

Modeling of Tomato Prices in Ashanti Region, Ghana, Using Seasonal Autoregressive Integrated Moving Average Model

ABSTRACT

The pricing of seasonal and perishable crops such as tomatoes is of paramount concern to emerging economies. In this paper, we have formulated a model for the prices of tomatoes in the Ashanti Region of Ghana. We applied time series on the tomatoes price data recorded over the period of 1994 to 2015, sourced from the Ministry of Food and Agriculture (MoFA) Research Directorate, Kumasi. We analyzed the data using Seasonal Autoregressive Integrated Moving Average (SARIMA) by regrouping the data into quarters for each year. Consequently, the best fitted model was found to be SARIMA (2, 1, 1). The model predicted that, in 2016 the prices of tomatoes will decline in all quarters, which is validated by first quarter price of tomatoes in 2016. Comparatively, prices of tomatoes in the same quarter of 2015 and 2014 were relatively high. We have formulated this model to assist stakeholders including government to take informed decisions and formulate pricing policies that will guarantee farmers stable price.

Keywords: Time series, Price of tomatoes and SARIMA models.

1 INTRODUCTION

Tomato is a fruit of the plant whose botanical name is *Lycopersicon esculentum* and is a member of the nightshade family. It is consumed in diverse ways including raw, as ingredient in dishes, sauces and in drinks because it is classified as vegetable (Alam *et al.*, 2007). Tomatoes and tomato-based foods provide a convenient matrix by which nutrients and other health-related food components are supplied to the body (Beecher, 1998). In Ghana, tomatoes form a very important component of food consumed as many dishes have them as part of the ingredients (Tambo and Gbemu, 2010). It is also confirmed that Ghana is second to Germany in terms of tomatoes consumption of an average of twenty five thousand (25,000) tones per year (Aryeetey, 2006).

In spite of the modest levels of beta-carotene and gamma-carotene in tomato products, due to their pro-vitamin A activity, a high consumption of the vegetable and its products results in a rich supply of vitamin A in the body. Lycopene, an antioxidant, purportedly fights the free radicals that interfere with normal cell growth and activity.

Tomato has valuable vitamins, for instance vitamin A and C, and also it contains fibers, and known to be cholesterol-free. Generally, the average size of tomato equal to 148g boasts only 35 calories. Approximately 20 – 50 mg of lycopene/100g of fruit weight can be found in tomato. Tomatoes are utilized at a higher rate in the developed countries than in the developing countries and hence it may be referred to as a luxury crop (Siddiqui and Dhua, 2010). It is a warm season crop with many varieties from which with special production practices, one can produce tomatoes in 60 days by choosing the right varieties. Choice of variety is also informed by the orderly production, earliness, shipping quality, and disease resistance, and usually requires large investment in time, labour and increase in intensity of management but rewarded by increased yields and profits (Robinson and Kolavalli, 2010).

1.1 Harvesting and Storage Challenges

The principles that dictate at which stage of maturity the fruit should be harvested are crucial to its subsequent storage (Siddiqui and Dhua, 2010). Maturation is indicative of the fruit being ready for harvest (FAO, 2008), while harvesting marks end of the growth cycle of tomatoes and the beginning of series of stages of activities that ensure that the consumer gets the vegetable in the preferred state and at the best of desired quality. Harvesting of fresh-market tomatoes is labour intensive and requires multiple pickings (Orzolek *et. al.*, 2006). Tomatoes for the wholesale market should usually be picked at the mature green to breaker stage to prevent the fruit from becoming overripe during long transportation and handling .It is recommended to leave tomatoes on the vine to become ripe before harvesting, if they can be sent to market quickly and in good condition. Usually, fresh market tomatoes are harvested by hand by a lot of growers using plastic buckets, and these buckets should be smooth without any sharp edges to damage the produce (Kitinoja, 2008, Orzolek *et. al.*, 2006).

The riper the tomato, the more susceptible it is to bruising and so good harvesting management technique should be applied to avoid losses (Hurst, 2010). Thus the harvest crews should carefully place fruits into picking containers instead of dropping them since dropping of more than 6 inches onto a hard surface can cause internal bruising which is evident until after the tomato is cut open.

In Ghana, the focus of the various stakeholders in the tomato industry has mostly been on improved production capacities of farmers. However, after investing so heavily in producing the vegetables, farmers produce are lost in the postharvest chain. In this sector, production seasonality, the dominance of rain fed agriculture, high perishability of the vegetable, lack of ready market, lack of a reasonable alternative uses of the vegetable and poor pricing are some problems faced by farmers (Robinson and Kolavalli, 2010).

It is also possible that, poor postharvest practices coupled with poor storage facilities account for the recurrent seasonal postharvest losses of tomatoes. Postharvest losses of tomatoes could also be attributed to traders' lack of appropriate postharvest skills of prolonging the shelf life of tomatoes in commercial quantities, unavailability of large scale processing factories in tomato producing districts to preserve the surplus produce for future use (Aryetey, 2006).

