

Original Research Article

<https://doi.org/10.20546/ijcmas.2020.910.368>

Forecasting of Vegetable Arrivals and Wholesale Prices of Koyambedu Market in Chennai Using SARIMA Approach

C. Tamilselvi* and G. Mohan Naidu

Department of Statistics and Computer Applications, S.V. Agricultural College,
Tirupati (A.P), India

*Corresponding author

ABSTRACT

Forecasting of vegetable arrivals and prices gain more significant because of its perishability and seasonality. Adequate information on prices and arrivals would bridge the gap between production and marketing. The secondary data regarding monthly arrivals (qtls) and prices (Rs/qrtl) of tomato and onion were collected for a period of 8 years (2011-2018). Forecasting was done using SARIMA models for next 6 months. SARIMA (3, 0, 3) (3, 0, 2)₁₂ and (2, 2, 2) (1, 0, 1)₁₂ were the best model for forecasting the tomato arrivals and prices while SARIMA (1, 2, 3) (1, 2, 2)₁₂ and (2, 0, 3) (2, 0, 3)₁₂ were the best model for forecasting the onion arrivals and prices in Koyambedu (Chennai) market respectively. The actual and predicted values of vegetable arrivals and prices were similar to each other except tomato arrivals. Tomato arrivals were high in the month of February (106369 qtls) whereas peak prices (Rs 1966/qrtl) were in the month of January. The peaks of onion arrivals (162534 qtls) were noticed in the Month of January whereas prices (Rs 4211 /qrtl) were high in June among the forecasted values.

Keywords

Tomato, Onion,
SARIMA, ACF,
PACF, \bar{R}^2 , RMSE

Article Info

Accepted:
24 September 2020
Available Online:
10 October 2020

Introduction

In India, the agriculture sector includes forestry and fishing is the main stay of the economy, contributing about 17% of the GDP and accounting for about 53% of employments. Under the purview of agriculture and allied activities in agriculture has become 33%. The share of plan outlay for horticulture which was 3.9% during the IX

plan, has increased to 4.6% during the XII plan. India has witnessed increase in horticulture production over the last few years. Significant progress has been made in area expansion resulting in higher production. Over the last decade, the area under horticulture grew by 2.6% per annum and annual production increased by 4.8%. During 2017-18, the production of horticulture crops was 311.71 Million tonnes from an area of

25.4 million hectares. The production of vegetables has increased from 101.2 million tonnes to 184.40 million tonnes since 2004-05 to 2017-18 and production of fruits has increased from 50.9 million tonnes to 97.35 million tonnes since 2004 to 2017-18 (Horticulture Statistics at a Glance, 2018).

The aim of this paper is to analyse the arrivals and price fluctuations of tomato and onion in Koyambedu market of Chennai using Seasonal Auto Regressive Moving Average (SARIMA) approach as the analysis tool. Vegetable arrivals and 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 analyse tool can give an early warning message of tomato and onion price fluctuation in the future.

Sampson *et al.*, (2013) argued that among the seasonal decomposition models of forecasting, the SARIMA models could enable to forecast the price of tomatoes in Turkey and found SARIMA (1, 0, 0) (1, 1, 1)₁₂ model as the most suitable. They reported that the highest tomato prices seasonality adjusted were in October. Gathondu (2014) 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 Heterostadicity (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.*, (2015) analysed the changes and future tendencies of the price of tomatoes with descriptive statistics and found that the

ARIMA models were suitable for price forecasting. Mohan Naidu and Srikala (2015) Seasonal indices were observed that generally the prices were low from January to May and it raise from June, and reaches the maximum in November. Based on highest coefficient of multiple determination (R^2) and Mean Absolute Percentage Error (MAPE), the estimated best model was Seasonal Auto Regressive Integrated Moving Average (SARIMA) (1,1,1) (1,0,0)₁₂ short term forecasts based on this model were close to the observed values.

Boateng *et al.*, (2017) found that the predictability of the model increased with 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.

Materials and Methods

Database

The secondary data pertaining monthly arrivals (qtls) and wholesale prices (Rs/qrtl) of tomato and onion were collected from various sources like Market Management committee – CMDA and Agmarknet for a period of 8 years (i.e. 2011-2018).

