

# Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya: Time Series Analysis Using Sarima Model

**Robert Mathenge Mutwiri**

School of Pure and Applied Sciences, Kirinyaga University, Kirugoya, Kenya

**Email address:**

[mutwiri@outlook.com](mailto:mutwiri@outlook.com)

**To cite this article:**

Robert Mathenge Mutwiri. Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya: Time Series Analysis Using Sarima Model. *International Journal of Statistical Distributions and Applications*. Vol. 5, No. 3, 2019, pp. 46-53. doi: 10.11648/j.ijstd.20190503.11

**Received:** June 23, 2019; **Accepted:** July 17, 2019; **Published:** August 5, 2019

---

**Abstract:** Price forecasting is more sensitive with vegetable crops due to their high nature of perishability and seasonality and is often used to make better-informed decisions and to manage price risk. This is achievable if an appropriate model with high predictive accuracy is used. In this paper, Seasonal Autoregressive Integrated Moving Average (SARIMA) model is developed to forecast price of tomatoes using monthly data for the period 1981 to 2013 obtained from the Ministry of Agriculture, Livestock and Fisheries (MALF) in the agribusiness department. Forecasting tomato prices was done using time series monthly average prices from January 2003 to December 2016. SARIMA (2, 1, 1) (1, 0, 1)<sub>12</sub> was identified as the best model. This was achieved by identifying the model with the least Akaike Information Criterion. The parameters were then estimated through the Maximum Likelihood Estimation method. The time series data of Tomatoes for wholesale markets in Nairobi are considered as the national average. The predictive ability tests RMSE = 32.063, MAPE = 125.251 and MAE = 22.3 showed that the model was appropriate for forecasting the price of tomatoes in Nairobi County, Kenya.

**Keywords:** Tomatoes, SARIMA, Autocorrelation Function, Akaike Information Criterion, Jarque-Bera Test

---

## 1. Introduction

In Kenya, the agriculture sector is the mainstay of the economy, contributing about 30% of the GDP and accounting for about 80% of employments. The total domestic value of the horticulture sector in 2012 amounted to Ksh 217 billion occupying an area of 662,835 ha with a total production quantity of 12.6 million tons. As compared to 2011, the total value, area and production increased by 6%, 9% and 38% respectively [13]. Vegetables contributed about 38% of the domestic value of horticulture with 287,000 ha under production and producing 5.3 million tons valued at Ksh 91.3 billion. Production increased by 13% while there was a slight reduction in value by 4% from 2011 levels. The increased production is occasioned by favourable weather conditions that resulting to high yield, thus reducing the value of vegetables. However, there was a drop in prices for commodities like Cabbages and Tomatoes thereby reducing the overall value for the year.

According to [13] tomato (*Lycopersicon esculentum mill*) is amongst the promising commodities in horticultural expansion and development in Kenya. It accounts for about

14% of the total vegetable produce and 6.72% of the total horticultural crops [12, 13]. Tomato is grown either on open field or under greenhouse technology. Open field production account for about 95% while greenhouse technology accounts for 5% of the total tomato production [13].

The aim of this paper is to analyze the price fluctuations of tomatoes in Nairobi County Kenya using SARIMA model as the analysis tool. Vegetable price analysis is used to formulate price stability policy and increased production. Accurate public information to farmers and market stakeholders like middlemen can inform policy forecasters to reduce price variance in other markets. The application of SARIMA as analysis tool can give an early warning message of tomatoes price fluctuation in the future.

## 2. Review of Previous Studies

According to [15] wholesale prices for vegetables are characterized by large seasonal variations, the degree and different timing of the changes. Due to the price fluctuations, vegetable producers normally have large losses, therefore the adaption of production to seasons, market research and

technological development should all be improved. Assis et al. [14] compared four different Univariate time series models: exponential smoothing, autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroscedasticity (GARCH) and the mixed ARMA and GARCH models. Dieng [16] investigated the performance of parametric models in forecasting selected vegetable prices in Senegal and suggested the use of the parametric ARIMA model as compared to the non-parametric models.

