https://doi.org/10.56093/JISAS.V78I2.4
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 78(2) 2024 107–113

Forecasting of Tomato Price in Karnataka using BATS Model

Vinay H.T., Pradip Basak, Arunava Ghosh, Sankalpa Ojha and Chowa Ram Sahu

Uttar Banga Krishi Viswavidyalaya, Cooch Behar

Received 04 December 2023; Revised 31 May 2024; Accepted 25 July 2024

SUMMARY

Tomato plays a vital role in Karnataka's agro-processing and food industries, contributing significantly to the state's economy. Even though tomato production in Karnataka is substantial, the state's market is characterized by price volatility. Tomato prices can undergo drastic fluctuations within short time periods, posing severe challenges to the farmers and consumers. To address these problems, time series models, such as Exponential Smoothing, ARIMA, SARIMA, BATS and TBATS have been implemented to forecast the tomato prices in Kolar market of Karnataka state using monthly wholesale prices data from the year 2010 to 2022. Among the applied models, BATS showed superior performance in terms of model validation criteria such as Root Mean Square Error and Mean Absolute Percentage Error.

Keywords: Tomato price; Exponential Smoothing; ARIMA; SARIMA; BATS and TBATS.

1. INTRODUCTION

Among vegetables, tomatoes hold a prominent place as one of the most consumed worldwide, trailing only behind potatoes and ahead of onions. It is also one of the important cash crops in the state of Karnataka. As per 3rd advanced estimate of 2021-22, area under tomato in Karnataka was 70.10 thousand hectares which contributes to 8.3% of India's total area coverage and it's production was 2104.68 thousand tons which contributes to 10% of India's total production (Anonymous, 2022).The perishable nature of tomatoes makes them a high-risk crop for farmers, as unpredictable price changes can lead to financial instability. However, a significant challenge with vegetable crop like Tomato, is their high perishability and the price volatility resulting from fluctuations in demand and production (Paul *et al.* 2023). Growers invest substantial time, effort, and resources into their tomato crops, and a sudden drop in tomato prices can result in significant losses. On the flip side, when tomato prices increase, consumers may experience price shocks that affect their household budgets and dietary preferences. Therefore, understanding the factors driving these price fluctuations is crucial for developing effective strategies to mitigate their impact and promote stability in the tomato market. Consequently, there is a growing demand for reliable price forecast models that can provide valuable insights into future price trends, helping to address these challenges and uncertainties. Dragan *et al.* (2015) analyzed changes and future tendencies of price parameters of tomato with descriptive statistics and found that the Autoregressive Integrated Moving Average (ARIMA) model was suitable for price forecasting. Boateng *et al.* (2007) formulated a model for tomato prices and found that predictability of the model increases with seasonal-ARIMA (SARIMA). Meena *et al.*(2022) used different arrival and price forecasting models like Moving Average, Weighted Moving Average, Exponential Smoothing (Simple Exponential Smoothing, Trend Analysis or Simple Linear Regression Model) to analyze the trends and ARIMA model to forecast the prices and arrivals. Seasonal patterns are common in most of the time series data, say for example, an hourly time series can exhibit a daily, weekly, monthly and yearly seasonality. The majority of time series models, such as Exponential Smoothing models, ARIMA, and

Corresponding author: Pradip Basak
E-mail address: pradip@ubkv.ac.in

SARIMA are developed to deal with simple seasonal patterns with a small integer-valued periodicity, such as 12 for monthly data or 4 for quarterly data. Many time series data exhibit complex seasonal patterns. For example, an hourly time series can exhibit a daily, weekly, monthly and yearly seasonality. To model such complex seasonal patterns, De Livera *et al.* (2011) introduced the BATS (Box-Cox transformation, ARMA errors, Trend, and Seasonal components) model as an alternative to traditional exponential smoothing. BATS extends homoscedastic ETS (E-Error, T-Trend and S-Seasonal) models to handle multiple seasonality by integrating a Box-Cox transformation for non-linearities and a residual ARMA adjustment for autocorrelation. This framework overcomes key limitations of traditional exponential smoothing, particularly for non-negative and non-linear time series, where the ETS framework often fails. Unlike ETS models, BATS can capture residual autocorrelation, providing more robust forecasts. A few time series models like BATS and TBATS were introduced to solve all of the aforementioned seasonal complexity. Kozuch (2023) analysed quarterly timber net prices of round wood species (Oak, Pine, Beech, Birch, Alder and Spruce) for the years 2005 to 2021 using ARIMA, exponential smoothing, BATS, TBATS, the Prophet model, ANN with RBF and ANN with MLP neural networks, based on accuracy measures they found that BATS and TBATS are efficient models for forecasting beech, birch and alder price of round wood species. Therefore, in this article application of BATS model have been considered to forecast the volatile tomato prices in Kolar market of Karnataka state.