Methodology

SARIMA models are an adaptation of Autoregressive Integrated Moving Average (ARIMA) models to specifically fit seasonal time series. The seasonal ARIMA model incorporates non-seasonal and seasonal factors in a multiplicative model and is

denoted as:

ARIMA (p, d, q)x(P, D, Q)_s

p = order of non- seasonal Auto Regressive (AR) order, d = order of non- seasonal difference, q = order of non- seasonal Moving Average (MA), P = order of seasonal Auto Regressive (SAR), D = order of seasonal difference, Q = order of seasonal Moving Average (SMA).

Without differencing operations, the model can be written as:

$$\Phi(BS)\varphi(B)(x_t - \mu) = \Theta(BS)\theta(B)wt(1)$$

The non-seasonal components are:

$$AR: \varphi(B) = 1 - \phi_1B - \dots - \phi_pB_p \quad (2)$$

$$MA: \theta(B) = 1 + \theta_1B + \dots + \theta_qB_q \quad (3)$$

The seasonal components are:

$$Seasonal\ AR: \Phi(BS) = 1 - \Phi_1BS - \dots - \Phi_PBS^P \quad (4)$$

$$Seasonal\ MA: \Theta(BS) = 1 + \Theta_1BS + \dots + \Theta_QBS^Q \quad (5)$$

Box and Jenkins bases the model selection on four stages i.e. Identification, Estimation, Diagnostic checking and Forecasting.

Identification

The autoregressive orders p and P and the Moving Average orders q and Q will be decided partly by visual inspection of the correlogram and partly by minimizing information criterions with the use of the HK-algorithm. Identification of the model was concerned with deciding the appropriate values of (p, d, q) (P, D, Q). It was done by observing Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) values up to 24 lags. The HK-algorithm then

performs an iterative procedure to select the model that minimizes the value of each criterion (AIC, RMSE, etc.).

Estimation of parameters

After tentatively identifying the suitable model, next step is to obtain Least Square Estimates of the parameters such that the error sum of squares is minimum.

$$S(\theta, \varphi) = \sum e_t^2(\theta, \varphi) \quad (6)$$

where, t = 1, 2, 3, . . . n.

There are fundamentally two ways of getting estimates for such parameters i.e. *Trial and error* and *Interactive method*.

Diagnostic checking

Examining ACF and PACF of residuals may show an adequacy or inadequacy of the model. If it shows random residuals, then it indicates that the tentatively identified model was adequate. Some of the diagnostic checks are over fitting method, Residual check, Box-Ljung test and minimum Akaike Information Coefficient (AIC). In this analysis, we have used the Residual check and Box-Ljung test.

Residual check

The residuals should behave like Gaussian white noise, that is appear random, homoscedastic and normal (Box and Jenkins, 1976, p. 324).The first part is a graphical check of the standardized residuals, meaning the residuals divided with their standard deviation. These should look random and homoscedastic. The number of outliers is also important where a good indication would be that about 95 percent of the residuals lie inside their 95 percent confidence interval ± 1.96 .

The Box-Ljung test

The Ljung and Box (BL) test was developed in 1978 and is used to test the randomness of the residuals. For this test the first step is to extract the residuals $\hat{\varepsilon}_t$ for the fitted model. The residuals are then used to derive the sample autocorrelations of the residuals with the following equation

$$\hat{r}_k = \frac{\sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^T \hat{\varepsilon}_t^2}, k = 1, 2, \dots \quad (7)$$

This equation is used until a set of autocorrelations $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m$ have been obtained. These are then used to test the null hypothesis of serially independent residuals versus the alternative hypothesis that they are not serially independent with the following test statistic Ljung and Box 1978, p.298.

$$\tilde{Q}(r) = T(T+2) \sum_{k=1}^m (T-k)^{-1} r_k^2 \quad (8)$$

which for an appropriate model was shown to be asymptotically distributed as a $\chi^2_{1-\alpha}(m)$ where m is the number of lagged autocorrelations included and α is the selected significance level.