1.2 Marketing of Tomatoes in Ghana

Unlike grains, the bulky and perishable nature of tomato, coupled with the lack of processing and storage facilities in Ghana, constitute critical impediments to the commodity's inter-market arbitrage by traders (Issahaku, 2012; Amikuzuno, 2009). Thus marketing of tomatoes has been a problem in recent times due to inadequate storage facilities, the perishable nature of tomatoes, inadequate funds, high cost of inputs, and high level of importation from other countries (Robinson and Kolavalli, 2010). Farmers get frustrated as a result of low prices of their product or lack of access to market opportunities and as results are unstable to pay back loans owed financial institutions. It appears that farmers are short-changed by the system and are not making any meaningful profit from tomatoes production and marketing. Market queens have taken undue advantage of the perishable nature of the commodity to decide price

for fresh tomato farmers and control the market for almost all fresh vegetables in Ghana (Kokutse, 2010).

The queens generally have the upper hand over the farmers they buy from. They do not hesitate to put extreme pressure on the farmers, who accuse them of cheating and pocketing the lion's share of profits. A programme's coordinator of the Peasant Farmers Association of Ghana notes that these tomato queens have been able to organize themselves into a sort of cartel and have used that to control the pricing of the produce to the disadvantage of the farmers. The market queens use the perishable nature of the produce to their advantage, forcing farmers to accept low prices so as not to lose out completely. What they do is that they wait till late afternoon after the farmers have packed their produce before they go to the farms and simply refuse to pay what the farmers ask for (Kokutse, 2010).

Considering that Ghana experiences annual gluts during the major seasons and when a high percentage of the harvested produce is lost (Addo, *et al* 2015), there is the need to model the price trend as frequent postharvest losses of tomatoes have dire consequences on farmers and consumers. As a result of these problems the study is intended to construct and interpret mathematical model that fits tomato prices using time series and then use the model to predict possible future prices of tomatoes. It is believed that farmers would use the model to be developed to take future decision concerning prevailing market price for tomatoes and respond appropriately with supply level to match with to avoid waste.

2 MATERIALS AND METHODS

This study focuses on estimating the future price of tomatoes in Ghana. To achieve this, we applied time series on the tomatoes price data recorded over the period of 1994 to 2015, which was obtained from the Ministry of Food and Agriculture (MoFA) Research Directorate, Kumasi. We modeled the data using Seasonal Autoregressive Integrated Moving Average (SARIMA) by regrouping the data into quarters for each year. SARIMA was chosen due to its suitability for modeling cases where we suspect a seasonal pattern in our data. In particular, tomatoes production in Ghana is seasonal and so SARIMA is appropriate to use.

2.1 Seasonal ARIMA Model

Seasonality usually causes the series to be non-stationary because the average values at some particular times within the seasonal span (months, for example) may be different than the average values at other times. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is $ARIMA(p, d, q) \times (P, D, Q)S$, with 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 could be written more formally as

$$\Phi(B^S)\phi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t$$

The non-seasonal components are:

131
$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$$

132
$$\text{MA: } \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

133 The seasonal components are:

134
$$\text{AR: } \Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}$$

135
$$\text{MA: } \Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS}$$

136 **Seasonal differencing** is defined as a difference between a value and a value with lag that is
137 a multiple of S .

138 For $S = 12$, which usually occurs with monthly data, has a seasonal difference of

139
$$(1-B^{12})x_t = x_t - x_{t-12}.$$

140 The differences (from the previous year) may be about the same for each month of the year
141 giving us a stationary series.

142 For $S = 4$, which occurs with quarterly data, also has a seasonal difference given by

143
$$(1-B^4)x_t = x_t - x_{t-4}.$$

144 Seasonal differencing removes seasonal trend and can also get rid of a seasonal random walk
145 type of non-stationarity.

146 **Non-seasonal differencing:** If trend is present in the data, we may also need non-seasonal
147 differencing. Often a first non-seasonal difference will “detrend” the data using $(1-B)x_t = x_t -$
148 x_{t-1} in the presence of trend (Penstate, 2012).

149 A crucial step in an appropriate model selection is the determination of optimal model
150 parameters. One criterion is that the sample ACF and PACF, calculated from the training data
151 should match with the corresponding theoretical or actual values (Hipel and McLeod, 1994;
152 Kihoro *et al.*, 2004). Other widely used measures for model identification are Akaike
153 Information Criterion (AIC) and Bayesian Information Criterion (BIC) which are defined
154 below (Faraway and Chatfield, 1998; Kihoro *et al.*, 2004):

155
$$AIC(p) = n \ln(\hat{\sigma}_e^2/n) + 2p$$

156
$$BIC(p) = n \ln(\hat{\sigma}_e^2/n) + p + p \ln(n)$$

157

158 Here n is the number of effective observations, used to fit the model, p is the number
159 of parameters in the model and $\hat{\sigma}_e^2$ is the sum of sample squared residuals. The optimal
160 model order is chosen by the number of model parameters, which minimizes either AIC or
161 BIC. Other similar criteria have also been proposed in literature for optimal model
162 identification.

