Prices at Kibaigwa Market

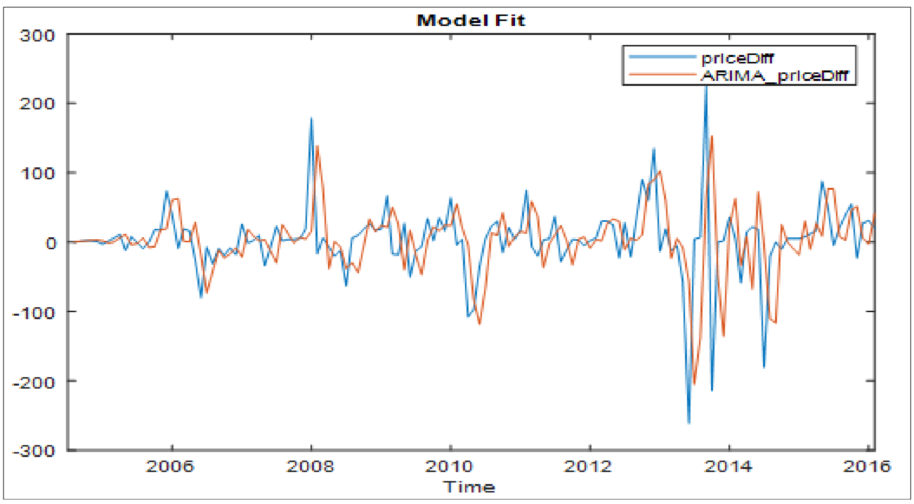

Fig 4.3.5: The Plot of Fitted ARIMA (2,2,3) Model

Forecasting with the Fitted Model at Kibaigwa Market

After diagnostic checking, the model fit can be used to predict the upcoming values of the variable of interest if it is adequate. However, before forecasting, we need to measure the accuracy of its predictions and it is completed by one-step-ahead forecasting. As a result, the fitted model ARIMA (2, 2, 3) was used to predict the maize prices for twelve months (January to December 2016). The observed and predicted values are shown in Table 4.3.4.

Table 4.3.4: One-step-ahead Forecast of Maize Prices at Kibaigwa Market

Period	Observed value	Forecast	Forecast error	Absolute error	squared error	Absolute % error
16-Jan	620	620.8566	-0.85664	0.85664	0.733832	0.138168
16-Feb	631	619.564	11.43602	11.43602	130.7825	1.812364
16-Mar	520	630.6651	-110.665	110.665	12246.74	21.28173
16-Apr	425	627.9105	-202.91	202.91	41172.47	47.74353
16-May	470	634.8685	-164.868	164.868	27181.46	35.0783
16-Jun	478	638.0066	-160.007	160.007	25602.24	33.47427
16-Jul	453	640.1817	-187.182	187.182	35037.1	41.32053
16-Aug	490	645.7389	-155.739	155.739	24254.64	31.78347
16-Sep	543	647.7328	-104.733	104.733	10969	19.28785
16-Oct	495	652.1118	-157.112	157.112	24684.18	31.7398
16-Nov	704	655.6458	48.35423	48.35423	2338.132	6.868499
16-Dec	829	658.83	170.17	170.17	28957.84	20.52714
Total	6658	7672.112	-1014.11	1474.033	232575.3	291.0556

Forecasting Accuracy of the Fitted Model

To assess the model predicting capability, the standard measures of forecast accuracy were obtained as shown in Table 4.3.5. The values of these measures were obtained using the procedures stated in methodological part and the forecast errors existing in Table 4.3.4.

Table 4.3.5: Measures of Forecasting Accuracy

Variables	Maize prices
Mean Squared error (MSE)	19381.28
Mean Absolute error (MAE)	122.84
Mean percentage error (MPE)	-19.39
Mean absolute percentage error (MAPE)	24.25

Analysis of forecasting errors

The analysis of forecast errors is very important because this aspect is used in evaluating the accuracy of future forecasts of the fitted model. To assess the model forecasting capability, we consider the standard measures of forecast accuracy in Table 4.3.5. These measures of forecast accuracy are evaluated as part of validation of the fitted models. The mean forecast error values for maize prices are not close to zero indicating that the forecasts produced by the fit ARIMA (2, 2, 3) are not unbiased. Both MAE and MSE indicate that there is variability in forecasting errors with the fitted model for maize prices. The relative forecast error shows that MAPE value for maize prices is 24.25%.