The efficiency of ARIMA and GARCH models were compared for modelling and forecasting the spot prices of Grams in India [23] and the SARIMA models were used to forecast the prices of Tomatoes in selected Indian states [19]. Gathondu [11] fitted four models to wholesale prices of major vegetables: tomato, potato, cabbages, kales and onions for markets in Nairobi, Mombasa, Kisumu, Eldoret and Nakuru in Kenya using Autoregressive Moving Average (ARMA), Vector Autoregressive (VAR), Generalized Autoregressive Condition Heterostadictity (GARCH) and the mixed model of ARMA and GARCH. In the study they found ARIMA (3, 1, 2) to be the best fitting model for tomatoes. The model failed to capture seasonal variability. Dragan et. al., [18] analyzed the changes and future tendencies of the price of tomatoes with descriptive statistics and found that the ARIMA models were suitable for price forecasting. Sampson et. al., [17] argued that among the seasonal decomposition models of forecasting, the Seasonal Autoregressive Integrated Moving average (SARIMA) models could enable producers to achieve better market prices by adopting the practice. In another study, [20] applied SARIMA models to forecast the prices of tomatoes in Turkey and found SARIMA (1, 0, 0) (1, 1, 1)<sub>12</sub> model as the most suitable. They reported that the highest tomatoe prices seasonally adjusted were in October. Boateng et al., [19] found that the predictability of the model increased with seasonal ARIMA (SARIMA). They noted wide fluctuations in prices of tomatoes in different months, prices sometimes increase 10 times compared to prices during peak harvest periods which implied that if farmers plan their area under tomatoes properly, sowing dates and sales by considering forecasted prices from the ARIMA models to receive increased prices, earnings may increase at least three to four times with 90% predictability of the forecast accuracy. According to [21], accurate prediction of agricultural prices is beneficial to correctly guide the circulation of agricultural products and agricultural production and realize the equilibrium of supply and demand of the agricultural area.

### 3. Material and Methods

#### 3.1. Data Overview

The wholesale price data is gathered from the Ministry of Agriculture, Livestock and Fisheries (MALF) in the agribusiness department which was collected by extension officers in the various wholesale markets. The data was

available on weekly prices and covered the four year period from 2003 to 2018 which was computed to obtain monthly average prices. Under this study, the average wholesale prices for markets Nairobi County is considered as the classical national average. The time series data is measured in Kenya shillings per Kilograms (Ksh/Kg) and the data ranged from January 2004 until December 2018.

#### 3.2. ARIMA Model

A generalization of ARMA models which incorporates a wide class of non-stationary time-series is obtained by introducing the differencing into the model. The simplest example of a non-stationary process which reduces to a stationary one after differencing is Random Walk. A process  $\{y_t\}$  is said to follow integrated ARIMA model denoted by ARIMA (p, d, q), if  $\nabla^d y_t = (1 - B)^d \varepsilon_t$ , is ARIMA (p, q). The model is written as Equation (1):

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t \quad (1)$$

The ARIMA methodology is carried out in three stages, viz. identification, estimation and diagnostic checking. Parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage and adequacy of selected model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration. An excellent discussion of various aspects of this approach is given in Box and Jenkins [3]. Most of the standard software packages, like SAS, and RGui contain packages and procedures for fitting of ARIMA models.

#### 3.3. SARIMA Models

SARIMA models are an adaptation of autoregressive integrated moving average (ARIMA) models to specifically fit seasonal time series. That is, their construction takes into account the underlying seasonal nature of the series to be modelled. Seasonality in a time series refers to a regular pattern of changes that repeats over in time-periods, where S defines the number of time-periods until the pattern repeats again. For monthly rainfall data  $S = 12$ . In a seasonal ARIMA model, seasonal AR and MA terms predict  $x_t$  using data values and errors at times with lags that are multiples of S (the span of the seasonality). The seasonal ARIMA model incorporates non-seasonal and seasonal factors in a multiplicative model and is denoted as:

$$\text{ARIMA}(p, d, q) \times (P, D, Q)_S$$

Where p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Without differencing operations, the model can be written as:

$$\Phi(BS)\varphi(B)(x_t - \mu) = \Theta(BS)\theta(B)w_t \quad (2)$$

The non-seasonal components are:

$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \quad (3)$$

$$\text{MA: } \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (4)$$

The seasonal components are:

Seasonal AR:

$$\Phi(BS) = 1 - \Phi_1 BS - \dots - \Phi_P B^P S \quad (5)$$

Season MA:

$$\Theta(BS) = 1 + \Theta_1 BS + \dots + \Theta_Q B^Q S \quad (6)$$

### 3.3.1. Model Identification in SARIMA

The first step of applying the model is to identify appropriate order of ARIMA (p, d, q) model. Identification of ARIMA model involves selection of order of AR (p), MA (q) and I (d). The order of d is estimated through I (1) or I (0) process. The model specification and selection of order p and q involves plotting of ACF and partial PACF or correlogram of variables at different lag lengths. The significance level of individual coefficients is measured by Box-Pierce Q statistics and jointly together by Ljung-Box LB statistics. The Box-Pierce Q statistics is defined as;

$$Q = \sum_{k=1}^m \widehat{\rho}_k^2 \sim \chi_m^2 \quad (7)$$

Where n = sample size and m is lag length. And Ljung Box (LB) Statistics is defined by:

$$LB = n(n+2) \sum_{k=1}^m \frac{\widehat{\rho}_k^2}{n-k} \sim \chi_m^2 \quad (8)$$

Where n = sample size and m is the lag length of the date. The possible SARIMA model is determined that best fit the data under consideration. SARIMA model is appropriate for stationary time series therefore, the data under consideration must satisfy the condition of stationarity that is the mean, variance and autocorrelation are constant over time.