3 FINDINGS

Preliminary analysis was performed on quarterly tomatoes price data (Appendix) recorded from 1994 to 2015 to remove any possible errors before the data was individually entered to be statistically analyzed using SPSS software. Box Jenkins ARIMA modeling procedure was applied to the raw data and time series plot was done to assess the stationarity of the data which provided the following results:

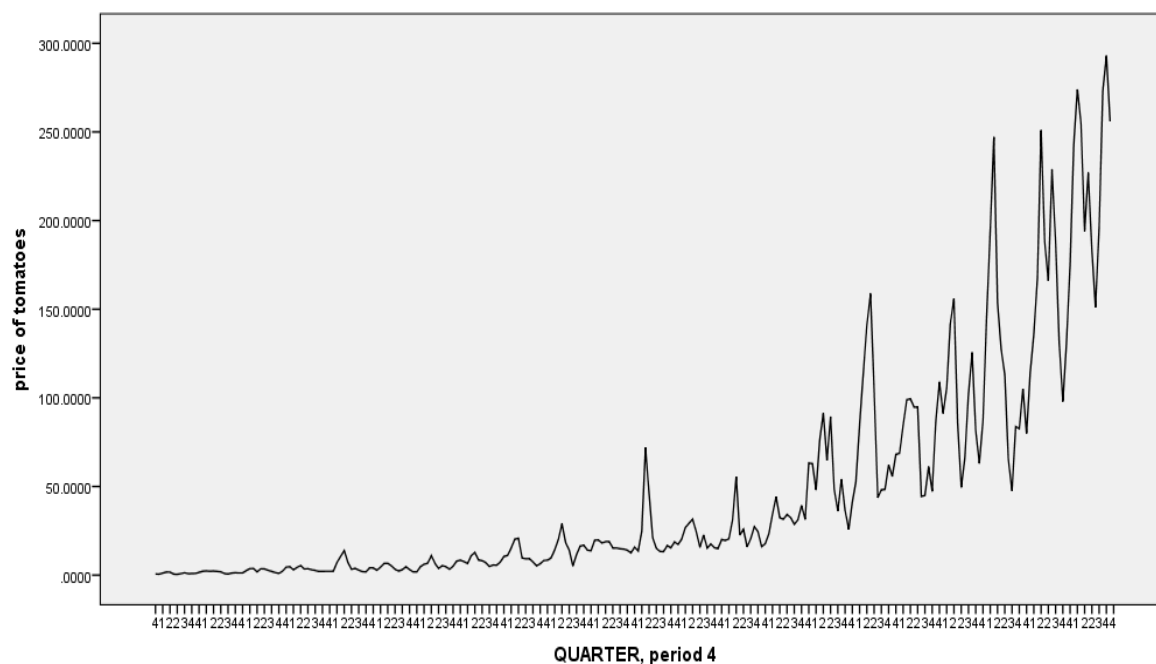


Figure 1: Price of tomatoes from 1994 to 2015

The price of tomatoes displayed in figure 1, indicates that the data are non stationary with seasonality. Thus the series does not exhibit a constant mean and variance.

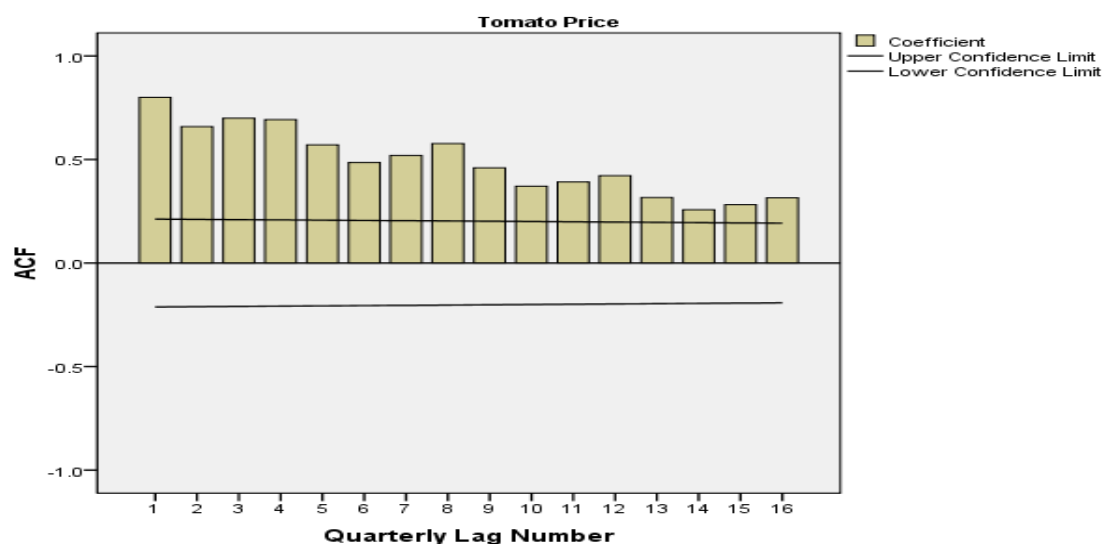


Figure 2: ACF plot of price of tomatoes

The autocorrelation function (ACF) plot of tomatoes price (see figure 2) ,also indicates that there are seasonality pattern which needs to be removed in order for the data to be stationary. Seasonality usually causes the series to be non-stationary because the average values at some particular times within the seasonal span are different from the average values at the other times.