2. MATERIALS AND METHODS

Time series data on monthly Tomato price (Rs/Quintal) in Kolar market of Karnataka were collected from AGMARKNET portal (https://agmarknet.gov.in/) from January, 2010 to December, 2022. Different time series models such as Simple Exponential Smoothing (SES), Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES), ARIMA, SARIMA, BATS and TBATS were applied in this price data.

2.1 Models

2.1.1 Simple Exponential Smoothing (SES)

The method of simple exponential forecasting takes the forecast for the previous period and adjusts it using the forecast error $(Y_t - F_t)$ (Brown, 1956).

$$F_{t+1} = F_t + \alpha\,(Y_t - F_t) \text{ or } F_{t+1} = \alpha Y_t + (1-\alpha)\,F_t$$

where, Y_t is the actual value and F_t is the forecasted value at time *t*, α is the smoothing parameter ranges from 0 to 1.

2.1.2 Double Exponential Smoothing (DES) or Holt's linear method

This method involves a forecast equation and two smoothing equations (one for the level and one for the trend) (Holt, 1957). The equations for the models are given as follows:

$$L_t = \alpha Y_t + (1-\alpha)\,(L_{t-1} + b_{t-1})$$

$$b_t = \beta[L_t - L_{t-1}] + (1-\beta)\,b_{t-1}$$

$$F_{t+h} = L_t + h b_t$$

where, L_t is the Level and b_t is the trend at time *t*, F_{t+h} is the forecast value for *h* period ahead, and α, β are smoothing parameters ranging from 0 to 1.

2.1.3 Triple Exponential Smoothing (TES) or Holt-Winter's Exponential Smoothing (H-WES)

Holt-Winter's Exponential Smoothing (H-WES) methods are widely used when the data shows trend and seasonality (Winter, 1960). The Holt-Winter's method depends on three smoothing equations: one for level, one for trend, and one for seasonality. The model equations are given as

$$L_t = \alpha\,(Y_t - S_t) + (1-\alpha)\,(L_{t-1} + b_{t-1})$$

$$b_t = \beta\,(L_t - L_{t-1}) + (1-\beta)\,b_{t-1}$$

$$S_t = \gamma\,(Y_t - L_t) + (1-\gamma)\,S_{t-s}$$

$$F_{t+h} = L_t + hb_t + S_{t-s+h}$$

where, s is the length of seasonality, S_t is the seasonal component at time t, and α, β and γ are level, trend and seasonal smoothing constants or the weights respectively, which lies between 0 and 1.

2.1.4 ARIMA

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors (Box and Jenkins, 1976), it can be expressed as;

$\varphi_p(B)(1-B)^d\, y_t = \theta_q(B)\varepsilon_t$ where, y_t and ε_t are the actual value and random error at time period *t*, *d* denotes the order of differencing, $\varphi_p(B)$ and $\theta_q(B)$

are the autoregressive operator (p) and moving average operator (q).

2.1.5 SARIMA

When time series data have seasonal component, SARIMA model is employed. SARIMA model is characterized by SARIMA (p, d, q) $(P,D,Q)_s$ and it is given by

$$\left(1-\varphi_p B\right)\left(1-\Phi_P B^s\right)\left(1-B\right)\left(1-B^s\right)y_t = \left(1-\theta_q B\right)\left(1-\Theta_Q B^s\right)\varepsilon_t$$

where, B is the backshift operator, $s$is seasonal lag, ε_t is sequence of independent normal error with mean 0 and variance σ^2, Φ's and φ's are respectively the seasonal and non-seasonal autoregressive parameters and Θ's and θ's are respectively the seasonal and non-seasonal moving average parameters. Here, p and q are orders of non-seasonal autoregressive and moving average parameters respectively, whereas P and Q are the seasonal autoregressive and moving average parameters respectively. Also 'd' and 'D' denote non-seasonal and seasonal differences respectively (Makridakis *et al.*1998).