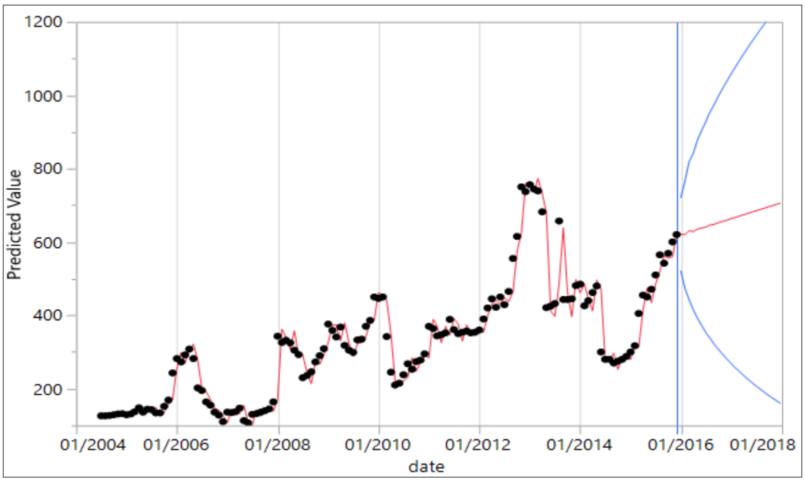

Fig 4.3. 6: Time Plot for Forecast Maize Prices at Kibaigwa Market

In figure 4.3.9 above, the two blue lines of forecast represent the 95% (lower and upper side) projection of the forecasting intervals. The maize prices are expected to increase with time from January 2016 to December 2016.

References

1. Adejumo AO, Momo AA. Modeling Box-Jenkins Methodology on Retail Prices of Rice in Nigeria, Int. J. Eng. Sci. 2013; 2(9):75-83.
2. Badmus MA, Ariyo OS. Forecasting cultivated areas and production of maize in Nigeria using ARIMA model. Asian J Agric. Sci. 2011; 3(3):171-176.
3. Box GEP, Jenkin GM. Time Series of Analysis, Forecasting and Control, Sam Franscico, Holden-Day, California. USA, 1976.
4. Contreras J, Espínola R, Nogales FJ, Coneja AJ. ARIMA Models to Predict Next-Day Electricity Prices. IEEE Transactions on power systems. 2003; 18(3):1014-1020.
5. Darekar A, Reddy AA. Forecasting of Common Paddy Prices in India, J Rice Research. 2017; 10(1):71-75.
6. Gertler P, Martinez S, Rawlings LB, Premand P, Vermeersch CMJ. Impact Evaulation in Practice, 2016. Doi.Org/10.1109/WI-IATW.2006.145

7. Kibona SE, Mbago MC. Forecasting Wholesale Prices of Maize in Tanzania Using Arima Model, 2017. https://doi.org/10.31559/glm2018.4.3.6
8. Kirimi J. Modelling the Volatility of Maize Prices Using Autoregressive Integrated Moving Average Model. Mathematical Theory and Modeling. 2016; 6(6):143-149. www.iiste.org,ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
9. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978; 65:67-72.
10. Manoj K, Anand Madhu. An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India. Studies in Business and Economics, 2017, 81-94.
11. Nkonya E, Xavery P, Akonaay H, Mwangi W, Anandajayasekeram P, Verkuijl H. Adoption of Maize Production Technologies in Northern Tanzania. Mexico D.F: International Maize and Wheat Improvement Center (CIMMYT), The United Republic of Tanzania, and the Southern African Center for Cooperation in Agricultural Research (SACCAR), 1998. ISBN: 970-648-003-X
12. Nathaniel Katinila H, Verkuijl W, Mwangi P, Anandajayasekeram, Moshi AJ. Adoption of Maize Production Technologies in Southern Tanzania. Mexico, D.F.: International Maize and Wheat Improvement Center (CIMMYT), the United Republic of Tanzania, and the Southern Africa Centre for Cooperation in Agricultural Research (SACCAR), 1998.
13. Nyangarika A, Mikhaylov A, Richter UH. Oil Price Factors : Forecasting on the Base of Modified Auto-Regressive Integrated Moving Average Model. International Journal of Energy Economics and Policy. 2019; 9(1):149-159. ISSN: 2146-4553.
14. Pierre G, Morales-Opazo C, Demeke M. Analysis and determinants of retail and wholesale staple food price volatility in developing countries, Scientific. January 2014, Food and Agricultural Organization of the UN (FAO), 2014, 3.
15. Pindyck RS, Rubinfeld DL. Econometric Models and Economic Forecasts, McGraw-Hill, 1981.
16. Rachana W, Suvarna M, Sonal G. Use of ARIMA models for forecasting pigeon pea production in India. Int. Rev. Bus. Financ. 2010; 2(1):97-107.
17. Rangsan Nochai, Titida Nochai. ARIMA model for forecasting oil palm price, Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Universiti Sains Malaysia, Penang, 2006.
18. Sohail A, Sarwar A, Kamran M. Forecasting total food grains in Pakistan. Department of Math and Stat, University of Agriculture, Faisalabad. J Eng. Appl. Sci. 1994; 13:140-146.
19. Urassa Justin K. Factors influencing maize crop production at household levels: A case of Rukwa Region in the southern highlands of Tanzania, African Journal of Agricultural Research. 2015; 10(10):1097-1106.
20. Venkatesh Panasa R, Vijaya Kumari, Ramakrishna G, Kaviraju S. Maize Price Forecasting Using Auto Regressive Integrated Moving Average (ARIMA) Model, International Journal of Current Microbiology and Applied Sciences. 2017; 6(8):2887-2895, ISSN: 2319-7706
21. Waryoba FD. Supply Response of Maize in Tanzania Supply Response of Maize in Tanzania MA. (Economics) Dissertation University of Dar Es Salaam, 2015. Https://Doi.Org/10.13140/RG.2.1.1454.4720.