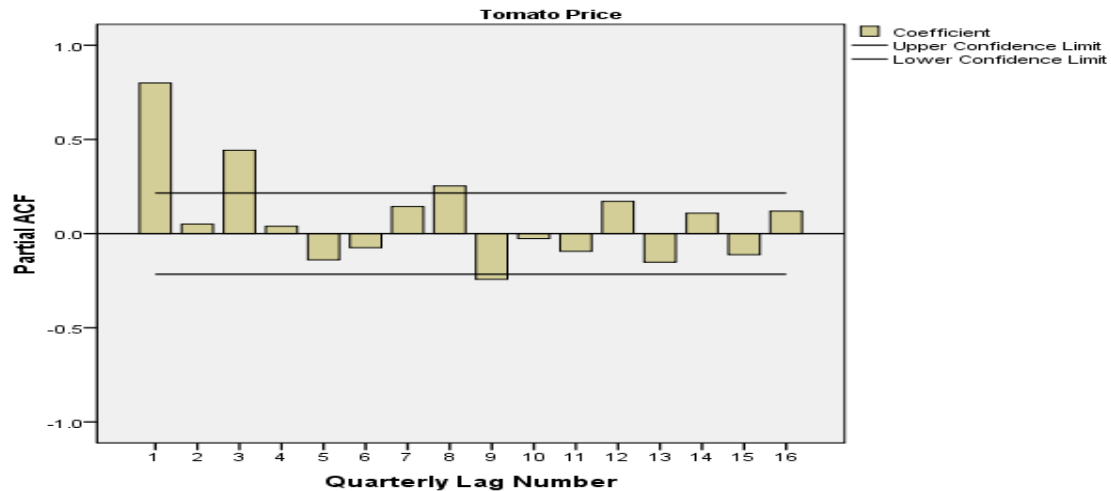


Figure 3: The PACF plot of price of tomatoes

The spikes (both positive and negatives) in the above partial autocorrelation function (PACF) autocorrelation function (ACF) plots of tomatoes price shown in figure 3, suggest the data is non-stationary with seasonality.

3.1 Data Stationarity Check

Statistically, we can determine non-stationarity of tomatoes price data by the Augmented Dickey fuller (ADF) test of the data as shown in table 1.

Table 1: Test for stationarity of the original series

Augmented dickey fuller test of stationarity		
Difference	<i>p</i> -value	test statistic
0	0.99	5.2267

From table 1, it can be seen that *p*-value (0.99) is far greater than the alpha level of (0.05, 0.01) indicating that the series is non-stationary. Therefore, the series in order to attain stationarity need to be seasonally differenced.

3.1.1 Data Differencing

There are trend and seasonal components as observed by the decomposition plot of the series, and therefore seasonal differencing could be applied to the price series and re-assess its stationarity status.

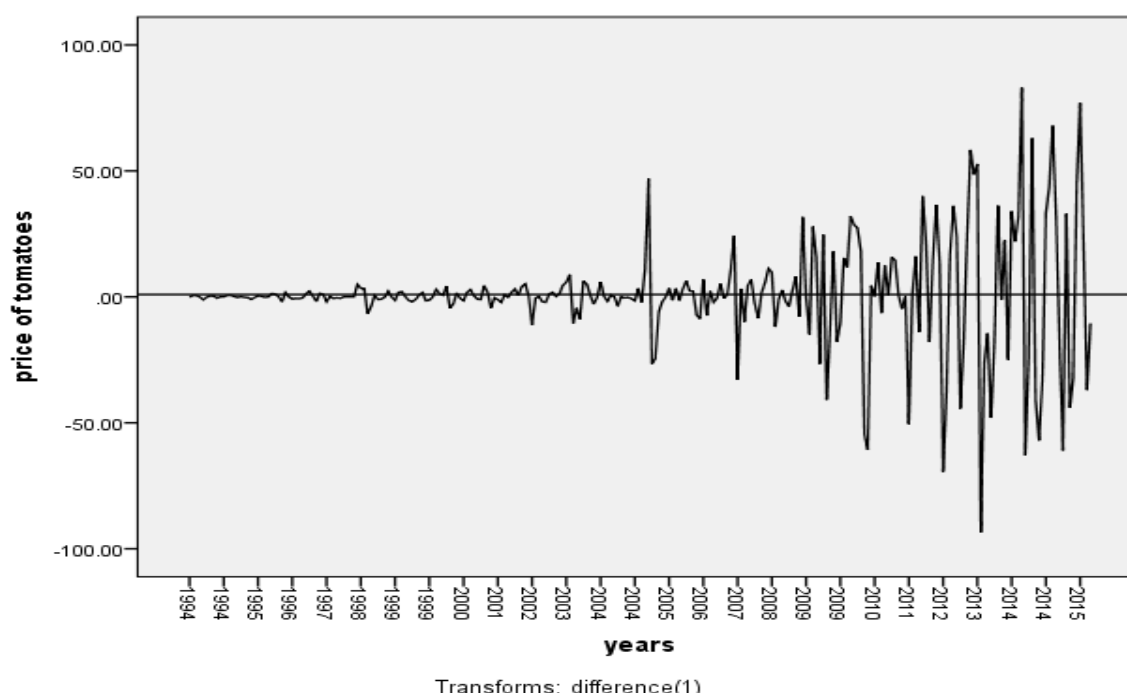


Figure 4: The Seasonal differenced of the Series

Figure 4, above indicates that the series is now stationary after the first order seasonal differencing even though it has not been statistically tested. Stationarity test is done using Augmented Dickey fuller (ADF) unit root test.