The number of lags should be a function of T for example the truncated value of $m = \sqrt{T}$ and that the degrees of freedom should be corrected for SARIMA models to $df = m - p - q - P - Q$. The critical value is included in (Ljung and Box, 1978) and then compared to the value of the test statistic. The null hypothesis of randomness is rejected for large values of the test statistic.

Forecasting

The principal objective of developing SARIMA model for a variable is to generate post sample period forecast for the same variable. For instance, let us consider the given Seasonal ARIMA (0, 1, 1) (1, 0, 1)₁₂ we can forecast the next step which is given by Cryer and Chan as:

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

$$z_t - z_{t-1} = \Phi(z_{t-12} - z_{t-13}) + \varepsilon_t - \theta\varepsilon_{t-1} - \Phi\varepsilon_{t-12} + \theta\varepsilon_{t-13} \quad (9)$$

$$z_t = z_{t-1} + \Phi z_{t-12} - \Phi z_{t-13} + \varepsilon_t - \theta\varepsilon_{t-1} - \Phi\varepsilon_{t-12} + \theta\varepsilon_{t-13} \quad (10)$$

The one step ahead forecast from the origin s given by

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

The next step is

$$\hat{z}_{t+1} = \hat{z}_{t+1} + \Phi z_{t-10} - \Phi z_{t-11} - \Phi\varepsilon_{t-10} + \theta\varepsilon_{t-11} \quad (12)$$

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;

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

The accuracy of forecasts for both Ex-ante and Ex-post is tested by using R-square, RMSE, MAPE, etc...

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

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

Results and discussion

The tentative models were selected by careful examination of ACF and PACF values up to 24 lags. The Moving Average (MA) terms and Auto Regressive (AR) terms were found with the help of ACF and PACF respectively. The MA terms were determined by number of non- zero coefficients in ACF and the AR terms were determined by number of non-zero coefficients in PACF (Table 1). Based on highest \bar{R}^2 and lowest RMSE, SARIMA (3, 0, 3) (3, 0, 2)₁₂& (2, 2, 2) (1, 0, 1)₁₂ were the best models for forecasting the tomato arrivals and prices whereas (1, 2, 3) (1, 2, 2)₁₂& (2, 0, 3) (2, 0, 3)₁₂ were the best model for forecasting the onion arrivals and prices in

Koyambedu (Chennai) market respectively. The values of all identified models were shown in Table 2 and parameter values in Table 3. The adequacy of the models can be found out by residual analysis. One basic assumption in ARIMA is the residuals should be independent and normal distributed. The p values of Ljung-box test for tomato arrivals and prices were 0.20 and 0.41 respectively (>0.05) indicates the independence of residuals. The p values of Ljung-box test for onion arrivals and prices were 0.213 and 0.830 respectively (>0.05) indicates the independence of residuals. Most of the residuals of ACF and PACF in Figures 1 to 4 are fall within 3σ limit indicate the best fitted model (i.e. residual are white noise).

The ex-post forecasted values for vegetable arrivals and prices were presented in Table 4 and Figures 5 to 8.

Table.1 ACF and PACF to identify the orders of SARMA (p, q) (P, Q)s, only positive lags are of interest

Terms	ACF	PACF
AR(p)	Exponentially decreasing or damped sine wave	Spikes to lag p then zero
MA(q)	Spikes to lag q then zero	Exponentially decreasing or damped sine wave
ARMA(p, q)	Exponentially decreasing or damped sine wave after q - p lags	Exponentially decreasing or damped sine wave after p - q lags
SAR(P)_s	Exponentially decreasing or damped sine wave for all lags times s	Spikes for lag P _s then zero
SMA(Q)_s	Spikes for lag Q _s then zero	Exponentially decreasing or damped sine wave for all lags times
SARMA(P, Q)_s	Exponentially decreasing or damped sine wave for all lags times s after lags (Q - P) _s	Exponentially decreasing or damped sine wave for all lags times s after lags (P - Q) _s

Table.2 The tentative SARIMA models for vegetable arrivals and prices in Koyambedu market