### 3.3.2. Parameter Estimation SARIMA

To estimate SARIMA models the ML method is used. Under the assumption of independent and distributed standardized  $z_t$ , the log-likelihood (LL) function of  $\{y_t(\theta)\}$  for a T observations sample, is given by:

$$\ln L(y_t, \theta) = \sum_{t=1}^T \left[ \ln[D(z_t(\theta), v)] - \frac{1}{2} \ln[\sigma_t^2(\theta)] \right] \quad (9)$$

where  $\theta$  is the vector of the parameters that have to be estimated for the conditional mean, conditional variance and density function.  $z_t$  is a sequence of independent and distributed random variables with mean as zero and variance as one. The approach of maximum likelihood (ML) requires the specification of a particular distribution for a sample of T observations  $y_t$ .

$$f(Y_t, Y_{t-1}, \dots, Y_t = y_t) = f(y_{T-1}, \dots, y_t | \Psi) \quad (10)$$

denote the probability density of the sample given the unknown parameters ( $n \times 1$ ) parameters  $\Psi$ . Following the

notation of Box and Jenkins,  $L(\Psi|y)$  with respect to derivatives to zero and using vector notation and suppressing y the result becomes teach of the unknown parameters of the vector  $\Psi$  the notation is the most appropriate. Setting the  $\frac{\partial(\varphi)}{\partial \varphi} = 0$ . As a rule, the likelihood equations are non-linear. Therefore, the ML estimates must be found in the course of an iterative procedure. This is true for the exact likelihood function of every Gaussian ARMA (p, q) process.

### 3.3.3. Forecasting with the SARIMA Model

Forecasting is the process of making statements about events whose actual outcomes have not yet been observed. It is an important application of time series. After the model has passed the entire diagnostic test, it becomes adequate for forecasting which the last step is in Box-Jenkins model building approach. For instance, let us consider the given Seasonal ARIMA (0, 1, 1) (1, 0, 1)<sub>12</sub> we can forecast the next step which is given by Cryer and Chan [24] as:

$$z_t - z_{t-1} = \Phi(z_{t-12} - z_{t-13}) + \varepsilon_t - \theta \varepsilon_{t-1} - \Phi \varepsilon_{t-12} + \theta \varepsilon_{t-13} \quad (11)$$

$$z_t = z_{t-1} + \Phi z_{t-12} - \Phi z_{t-13} + \varepsilon_t - \theta \varepsilon_{t-1} - \Phi \varepsilon_{t-12} + \theta \varepsilon_{t-13} \quad (12)$$

The one step ahead forecast from the origin t is given by

$$\hat{z}_{t+1} = z_t + \Phi z_{t-11} - \Phi z_{t-12} - \theta \varepsilon_t - \Phi \varepsilon_{t-11} + \theta \varepsilon_{t-12} \quad (13)$$

The next step is

$$\hat{z}_{t+2} = \hat{z}_{t-1} + \Phi z_{t-10} - \Phi z_{t-11} - \Phi \varepsilon_{t-10} + \theta \varepsilon_{t-11} \quad (14)$$

and so on. The noise term  $\varepsilon_{13}, \varepsilon_{12}, \varepsilon_{11}, \dots, \varepsilon_1$  (as residuals) will enter into the forecasts for lead times  $l = 1, 2, \dots, 13$ , but for  $l > 13$  the autoregressive part of the model takes over;

$$\hat{z}_{t+l} = \hat{z}_{t-l+1} + \Phi z_{t+l-12} - \Phi z_{t+l-13}, \text{ for } l > 13 \quad (15)$$

### 3.3.4. Forecasting Performance

The accuracy for each model can be checked to determine how the model performed in terms of in-sample forecast. In terms of out sample forecasting, some of the observations are left out during model building. The accuracy of the model can be compared using forecast measure or some statistic such as mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean square error (MSE) etc. The model with the minimum of MAE or RMSE is considered to be the best for forecasting. The mathematical expressions are defined as:

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| = \frac{1}{T} \sum_{t=1}^T |e_t| \quad (16)$$

$$MSE = \frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2 = \frac{1}{T} \sum_{t=1}^T (e_t)^2 \quad (17)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} = \sqrt{\frac{1}{T} \sum_{t=1}^T (e_t)^2} \quad (18)$$

where  $y_t$  is the actual observation,  $\hat{y}_t$  is fitted or the forecast value and T is the sample size. If we have perfect forecast then  $MAE = MSE = RMSE = 0$ . The smaller the value the better the prediction and the big the value the poorer the

predictive power of the model.

## 4. Empirical Results

The data used in this study is average monthly prices of tomatoes in Nairobi County from 2003 to 2016. The wholesale price data was gathered from the Ministry of Agriculture, Livestock and Fisheries (MALF) in the agribusiness department which was collected by extension officers in the various wholesale markets. Figure 2 shows the plot of Kenya's monthly price of tomatoes.