3.1.2 Stationarity Confirmation by ADF test

Augmented Dickey-Fuller Unit Root Test was done on the entire tomatoes price data after first order differencing, and results of the test displayed as follows:

Table 2: Test of stationarity by ADF test.

Augmented dickey fuller test of stationarity			
Difference	p-value	test statistic	lags
1	0.01	-0.84	24
0	0.99	5.23	24

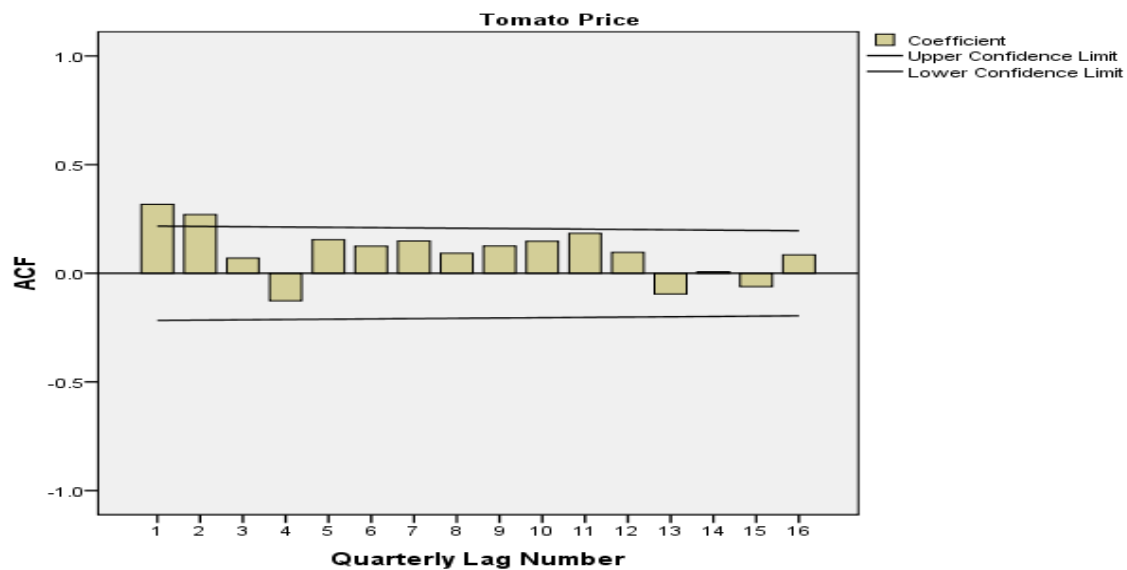
From table 2, it can be seen that the p-value (0.01) of the test is far less than the alpha level of (0.05) indicating the series is stationary after the first order differencing.

3.2 Model Identification

Now, we aim at finding an appropriate seasonal ARIMA model that best fit the tomato price data based on the ACF and PACF plots. Thus, we tentatively identify the numbers of AR and/or MA terms that are needed to correct any autocorrelation left in the series by looking at the ACF and PACF plots of the seasonal differenced series.

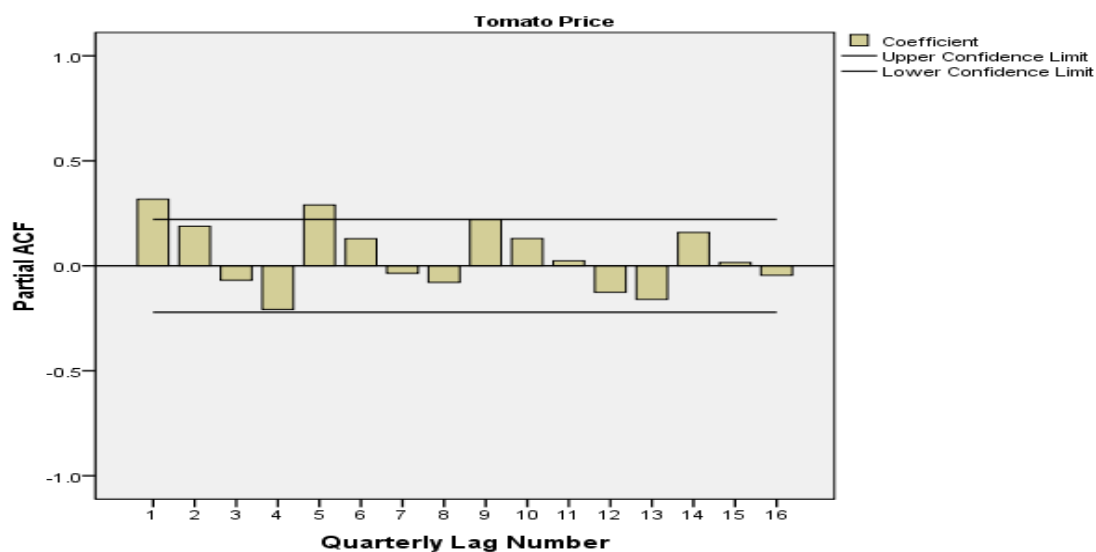
3.2.1 Identification of the Order of AR/MA in the Model

232 After the tomatoes price data has been stationarized by first order seasonal differencing, the
 233 next step in fitting the proposed SARIMA model is to determine whether AR or MA terms
 234 are needed to correct any autocorrelation that remains in the seasonal differenced tomatoes
 235 price series.