2.1.6 BATS (B: Box-Cox transformation A:ARIMA errors T: Trend S: Seasonal components) model

Box-Cox transformation is a power transformation that helps make the series stationary, by stabilizing the variance and mean over time. BATS model is developed by extension of Double-Seasonal Holt-Winter's (DSHW) method with Box-Cox transformation, ARMA errors, Trend, and multiple seasonal patterns (De Livera, 2012).

$$y_t^{(\omega)} = \begin{cases} \dfrac{y_t^{\omega}-1}{\omega}; \omega \neq 0 \\ \log y_t \quad \omega = 0 \end{cases}$$

$$y_t^{(\omega)} = l_{t-1} + \varnothing b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \varnothing b_{t-1} + \alpha d_t$$

$$b_t = \left(1-\varnothing\right)b + \varnothing b_{t-1} + \beta d_t$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t$$

$$d_t = \sum_{i=1}^{p}\varphi_i d_{t-i} + \sum_{i=1}^{q}\theta_i \varepsilon_{t-i} + \varepsilon_t$$

where, $y_t^{(\omega)}$ represents Box-Cox transformed observations with a parameter ω at time t, $m_1, \ldots, m_T$ denote the seasonal periods, l_t is the local level at time t, b is the long-run trend and b_t is the short-runtrend at time t, $s_t^{(i)}$ indicates the i^{th} seasonal component at time t, d_t represents an ARMA (p, q) process, ε_t is a Gaussian white-noiseprocesswithzeromeanandconstantvariance σ^2, and the smoothing parameters are given by α, β, and γ_i for i=1, ..., T. The model was represented by BATS (ω, (p, q), $\varnothing$, $m_1, m_2, \ldots, m_T$), where, ω is the Box–Cox transformed value, (p, q) is ARMA components, $\varnothing$ dampening parameter, m_i represents i^{th} seasons.

2.1.7 TBATS (T: Trigonometric B: Box-Cox transformation A: ARIMA errors T: Trend S: Seasonal components) model

For high frequency and non-integer seasonality BATS model are not efficient, therefore, to overcome this problem, TBATS was introduced as an extension of BATS model by adapting the following equations (De Livera *et al.* 2011):

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin\lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos\lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

where, $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$,are the smoothing parameters, $\lambda_j^{(i)} = \dfrac{2\pi j}{m_i}$, $s_{j,t}^{(i)}$ describe the stochastic level of the i^{th} seasonal component, $s_{j,t}^{*(i)}$ describe the stochastic growth of the i^{th} seasonal component, k_i is the number of harmonics required for the i^{th} seasonal component, $k_i = \dfrac{m_i}{2}$ for even values of m_i, and $k_i = \dfrac{(m_i-1)}{2}$ for odd values of m_i.

2.2 Model evaluation criteria

The model performance was assessed based on the following criteria: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

RMSE $= \sqrt{\dfrac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$ and

$$\textbf{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| * 100$$

where, n is number of observations, Y_t and $\hat{Y}_t$ are the actual and forecasted price at time t. The measurements like RMSE and MAPE are employed to gauge how accurate the model's predictions are; a lower value for these metrics denotes a more accurate prediction from the model.

3. RESULTS AND DISCUSSION

The monthly price data from January, 2010 to December, 2022 is split into training and testing set as 80:20. The training dataset is used to build the SES, DES, TES, ARIMA, SARIMA, BATS, and TBATS models and testing data set is used set for evaluating forecasting performance of these models. Based on model evaluation criteria *i.e.,* RMSE and MAPE, the best fitted model has been used for the forecasting of monthly tomato price for the year 2023 in Kolar market of Karnataka. The complete methodology of this study is depicted in the Fig. 1.