The prices of tomatoes vary along the year with the time period from 2003 to 2016 (Figure 1). The prices show a declining tendency during the first quarter of the year. They are picked in April and prices decline in June because of poor weather and the excess supply of tomatoes. Moreover, supply to the wholesale market is on the increase and picked in the May and June (Figure 1). The trend of wholesale price of tomatoes has an increasing trend after the period of June and

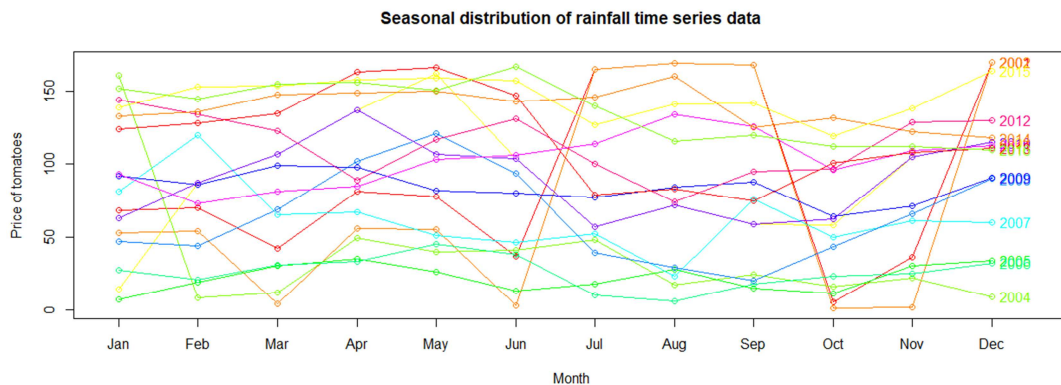
continues in the last quarter of the year. The average real tomato price is KSh 87.27/kg in time period between 2003 and 2016. And the lowest tomato price was KSh 61.875 per 64kg crate in October and the highest price was KSh 100.8125 per 64kg crate in May (Table 1). Figure 1 shows the presence of trend and seasonality in each time series of tomatoes price in Nairobi, county.

Table 1 shows the monthly descriptive statistics of tomatoe prices. The average monthly price was highest in December and lowest in October.

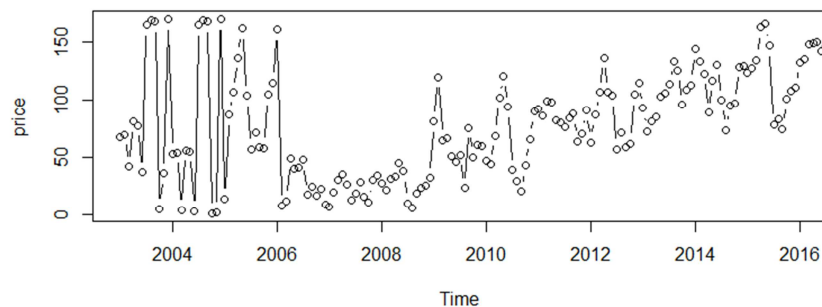
As observed from Figure 4, tomato prices do not indicate a significant trend. This indicates that the series is in a stationary structure. In fact, null hypothesis was rejected in Augmented Dickey-Fuller (ADF) test, which was performed to determine if the series is stationary or not. This shows that the series does not have a root unit which means it is stationary (Table 2).

**Table 1.** Descriptive analysis of monthly wholesale prices of tomatoes in Nairobi County

Month	N	Minimum	Maximum	mean	std. dev
January	16	7	161	87.375	50.50132
February	16	8	153	85.3125	47.39862
March	16	4	155	85.125	51.04883
April	16	33	163	99.6875	45.21168
May	16	26	166	100.8125	48.23169
June	16	3	167	88.1875	53.05119
July	16	10	165	87.125	50.76728
August	16	6	169	86.0625	56.43577
September	16	15	168	86.125	51.95495
October	16	1	132	61.875	43.4233
November	16	2	138	77.5625	43.87857
December	16	9	170	102	48.26179



**Figure 1.** Seasonal distribution of prices of tomatoes in Nairobi County.



**Figure 2.** Time series plot of prices of tomatoes in Nairobi County.

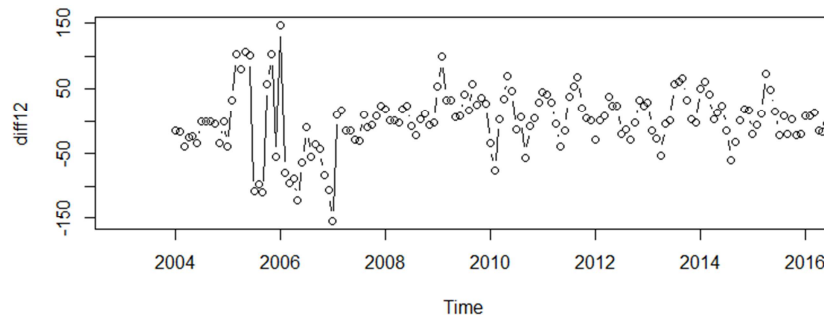


Figure 3. Time series plot of differenced prices of tomatoes in Nairobi County.