Type	Models	\bar{R}^2	RMSE
Tomato arrivals	(1, 0, 3) (2, 0, 2) ₁₂	0.713	19571.523
	(2, 0, 3) (1, 0, 3) ₁₂	0.697	19674.257
	(3, 0, 3) (3, 0, 2)₁₂	0.723	19537.798
Tomato prices	(2, 2, 2) (1, 0, 2) ₁₂	0.730	1021.575
	(2, 2, 2) (1, 0, 1)₁₂	0.731	1015.079
	(2, 2, 2) (1, 0, 3) ₁₂	0.729	1031.299
Onion arrivals	(1, 2, 2) (1, 2, 3) ₁₂	0.874	51707.021
	(1, 2, 3) (1, 2, 2)₁₂	0.876	51349.099
	(1, 2, 3) (1, 2, 3) ₁₂	0.875	51718.364
Onion prices	(2, 0, 3) (2, 0, 3)₁₂	0.753	1023.359
	(2, 0, 1) (2, 0, 3) ₁₂	0.742	1033.758
	(3, 0, 2) (2, 0, 3) ₁₂	0.748	1035.061

Table.3 SARIMA parameters for vegetable arrivals and prices in Koyembedu market

Tomato				Onion				
Model	Component		Estimate	Model	Component		Estimate	
Arrivals (3, 0, 3) (3, 0, 2)₁₂	Constant	Lag	95590.316	Arrivals (1, 2, 3) (1, 2, 2)₁₂	Constant	Lag	144820.32	
	AR	1	0.700		AR	1	-0.874	
		2	0.171		MA	1	0.696	
		3	-0.145			2	0.972	
	MA	1	0.577			3	-0.723	
		2	0.424		SAR	1	-0.064	
		3	-0.492		SMR	1	1.325	
	SAR	1	0.205			2	-0.325	
		2	-0.194		Prices (2, 0, 3) (2, 0, 3)₁₂	Constant		3316.183
		3	-0.059			AR	1	-0.200
	SMR	1	-0.269				2	0.452
		2	-0.841			MA	1	-1.326
				2		-0.631		
				3		-0.259		
Prices (2, 2, 2) (1, 0, 1)₁₂	Constant		1872.733	SAR	1	-1.863		
	AR	1	0.617		2	-0.973		
		2	-0.537	SMR	1	2.020		
	MA	1	1.789		2	-1.315		
		2	-0.789		3	0.237		
	SAR	1	0.633					
	SMR	1	0.153					

Table.4 Forecasted values for vegetable arrivals and prices

Forecasted Months (2019)	Tomato		Onion	
	Arrivals (qtls)	Prices (Rs/ctl)	Arrivals (qtls)	Prices (Rs/ctl)
January	98690	1966	162534	2338
February	106369	1471	154697	2537
March	99296	678	136522	3439
April	83343	464	102567	3305
May	102076	717	134123	3949
June	98478	1229	136852	4211