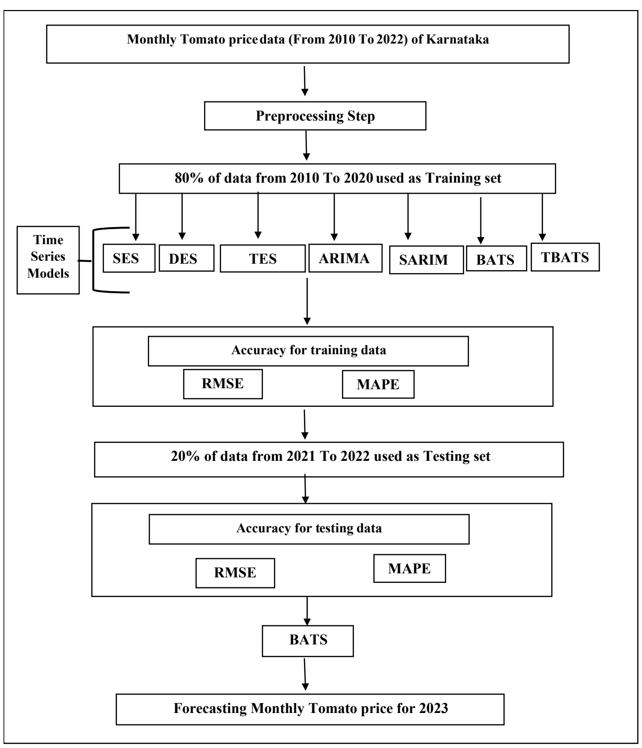

Fig. 1. Schema of the best model for forecasting

3.1 Descriptive statistics and seasonal indices

The descriptive statistics of tomato prices in the Kolar market is presented in Table 1. According to Table 1, the average tomato price in Kolar market is Rs.1058 /quintal. Since the CV is greater than 50%, we can deduce that the price variability is slightly higher. The time series is positively skewed and leptokurtic. The monthly seasonal index values are shown in Table 2. Seasonal indices have greater values from May to August and October to December. Time plot of the average monthly tomato price for the original series is depicted in Fig. 2.

Table 1. Descriptive Statistics of Tomato prices in Kolar market of Karnataka

Statistics	Price
Observations	156
Mean (Rs/quintal)	1058
Minimum Median	214 830
Maximum	4114
Standard Deviation	696
Coefficient of Variation	65
Skewness	1.72
Kurtosis	3.37

Table 2. Seasonal Indices for tomato prices

Months	Seasonal Index
January	0.92
February	0.62
March	0.59
April	0.69
May	1.16
June	1.29
July	1.44
August	1.08
September	0.99
October	1.05
November	1.16
December	1.01

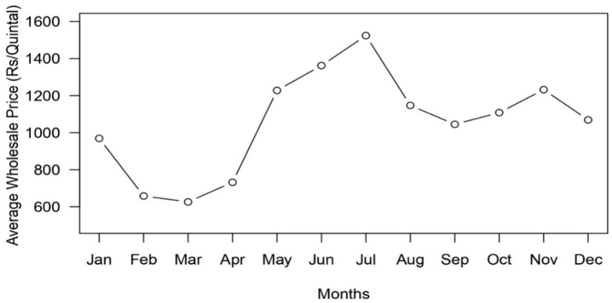

Fig. 2. Monthly average wholesale price of Tomato in Kolar market, Karnataka

3.2 Fitting of Exponential Smoothing models

Different combination of smoothing parameters α, β and γ are tried by grid search method to get suitable Exponential Smoothing parameters. The estimates of model parameter and its standard error are given in Table 3.

Table 3. Parameter estimates of SES, DES and TES model

Models	Parameters	Estimates	AIC
SES	Alpha (Level)	0.9999	2325.906
DES	Alpha (Level)	0.9999	2330.059
	Beta (Trend)	0.00001	
TES	Alpha (Level)	0.9999	2307.802
	Beta (Trend)	0.0246	
	Gamma (Seasonal)	0.00001	

3.3 Fitting of ARIMA model

The data was checked for stationarity by using ADF test, first differencing is used to transform the data from non-stationary to stationary and the results was presented in Table 4. ARIMA (0, 1, 3) is found to be the best fit model by using "auto. arima" function in R software. The parameter estimates of fitted ARIMA model are furnished in Table 5 along with their significance level.

Table 4. Results of Augmented Dicky Fuller test

Data	ADF test	Lag order	P-value
Original	-3.03	11	0.149
Differenced	-6.23	11	0.010

Table 5. Parameter estimates of ARIMA (0, 1, 3) model

Parameters	Estimate	S.E.	p-value
MA1	-0.09	0.08	0.317
MA2	-0.46	0.07	0.000***
MA3	-0.38	0.08	0.000***

***: Significant at 0.1%

3.4 Test for Seasonality

To assess the seasonal variation in data, Kruskal-Wallis Seasonality test is applied. The test statistics value is 20.38 which corresponds to p-value of 0.04 indicating that there is seasonality in the data.