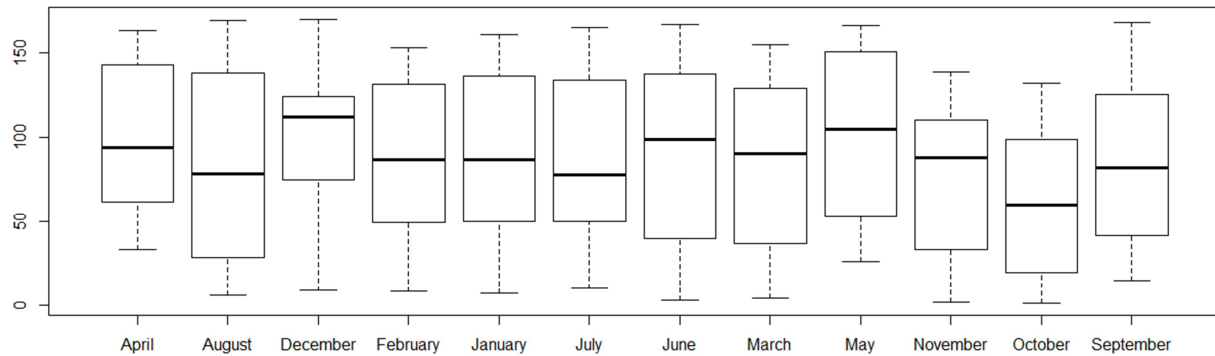


Figure 4. Monthly distribution of prices of tomatoes in Nairobi County.

Table 2. ADF Test Result for Unit Root.

Dickey-Fuller	Lag Order	P value
-3.8619	5	0.01729

H0: Tomatoe series has a unit root (there is non-Stationarity and no seasonality).

H1: The series is stationary.

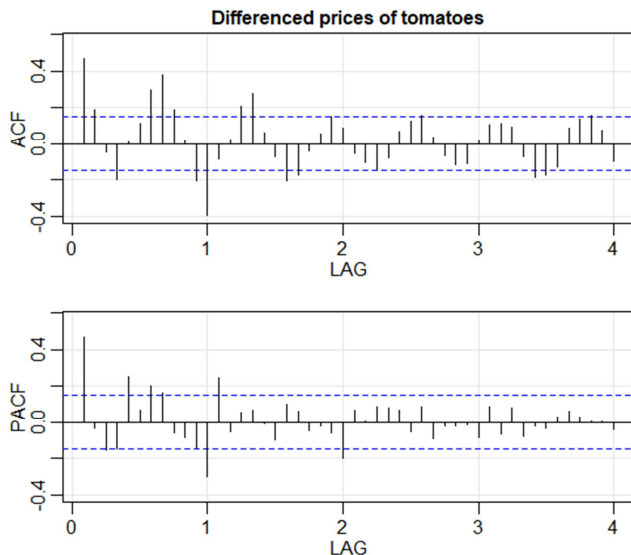


Figure 5. ACF and PACF plot of prices of tomatoes in Nairobi County.

ACF and PACF values of the series from which the seasonal differences are taken are presented in Figure 6. The seasonal spikes at ACF and PACF after 1 lag (12, 24,...) are observed as being cut off after taking the seasonal difference of the series. This also indicates the seasonal model of AR

(1) and MA (1). Therefore, to include the model of (1, 1, 1) to the part (P, D, Q) of the model will be formed can be considered as one of the best possibilities among the alternative choices. At the non-seasonal part of the model (p, d, q); the discontinuation of PACF value after 1 lag indicates that the addition of the AR (1) term may be appropriate (see Figure 5). On the other hand, even the discontinuations occurs after 1 lag at ACF values, these values are observed to be increased after a certain lag. Therefore, there is no clarity for the MA term at the non-seasonal part of the model. In this case, two alternatives to be taken into account occur for the non-seasonal part of the model. One of these alternatives is the model which MA term is not added to (1, 0, 0), and the other one which MA term is added (1, 1).

Nine possible alternative models were analyzed based on the seasonal part of the model in order to select the SARIMA model which will be used to forecast the prices of tomatoes. The analyzed models are compared according to the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC). The model selected should have the smallest AIC and SBC values (Wang and Lim, 2005). SARIMA (2, 1, 1)  $\times$  (1, 0, 1)<sub>12</sub> model has comparatively lower AIC and SBC values (Table 3). Therefore, this model was selected as the most suitable model or the best fit model from amongst the four models.

**Table 3.** Akaike Information Criteria (AIC), Schwarz Bayesian Criterion (SBC), and considering different SARIMA (p, 0, q) (1, 1, 1)<sub>12</sub> models.