236
 237
 238

Figure 5: The ACF of Order One Seasonal Differenced Series



239
 240

Figure6: The PACF of the Seasonal Differenced Series

241 Based on the ACF and PACF correlograms shown in both figure 5 and figure 6 and the
 242 tentative SARIMA model, would consist of the combination of non-seasonal and seasonal
 243 differencing of order one.

244
 245

246 **Table 3: Tentative SARIMA Models**

MODELS	BIC
SARIMA(0,1,1)x(0,1,1)	5.969
SARIMA(1,0,1)x(0,1,1)	6.086
SARIMA(1,0,0)x(0,1,1)	6.198
SARIMA(0,0,1)x(0,1,1)	6.296

247
 248 From table 3, we select SARIMA (0,1,1)x(0,1,1) as the best model for forecasting tomato
 249 price in Ashanti Region, Ghana based on the fact its BIC(5.969) value is smallest among all
 250 the other values as criteria for model selection.

Table 4. SARIMA (0, 1, 1)x(0,1,1) Model Parameter Estimates

			Estimate	SE	t	Sig.
Tomato	No	Difference	1			
PriceModel_1	Transformation	MA Lag 1	0.705	0.089	7.942	0.000
		Seasonal Difference	1			
		MA, Lag 1	0.627	0.116	5.404	0.000
		Seasonal				

251 Table 4, contains the estimates of the model parameters. The p -values for the two (2) MA
 252 estimates are all 0.000 which indicate that the parameters are statistically significant and that
 253 the selected SARIMA model fit the tomato price data and is good for forecasting.

254

255 3.2.2 Model Estimation

256 The forecasting equation of the model can be written as

$$257 \hat{y}_t = y_{t-1} + y_{t-s} - y_{t-s-1} - \phi_1 e_{t-s} - \phi_1 e_{t-1} + \phi_1 \phi_1 e_{t-s-1}$$

$$258 \hat{y}_t = y_{t-1} + y_{t-4} - y_{t-5} - 0.627 e_{t-4} - 0.705 e_{t-1} + 0.442 e_{t-5}$$

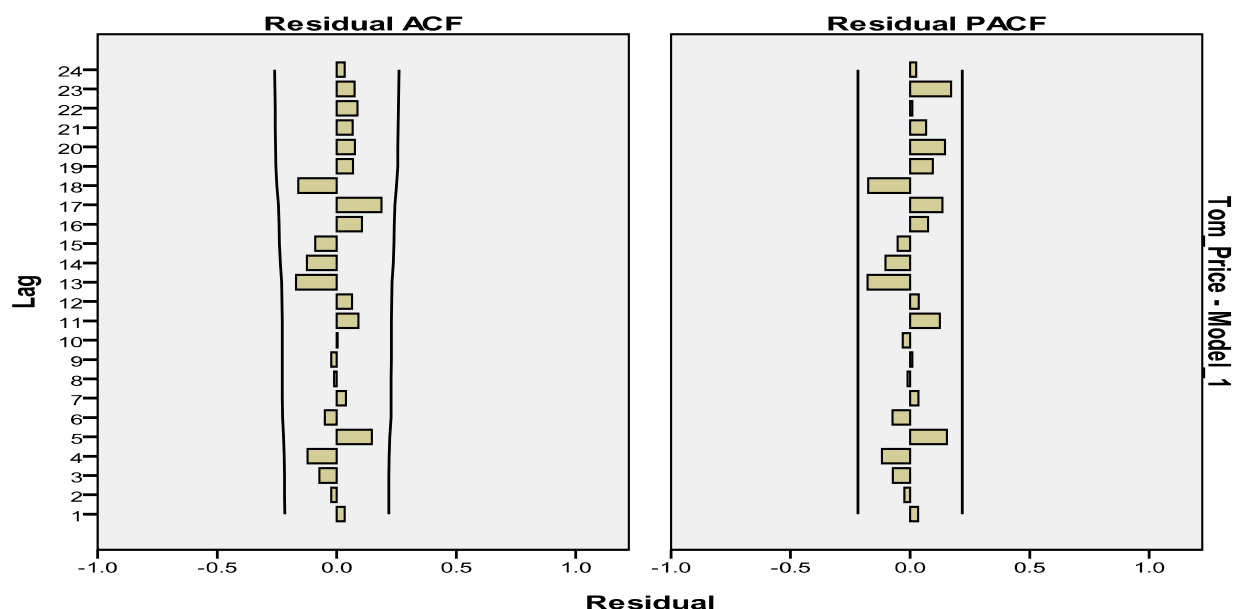
259 where ϕ_1 is the MA(1) (non-seasonal) coefficient and ϕ_1 is the SMA(1) (seasonal) coefficient.
 260 This seasonal forecasting model also consists of the adding multiples of the errors at lags 1, 4,
 261 and 5. The coefficient of the lag-5 error is the product of the MA(1) and SMA(1) coefficients.
 262 The model also effectively applies exponential smoothing to seasonality which makes it very
 263 good for making long-term forecasts.