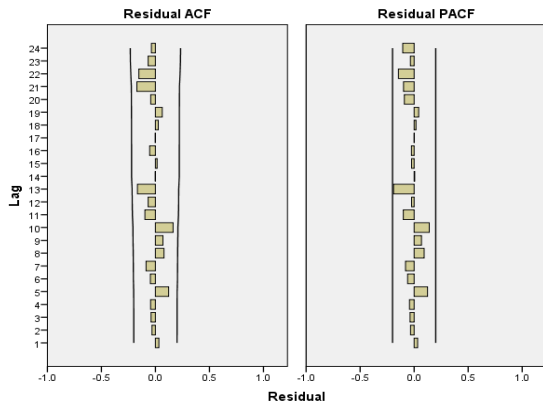


Figure.1 Residual ACF and PACF for tomato arrivals

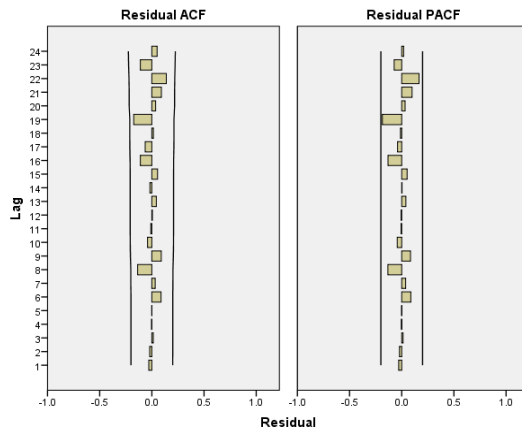


Figure.2 Residual ACF and PACF for onion arrivals

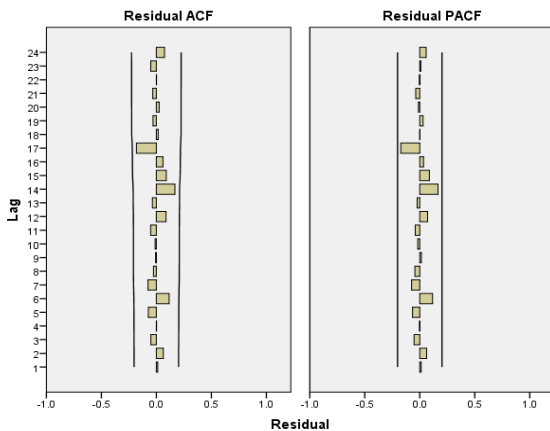


Figure.3 Residual ACF and PACF for tomato prices

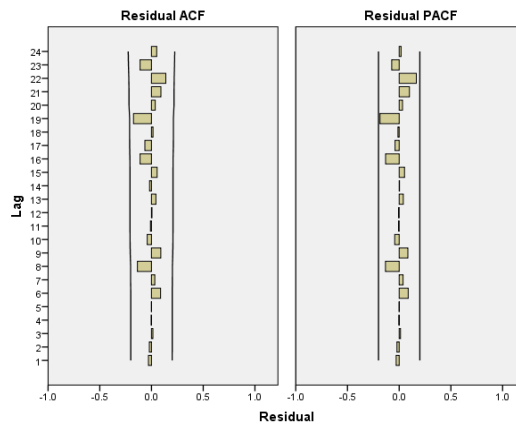


Figure.4 Residual ACF and PACF for onion prices

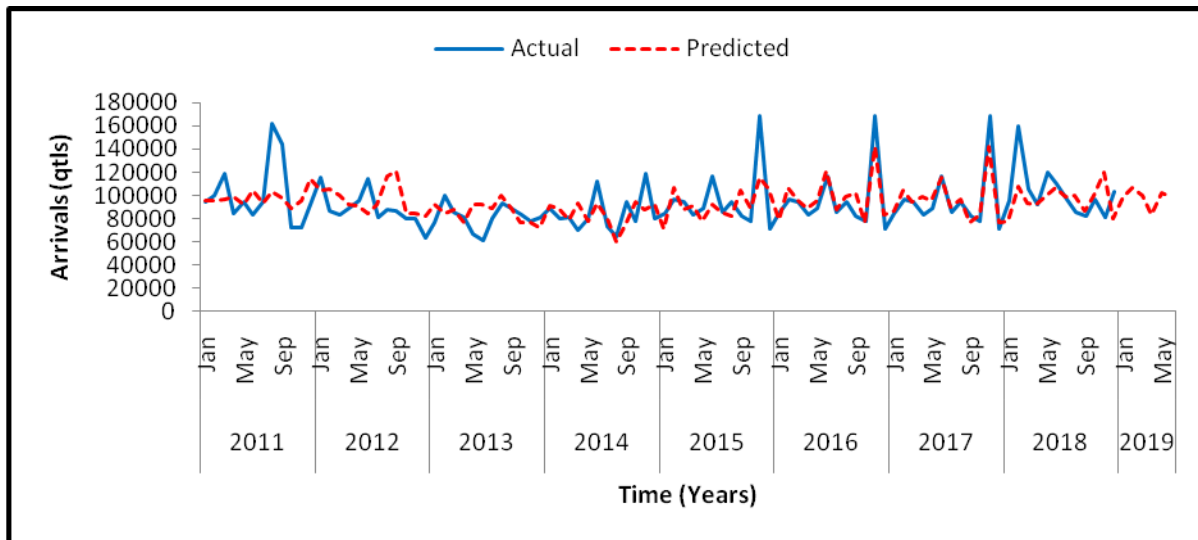


Figure.5 Ex- post forecast of tomato arrivals

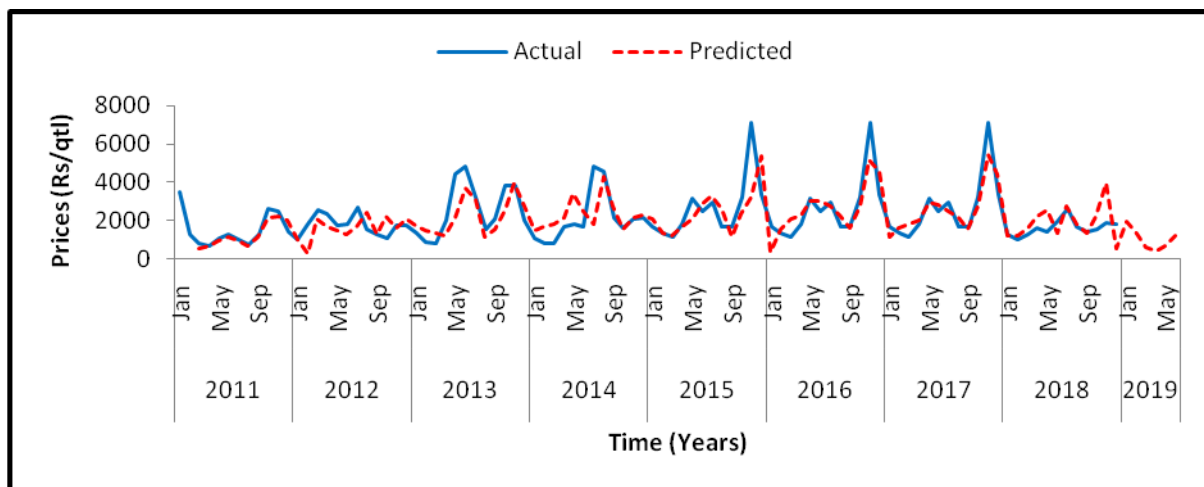


Figure.6 Ex- post forecast of tomato Prices

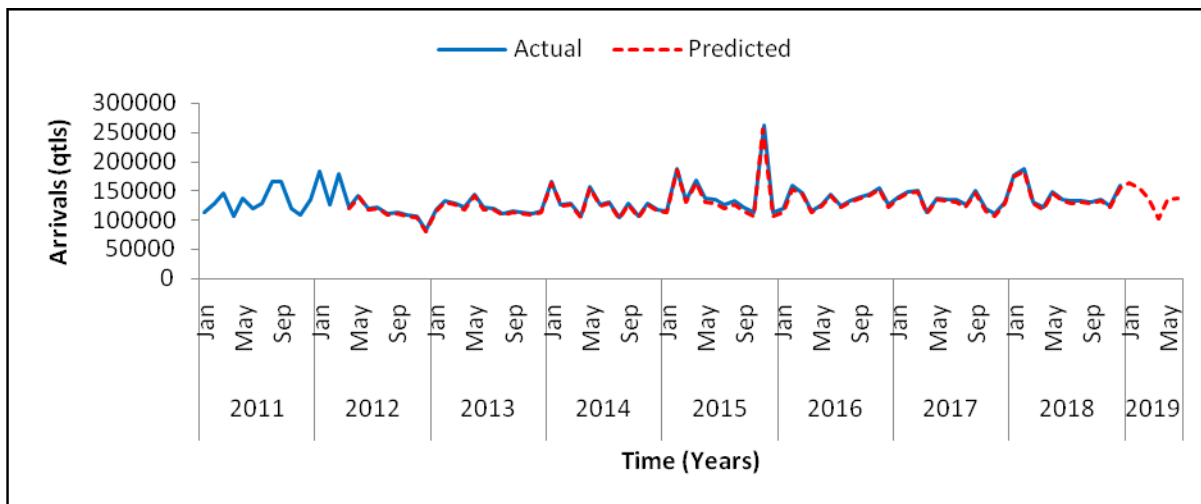


Figure.7 Ex- post forecast of onion arrivals

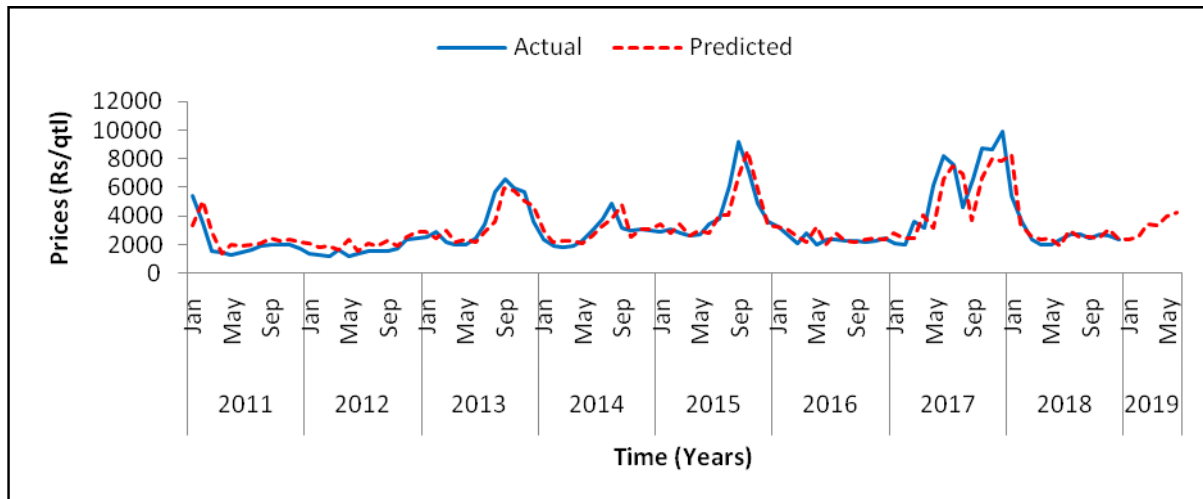


Figure.8 Ex- post forecast of onion Prices

It is concluded, actual and forecasted values were often similar because of high fluctuations in tomato arrivals. Tomato arrivals were high in the month of February (106369 qtls) and May among the forecasted values. The actual and predicted values of tomato prices were similar