Model	DF	AIC	AICc	BIC
SARIMA (0, 0, 1), (0, 0, 1) <sub>12</sub>	189	8.20495	8.21648	7.255851
SARIMA (1, 0, 0), (0, 0, 1) <sub>12</sub>	189	8.08220	8.09373	7.133096
SARIMA (0, 0, 0), (0, 0, 1) <sub>12</sub>	190	8.529216	8.540300	7.563148
SARIMA (1, 0, 0), (0, 0, 2) <sub>12</sub>	188	8.054761	8.086686	7.122625
SARIMA (2, 1, 1), (1, 0, 1) <sub>12</sub>	185	8.0026	8.063865	7.103232
SARIMA (1, 1, 1), (0, 1, 1) <sub>12</sub>	176	8.052334	10.00498	9.039445
SARIMA (1, 1, 1), (0, 1, 1) <sub>12</sub>	174	8.052956	8.065738	7.137787
SARIMA (1, 1, 1), (1, 1, 2) <sub>12</sub>	174	8.052956	8.065738	7.137787
SARIMA (1, 1, 1), (0, 1, 2) <sub>12</sub>	176	8.112171	8.123702	7.16307

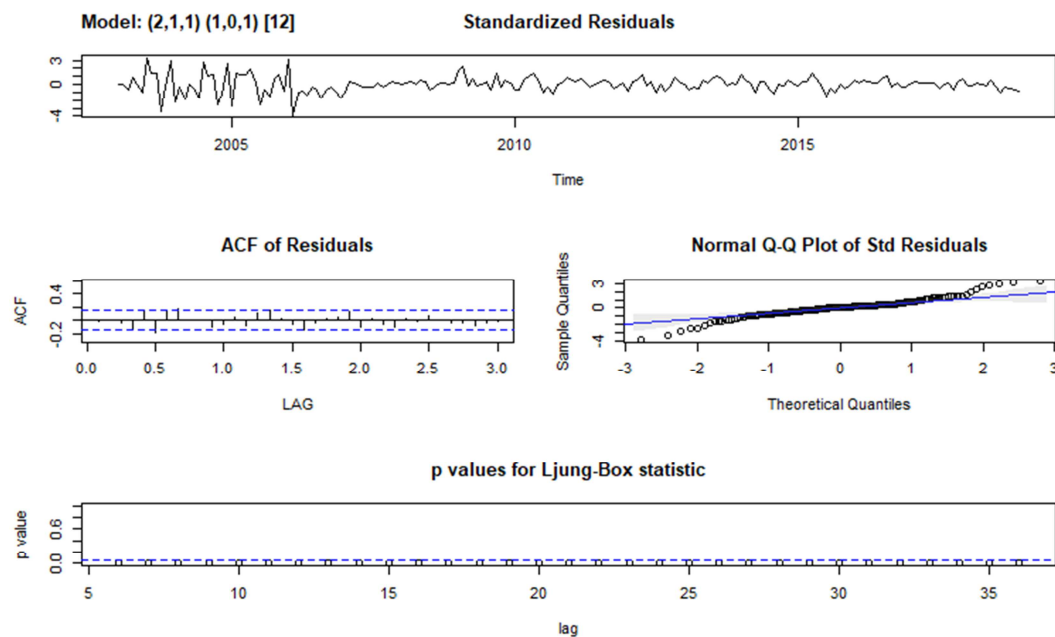
Considering this model, the autoregressive and seasonal parameters were estimated respectively (Table 4). Although the constant term in the estimated SARIMA model is not significant at the different levels, the autoregressive and seasonal parameters are significant at the 1% level.

After estimating the parameters of this model, further analysis was done with the selected SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> model to check whether the residuals of the model are independent. The autocorrelation and partial autocorrelation up to 36 lags were computed and their significance was tested using Box-Ljung test. It is evident from Figure 6 that the values of the SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> residuals lie within the upper and lower confidence limits. Panel (b) shows p-values for the Ljung-Box statistics. Given the high p-values associated with the statistics, we cannot reject the null hypothesis of independence in this residual series. The results indicate that none of these correlations are

significantly different from zero at a 95% confidence level. This shows that the selected SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> model is appropriate model for the monthly tomato price forecasting.

The residuals were checked to find out if they followed a white noise process. This was achieved by plotting the residual Q-Q and normality test plots as shown in Figure 7. The Q-Q plot is reasonably straight so normality is okay. The histogram shows a bell shaped distribution with a p-value = 0.0639237 > 0.05 which is an indicator for normality.

In addition, the ACF plot of the residuals in Figure 5 shows that for the first 20 lags, all sample autocorrelations fall inside the 95% confidence bounds indicating the residuals appear random. The forecasting evaluation statistics in Table 4 reveals that SARIMA model is appropriate in forecasting tomatoes prices in Kenya.

**Figure 7.** Graphical diagnostics for assessing the SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> model fit.

The selected SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> 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 predicted prices were compared with the observed prices (Table 5). The predicted real tomato prices are close to the observed prices, except for the months of March and April. This result

indicates that the model provides an acceptable fit to predict the tomato prices.