264

265 3.3 Check of Model Fitness

266 We checked the model fitness using ACF and PACF plots, distribution of residuals by Ljung
 267 Box-Test plot. Ljung Box-Test is a type of statistical test that assesses whether any of a group
 268 of autocorrelations of a time series are different from zero. The following graphs of tomato
 269 price residual plots are presented to evaluate the autocorrelations of the selected model.

270



271

272 **Figure 7: Standardized Residual, ACF and Ljung Box-Test statistic plot of the model.**

273 From figure 7 above, we found that spikes are all falling within the significance limits which
 274 indicates that the model residual (errors) not auto-correlated and therefore the fitted model is
 275 valid and could be used for making forecast. Thus the residuals are normally and
 276 independently distributed.

277

278 3.4 Forecast

279

280 The model we selected was estimated using the in-sample tomato price series and
 281 realistically, to test ability of the model, one has to compare its forecasting performance
 282 against out-of-sample data. Thus, we have to withhold some of the sample data from the
 283 model identification and estimation process, then use the model to make predictions for the
 284 hold-out data in order to see how accurate they are and to determine whether the statistics of
 285 their errors.

286

Table 5. Forecast from 2015 to 2016 by SARIMA(2,1,1)x(0,1,1)

Model		Q3 2015	Q4 2015	Q1 2016	Q2 2016	Q3 2016	Q4 2016
Tom_Price- Model_1	Forecast	189.46	169.76	228.21	261.79	222.12	202.84
	Actual	176.67	274.00	00.00	000.0	000.00	000.00
	Difference	6.91	6.91	45.79	000.00	000.00	000.00

The forecast values of the selected model are displayed in table 5. We found out that the difference between the actual data values and the estimated (forecast) values from the selected model are very small. This indicates that the forecast values are valid and could reliably be used in making decisions.

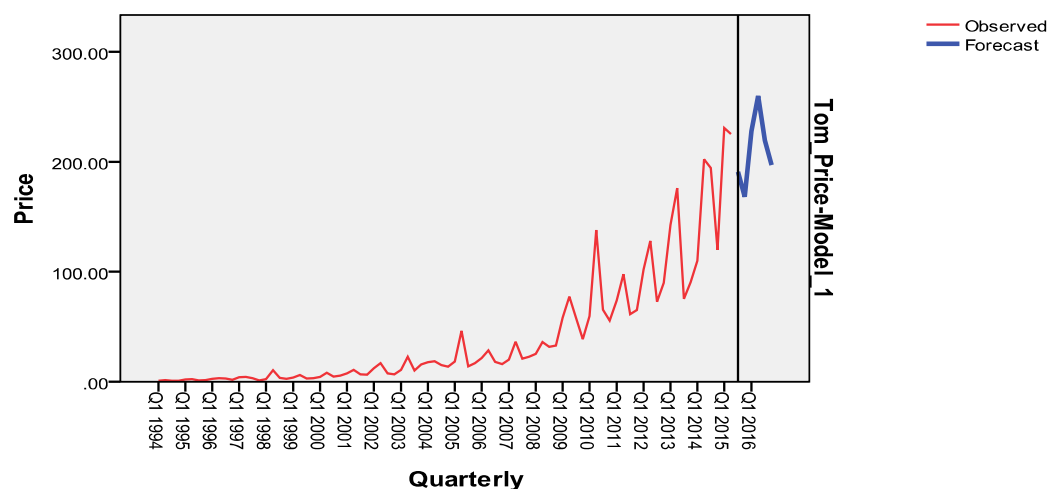


Figure 8: Forecast Trend of Tomato Prices for 2016

From the forecast plot above (Figure 8), it follows that price of tomatoes in the region will decline throughout the year.

4 DISCUSSIONS AND CONCLUSIONS

In the study, we have formulate model for prices of tomatoes in the Ashanti Region using the technique of time series based on tomato price data recorded over time from 1994 to 2015, from Ministry of Food and Agriculture (MoFA) Research Directorate Kumasi, Ashanti Region. The monthly data was regrouped into quarters for each year and was subsequently analyzed using Seasonal Autoregressive Integrated Moving Average (SARIMA). From the analysis, the best fitted model was found to be SARIMA (0, 1, 1)x(0,1,1).

The model predicted that in 2016, the prices of tomatoes will decline in all quarters all things being equal. Thus the price of tomatoes in the first quarter for instance in the year 2015 was GH¢ 230.67 compare to GH¢140.33 in the same period in 2016 and continues in same manner for subsequent quarters.

However, in 2014, there was an increase in prices from the first quarter to the second quarter, while a decrease was seen from the second quarter to the fourth quarter.

We believe that stakeholders would use the model developed in this study to take decisions. We also recommended that Government would use it and come upwith price control policies that will guarantee farmers a stable price. Due to the fact that tomatoes are perishable products, producing more when the prices are low, means that farmers will be forced to sell at a lower price which will at the end reduce their revenue. Therefore, Government is also advised to set up processing industry. Consumers of tomatoes should also buy more of the products during this period due to the fact that prices are expected to decrease.