3.5 Fitting of SARIMA model

SARIMA (0, 1, 3) $(0, 0, 2)_{[12]}$ is chosen as the best fit model based on the lowest AIC and BIC values while keeping the maximum orders for p, q, P, and Q at three. The parameter estimates of fitted SARIMA model are furnished in Table 6 along with their significance level.

Table 6. Parameter estimates of the SARIMA (0, 1, 3) $(0, 0, 2)_{[12]}$ model

Parameters	Estimate	S.E.	p-value
MA1	-0.07	0.08	0.409
MA2	-0.46	0.07	0.000***
MA3	-0.39	0.08	0.000***
SMA1	-0.05	0.09	0.575
SMA2	0.15	0.09	0.091

***: Significant at 0.1%

3.6 Fitting of BATS model

The grid search approach is adopted to test several combinations of smoothing parameters, and Box-Cox transformed value, and finally, BATS (0.001, {3, 1}, -, {12}) model is chosen on the basis of lowest AIC value. The Box-Cox transformation (ω) is found to be 0.001 that means a slight transformation is used to convert the data into stationary, ARMA(3, 1) model with the lowest AIC value of 2189.23 is determined to be the best, as shown in Table 7.

Table 7. Parameter estimates of BATS model

Model	*Box-Cox transformation (Omega)	Smoothing parameter			Phi	ARMA coefficients		Prediction error	
		Alpha	Beta	Gamma		AR coefficients	MA coefficients	Sigma	AIC
BATS (0.001, {3, 1}, -, {12})	0.001	0.058	-	-0.123	-	1.144 (AR1)	-0.230 (MA1)	0.335	2189.23
						-0.776 (AR2)			
						0.358 (AR3)			

Table 8. Parameter estimates of TBATS model

Model	Omega	Smoothing parameter				Phi	ARMA coefficients		Prediction error	
		Alpha	**Beta**	**Gamma-1**	**Gamma-2**		**AR coefficients**	**MA coefficients**	**Sigma**	**AIC**
TBATS (0.009, {3, 0}, 0.908, {<12, 2>})	0.009	-0.109	0.023	-0.0001	0.0001	0.908	0.987 (AR1)	-	0.395	2206.33
							-0.509 (AR2)			
							0.290 (AR3)			

3.7 Fitting of TBATS model

The grid search method is used to evaluate various combinations of smoothing parameters, damping parameter, and Box-Cox transformed value, and finally, TBATS (0.009, {3, 0}, 0.908, {12, 2}) model with the lowest AIC value is chosen. The parameter estimates of the model are presented in Table 8.

3.8 Forecasting performance evaluation

BATS model performed better than the other six models, with the minimum RMSE of 384.35 and MAPE of 26.31% under training set as presented in Table 9. Similarly, in the testing set, BATS is found to be the best fitted model on the basis of lowest RMSE of 233.06 and MAPE of 30.06%, and the results are given in Table 9. Therefore, BATS model is used to forecast the monthly wholesale prices of tomato for the year 2023. The plot of observed vs fitted values from BATS model is presented in Fig. 3.

Table 9. Model accuracy evaluation

Models	Training set		Testing set	
	RMSE	**MAPE (%)**	**RMSE**	**MAPE (%)**
SES	570.28	39.97	348.39	52.07
DES	570.61	40.77	363.37	51.05
TES	493.49	37.20	355.34	38.45
ARIMA	479.17	34.51	305.40	42.59
SARIMA	479.12	34.47	305.31	42.49
BATS	**384.35**	**26.31**	**233.29**	**30.06**
TBATS	430.13	30.56	256.01	33.54