After obtaining satisfactory forecasting results over a short period, the selected SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> model was employed to forecast stream flow over a longer period. Table 5 displays a forecast of the monthly real tomato prices in Nairobi County, Kenya for 2019. The forecasted tomato



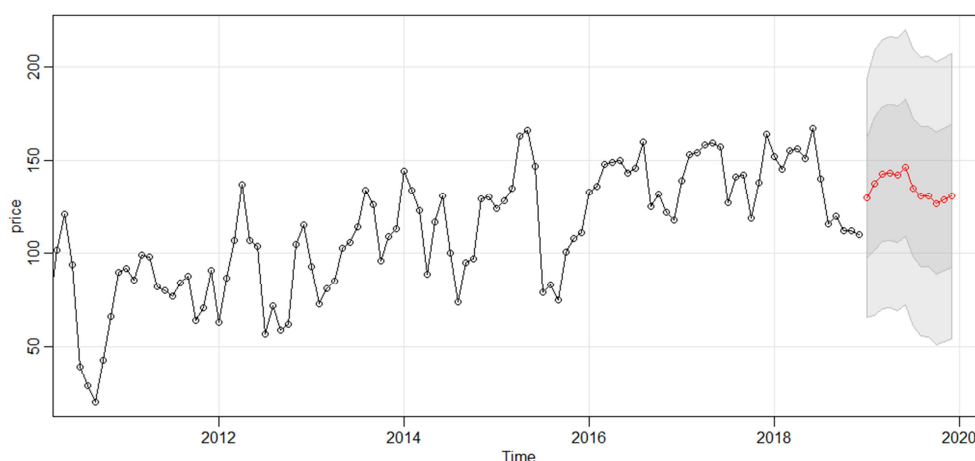
prices were compared with the observed prices in Figure 8. As evident from Figure 8, SARIMA (2, 1, 1) x (1, 0, 1)<sub>12</sub> model is able to capture the flow trend. The forecast series tracks the actual series quite well during the period. The accuracy of this model is calculated based on the MAPE. The outcome shows that the proposed model can forecast the real tomato prices with an accuracy of MAPE value 125.251. MAPE is 125.251%, meaning that the forecasts are off by about 125% on average. The error at the estimation is at an

acceptable level considering the extraordinary factors. Changes of conditions in the market entry of exporting countries immediately impact the tomato export and the prices (entry price, request for active ingredient, etc.).

As shown in Figure 8; the structure of fluctuation of the prices predicted for the following three years is similar to previous years, but it exhibits a more stationary structure. This result indicates that important changes will not occur in tomato prices until the end of 2014 under normal conditions.

**Table 4.** Best-fit models based on different criteria of monthly prices of tomato in Nairobi County Kenya.

Tool	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
SARIMA	1.3491	32.063	22.3	-102.1	125.251	0.7111	-0.007



**Figure 8.** Forecast and Original Observations for the Real Tomato Prices.

**Table 5.** Forecast and Original Observations for the Real Tomato Prices.

Month Year	Point Forecast	Lower 99.5% CI	High 99.5% CI
Jan 2019	128.7222	38.48605	218.9584
Feb 2019	136.3064	36.33723	236.2756
Mar 2019	141.0162	39.55868	242.4737
Apr 2019	141.8015	39.67742	243.9256
May 2019	140.5944	37.7994	243.3895
Jun 2019	144.2263	40.70455	247.7481
Jul 2019	132.889	28.62298	237.155
Aug 2019	128.281	23.27284	233.2892
Sep 2019	128.1427	22.39829	233.8871
Oct 2019	124.2352	17.76022	230.7103
Nov 2019	126.0878	18.88727	233.2883
Dec 2019	128.0736	20.15252	235.9947
Jan 2020	133.0322	18.97371	247.0907
Feb 2020	134.9403	18.36067	251.5198
Mar 2020	137.4491	19.6157	255.2825
Apr 2020	138.0146	19.15192	256.8772
May 2020	137.3084	17.42263	257.1942
Jun 2020	139.6573	18.73875	260.5758
Jul 2020	132.3903	10.44114	254.3395
Aug 2020	129.4349	6.462763	252.4071
Sep 2020	129.3457	5.359189	253.3322
Oct 2020	126.8403	1.847913	251.8328
Nov 2020	128.0281	2.037771	254.0183
Dec 2020	129.3013	2.320955	256.2816

## 5. Conclusion

The results obtained from this study shows that the prices

of tomatoes in Nairobi County have not shown any trends towards an increase or decrease, in other words the series is stationary. The forecasts predicted from SARIMA (2, 1, 1) (10, 1)<sub>12</sub> model which was chosen in order to determine the

course of the prices of the next 2 years show that any significant changes will not occur by the end of 2019. Better price forecast methods, which take into consideration of seasonality, need to be developed to accurately forecast tomato price information. When there are huge price fluctuations, there is a benefit for dissemination of accurate price forecasts among stakeholders to make informed decisions relating to land in production, marketing, trade, and storage. Adoption of cooperative, and contract, farming to shift price risk from farmers to large retailer can stabilize farmer income. Farmers can use forecast price information to hedge their positions by storing tomatoes in cold storage, selling them in other markets where prices are higher, and processing them to make tomato paste, tomato sauce, and ketchup if forecasted prices are too low.