to each other. There was a peak in tomato prices (Rs 1966/qrtl) in the Month of January among the forecasted values. In onion, actual and forecasted values for arrivals and prices were similar to each other. The peaks of arrivals (162534 qtls) were noticed in the Month of January whereas prices (Rs 4211 /qrtl) were high in June among the forecasted values. The above mentioned SARIMA models will be useful for forecasting the future prices and arrivals. Policies makers will make use of these forecasting to formulate sound policies for the benefit of both producers and consumers.

References

- Boateng, F.O., Amoah-Mensah, J., Anokye, M., Osei, L., and Dzebre, P. 2017. Moseling of tomato prices in Ashanti region, Ghana, using Seasonal autoregressive integrated moving average model. *Journal of Advances in Mathematics and computer Science*, 1-13.
- Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control*, San. Francisco: Holden-Day (Revised edition). 19 and 324.
- Cryer, J. D., and Chan, K. S. 2008. Time series regression models. *Time series analysis: with applications in R*, 249-276.
- Dragan, I., M. Beba, N. Nebojsa, and V. Natasa. 2015. Analysis and prediction of tomato price in Serbia. *Econ. Agric.* 62(4): 951-962.
- Gathondu, E. K. 2014. Modelling of wholesale prices for selected vegetables using time series models in Kenya (Mastersthesis, University of Nairobi).
- Mohan Naidu, G and Sri Kala, M. 2015. Potato price forecasting using seasonal ARIMA model. *Research Journal of Agricultural Sciences*, 6(2): 304-305.
- Robert Mathenge Mutwiri. 2019. 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, 46-53.
- Sampson, W., Suleman, N., and 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.

How to cite this article:

Tamilselvi, C. and Mohan Naidu, G. 2020. Forecasting of Vegetable Arrivals and Wholesale Prices of Koyambedu Market in Chennai Using SARIMA Approach. *Int.J.Curr.Microbiol.App.Sci.* 9(10): 3059-3068. doi: <https://doi.org/10.20546/ijcmas.2020.910.368>