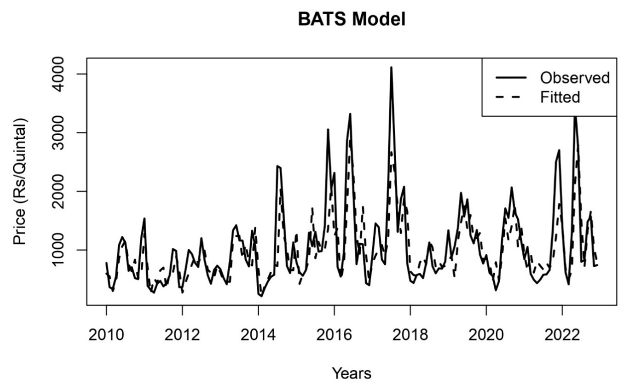

Fig. 3. Plot showing observed vs fitted values by BATS model

3.9 Residual diagnostics

Residual diagnostics were conducted on the best-fitted model (BATS model). The Box-Ljung test results indicate that the residuals are random in nature. To test the normality, Shapiro-Wilk test has been applied and from the results depicted in Table 10, it can be concluded that the null hypothesis is accepted, as the p value is more than 0.05, indicating that the residuals are normally distributed. To check the presence of ARCH (Autoregressive Conditional Heteroskedasticity) effect, ARCH-LM test has been carried out and from the results it is confirmed that there is no existing ARCH effect in residuals, *i.e.,* residuals are homoscedastic in nature.

Table 10. Residual diagnostics

Diagnostic test	Test statistic	p-value
Box-Ljung	0.002	0.968
Shapiro-Wilk	0.994	0.896
ARCH-LM	15.35	0.223

3.10 *Forecasting monthly wholesale price of Tomato*

Forecasted values of monthly wholesale prices of tomato for the year 2023 based on BATS model are presented in Table 11 and the results revealed that in the month of July, 2023 the price of Tomato is expected to be maximum followed June, 2023.

4. CONCLUSIONS

In this article, monthly wholesale prices of tomato are modelled by fitting Exponential smoothing, ARIMA, SARIMA, BATS, and TBATS models. It is found that the BATS model outperforms the other models in terms

Table 11. Forecasted monthly wholesale prices of Tomato

Month-year	Price (Rs/quintal)
Jan-23	848.78
Feb-23	708.62
Mar-23	747.88
Apr-23	719.97
May-23	983.06
Jun-23	1439.64
Jul-23	**2012.92**
Aug-23	1380.55
Sep-23	1206.05
Oct-23	1243.74
Nov-23	1125.94
Dec-23	873.62

of model evaluation criteria such as RMSE and MAPE. Monthly wholesale prices of tomato are forecasted by using BATS model for the year 2023 and it is found that the price of tomato is expected to be higher in the month of July followed by June. Therefore, with the help of this study, farmers of Karnataka state may be advised when to plan their crops such that they can maximize profits from tomato farming.

ACKNOWLEDGEMENTS

The authors are grateful to Agricultural Marketing Information Network (AGMARKNET) for providing the data.

REFERENCES

Anonymous., *Agricultural Statistics at a Glance 2022* (2023). Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India.

Boateng, F.O., John, A.M., Martin, A., Lot, O. and Paul, D. (2017). Modeling of tomato prices in Ashanti region, Ghana, using seasonal autoregressive integrated moving average model. British Journal of Mathematics & Computer Science, **20**, 1-13.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and control*. Revised Edition, San Francisco: Holden Day.

Brown, R.G. (1956). *Exponential smoothing for predicting demand*. Cambridge mass Arthur d little.

De Liver, A.M., Hyndman, R.J. and Snyder, R.D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, **106**, 1513-1527.

De Livera, A.M. (2012). Automatic forecasting with a modified exponential smoothing state space framework. Monash Econometrics and Business Statistics Working Papers, **10**, 1-6.

Dragan, I., Beba, M., Nebojsa, N. and Natasa, V. (2015). Analysis and prediction of tomato price in Serbia. *Economics of Agriculture*, **62**, 951-962.

Holt, C.C. (1957). Forecasting trends and seasonals by exponentially weighted moving averages. Office of Naval Research, Research memorandum, **52**, 5-10.

https://apeda.gov.in/apedawebsite/#home (accessed on 10 October 2023).

Kozuch, A., Cywicka, D. and Adamowicz, K. (2023). A Comparison of Artificial Neural Network and Time Series Models for Timber Price Forecasting. *Forests*, **14**, 177.

Makridakis, S., Steven, C. W. and Hyndman, R.J. (1998). *Forecasting methods and applications*. 3rd edition, John Wiley and sons New York,

Meena. S., Peram, N., Sharma, A. and Surliya, V. (2022). Price forecasting of tomato by using moving average forecasting model, weighted moving average forecasting model, simple exponential smoothing forecasting model and ARIMA model. *Journal of applied management and entrepreneurship*, **16**, 25-37.

Paul, R.K., Das, T. and Yeasin, M. (2023). Ensemble of time series and machine learning model for forecasting volatility in agricultural prices. *National Academy Science Letters*, 1-4.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, **6**, 324-342.