## References

- [1] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74: 427-431.
- [2] Bollerslev, T. (1986). Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31: 307-27.
- [3] Box, G. E. P., Jenkins G. M., Reinsel, G. C. and Ljung, G. M., 2015, *Time Series Analysis: Forecasting and Control (5th Ed.)*, John Wiley & Sons, Inc., Hoboken, New Jersey.
- [4] Zakoian, J. M., 1994. Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, Vol. 18, pp. 931-955.
- [5] Glosten, L. R., Jagannathan, R. and Runkle, D., 1993. On the Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, Vol. XLVIII, No. 5, pp. 1779-1801.
- [6] Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying Covariances, *Econometric Reviews*, 11, 143-173.
- [7] Zhang, J. H., Kong, F. T., Wu, J. Z., Zhu, M. S., Xu, K., and Liu, J. J., 2014. Tomato prices time series prediction model based on wavelet neural network. In *Applied Mechanics and Materials* (Vol. 644, pp. 2636-2640).
- [8] Muhammad Aamir, S., 2017. Price Forecasting Model for Perishable Commodities: A Case of Tomatoes in Punjab, Pakistan.
- [9] Fatima, A., Abid, S. and Naheed, S., 2015. Trends in Wholesale Prices of onion and potato in major markets of Pakistan: A time series Analysis. *Pakistan Journal of Agricultural Research*, 28 (2).
- [10] Reddy, A. A., 2019. Price Forecasting of Tomatoes. *International Journal of Vegetable Science*, 25 (2), pp. 176-184.
- [11] Gathondu, E. K. (2014). *Modelling of wholesale prices for selected vegetables using time series models in Kenya* (Masters thesis, University of Nairobi).
- [12] Ruttoh J K, Bett E K, and Nyairo N (2018). *Empirical analysis of structure and conduct of tomato marketing in Loitoktok, Kajiado County, Kenya*. *International Journal of Agricultural Extension and Rural Development* ISSN 3254-5428 Vol. 6 (4), pp. 628-638, April, 2018.
- [13] Geoffrey, S. K., Hillary, N. K., Kibe, M. A., Mariam, M., & Mary, M. C. (2014). Challenges and strategies to improve tomato competitiveness along the tomato value chain in Kenya. *International Journal of Business and Management*, 9 (9), 205.
- [14] Assis, K., Amran, A., & Remali, Y. (2010). Forecasting cocoa bean prices using univariate time series models. *Researchers World*, 1 (1), 71.
- [15] Fenyves, V., Orban, I., Dajnoki, K., & Nabradi, A. (2010). Evaluation of different predicting methods in forecasting Hungarian, Italian, and Greek lamb prices. *Food Economics–Acta Agricult Scand C*, 7 (2-4), 192-196.
- [16] Dieng, A. (2008). Alternative forecasting techniques for vegetable prices in Senegal. *Revue sénégalais de recherches agricoles et agroalimentalress*, 1 (3), 5-10.
- [17] Sampson, W., Suleman, N., & Gifty, A. Y. (2013). Proposed seasonal autoregressive integrated moving average model for forecasting rainfall pattern in the Navrongo Municipality, Ghana. *Journal of Environment and Earth Science*, 3 (12), 80-85.
- [18] Ivanišević, D., Mutavdžić, B., Novković, N., & Vukelić, N. (2015). Analysis and prediction of tomato price in Serbia. *Economics of Agriculture*, 62 (4), 951-962.
- [19] Boateng, F. O., Amoah-Mensah, J., Anokye, M., Osei, L., & Dzebre, P. (2017). Modeling of tomato prices in Ashanti region, Ghana, using seasonal autoregressive integrated moving average model. *Journal of Advances in Mathematics and Computer Science*, 1-13.
- [20] Adanacioglu, H., & Yercan, M. (2012). An analysis of tomato prices at wholesale level in Turkey: an application of SARIMA model. *Custos e@ gronégocio on line*, 8 (4), 52-75.
- [21] Zhang, H., & Rudholm, N. (2013). Modeling and forecasting regional GDP in Sweden using autoregressive models. *Dalama University*.
- [22] Box, G. E. P. and Jenkins, G. M. (2016). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- [23] Bhardwaj, S. P., Paul, R. K., Singh, D. R., & Singh, K. N. (2014). An empirical investigation of ARIMA and GARCH models in Agricultural Price Forecasting. *Economic Affairs*, 59 (3), 415.
- [24] Cryer, J. D., & Chan, K. S. (2008). Time series regression models. *Time series analysis: with applications in R*, 249-276.