REFERENCES

- Addo, J. K., Osei, M. K., Mochiah, M. B., Bonsu, K. O., Choi, H. S. and Kim, J. G. (2015). Assessment of Farmer Level Postharvest Losses Along The Tomato Value Chain In Three Agro-Ecological Zones Of Ghana. *International Journal of Research in Agriculture and Food Sciences*. Vol. 2, No.9
- Alam, T., Tanweer, G. and Goyal, G. K. (2007). Stewart Postharvest Review, Packaging and storage of tomato puree and paste. Research article, Volume 3, Number 5, October 2007, pp. 1-8(8). Publisher: Stewart Postharvest Solutions. DOI: 10.2212/spr.2007.5.1
- Amikuzuno, J. (2009). Spatial Price Transmission and Market Integration between Fresh Tomato Markets in Ghana: Any Benefits from Trade Liberalisation? Georg-August-Universität Göttingen, Agricultural Economics and Rural Sociology, Germany. <http://www.tropentag.de/2009/proceedings/node368.html>. Retrieved September, 12, 2016.
- Aryeetey, E. (2006). ISSER-Merchant Bank Development Seminar Series. Ghanaweb. <http://www.ghanaweb.com/GhanaHomePage/election2008/artikel.php?ID=101256>.
- Beecher, B. R. (1998). Nutrient content of tomatoes and tomato products. National Center for Biotechnology Information, U.S. National Library of Medicine. (Online) <http://www.ncbi.nlm.nih.gov/pubmed/9605204>. Accessed: March 3, 2016.
- Faraway, J. and Chatfield. C. (1998). "Time series forecasting with neural networks: a comparative study using the airline data", *Applied Statistics* 47, pages: 231–250.
- Food and Agriculture Organization (FAO), (2008). FAOSTAT. <http://faostat.fao.org>. (Online) Accessed: April 27, 2016.
- Hipel, K. W. and McLeod, A. I. (1994). "Time Series Modelling of Water Resources and Environmental Systems", Amsterdam, Elsevier.
- Hurst, W. C. (2010). Harvest, Handling and Sanitation Commercial Tomato Production Handbook B1312. CAES Publications. Univ. of Georgia. Accessed: April 2, 2016. Available http://www.caes.uga.edu/publications/pubDetail.cfm?k_id=7470.
- Kihoro, J. M., Otieno, R. O. and Wafula, C. (2004). "Seasonal Time Series Forecasting: A Comparative Study of ARIMA and ANN Models", *African Journal of Science and Technology (AJST) Science and Engineering Series* Vol. 5, No. 2, pages: 41-49.
- Kitinoja, L. (2008). Causes and Sources of Postharvest Problems. Postharvest Training CD Rom\Sample Presentations. Pp 1 -19
- Kokutse, F. (2010). Tomato Queens Short-Change Farmers. <http://ipsnews.net/africa>, Retrieved August 17, 2010.
- Issahaku, H. (2012). An Analysis of the Constraints in the Tomato Value Chain. *International Journal of Business and Management Tomorrow*, vol. 2(10). pp 1-8.

Orzolek, M. D., Bogash, M. S., Harsh, M. R., Lynn, F., Kime, L. F., Jayson, K., Harper, J. K. (2006). Tomato Production. Agricultural Alternatives Pub. Code # UA291. Pp. 2-3.

Penstate, (2012). *Applied Time Series Analysis: Seasonal ARIMA Models*. Retrieved October 12, 2010, from <http://onlinecourses.science.psu.edu>.

Robinson, J. Z. E. and Kolavalli, S. L. (2010). The Case of Tomato in Ghana: Productivity, Development and Strategy Governance Division, IFPRI, Ghana GSSP Working Paper No. 19

Siddiqui, M. W., and Dhua, R. S. (2010). Eating artificially ripened fruits is harmful. *Current Science*, 99(12), pp 1664 –1668.

Tambo, J. A. and Gbemu, T. (2010). Resource-use Efficiency in Tomato Production in the Dangme West District, Ghana. Conference on International Research on Food Security, Natural Resource Management and Rural Development. Tropentag, ETH Zurich.

APPENDIX

Table A1: Quarterly Price of Tomatoes in Ashanti Region, Ghana

YEAR	QTR 1	QTR 2	QTR 3	QTR 4
1994	0.18	1.34	0.88	0.9
1995	2.07	2.21	1.14	1.27
1996	2.48	3.05	2.79	1.55
1997	4.11	4.43	3.12	0
1998	2.37	10.53	3.33	2.66
1999	3.79	6.18	2.89	3.25
2000	4.25	8.13	4.63	5.43
2001	7.53	10.79	6.69	6.24
2002	12.31	16.94	8.64	6.65
2003	10.82	22.54	10.2	15.8
2004	17.77	18.65	15.16	13.75
2005	18.17	46.25	13.91	16.96
2006	21.54	28.41	17.86	15.98
2007	20.09	36.45	20.79	22.67
2008	25.22	36.05	31.79	33.97
2009	58.05	77.31	58.01	38.78
2010	59.71	137.78	65.42	55.42
2011	73.81	97.72	61.42	65.22
2012	101.97	128.12	72.64	90
2013	142.52	176.08	75.2	90.47
2014	110	202.33	194.33	120
2015	230.67	225.33	176.67	274