Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods

Sourav Kumar Purohit, Sibarama Panigrahi, Prabira Kumar Sethy & Santi Kumari Behera

APPLIED ARTIFICIAL INTELLIGENCE
2021, VOL. 35, NO. 15, 1388–1406
https://doi.org/10.1080/08839514.2021.1981659

Taylor & Francis
Taylor & Francis Group

Check for updates

Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods

Sourav Kumar Purohit [a], Sibarama Panigrahi [a], Prabira Kumar Sethy [b], and Santi Kumari Behera [a]

[a]Department of Computer Science Engineering and Application, Sambalpur University Institute of Information Technology, Odisha, India; [b]Department of Electronics, Sambalpur University, Odisha, India

ABSTRACT

Accurate prediction of crop prices assists farmers to decide the best time to sell their produce so as to get maximum benefit and assists Government for post-harvest storage and management of the produce so as to stabilize the price volatility throughout the year. At the same time, pricing of crop depends on various factors including the amount of cultivation, demand of consumers, climate, etc. Hence, the prediction of crop prices is a challenging and important problem. Inspired from this, in this study, we have proposed two additive hybrid methods (Additive-ETS-SVM, Additive-ETS-LSTM) and five multiplicative hybrid methods (Multiplicative-ETS-ANN, Multiplicative-ETS-SVM, Multiplicative-ETS-LSTM, Multiplicative-ARIMA-SVM, Multiplicative-ARIMA-LSTM) to predict the monthly retail and wholesale price of three most commonly used vegetable crops of India, namely, tomato, onion, and potato (TOP). The obtained results are compared with two most promising statistical models, three leading machine learning models and five hybrid methods existing in the literature. Extensive statistical analyses of simulation results considering mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), and root mean square error (RMSE) confirm the superiority of the hybrid methods in predicting the TOP prices.

ARTICLE HISTORY
Received 1 April 2021
Accepted 13 September 2021

Introduction

Vegetable crops play a vital role in India's economy, enhancing income of rural households. Cultivation of fruits and vegetables is a labor intensive task and hence generates a lot of employment in the rural areas. India has diverse kinds of soil and climate, across several agro-ecological regions that make production of a wide variety of horticultural crops possible. Vegetables, fruits, tubers, roots, ornamental plants, medicinal, aromatic, flowers, condiments, spices, mushrooms, and plantation crops form the major part of the total produce in agricultural domain of the country (available at: www.businessworld.in, accessed August 17, 2020).

CONTACT Sibarama Panigrahi spanigrahi@suiit.ac.in Department of Computer Science Engineering and Application, Sambalpur University Institute of Information Technology, Odisha, 768019, India

Tomato, onion, and potato (TOP) crops are considered as a regular food commodity across India. However, the management of TOP crops is at high risk because they are only abundantly available in short spells, which is coupled with the problems of storage, transportation and the consequent post-harvest losses. Hence, it poses a challenge to marketing the produce (available at: www.nhb.gov.in, accessed March 20, 2020) and wide price fluctuations in the market. The volatility of TOP crop prices causes a chaos in the households of this country (Kumar and Joshi 2016). Operation Green is a project introduced by the Ministry of Food Processing Industries of Government of India with the target to stabilize the supply of TOP crops in India by ensuring their availability throughout the year without price volatility (Kumar et al. 2012).

Statistical forecasting models have been widely used to assist decision maker to plan the future more efficiently and effectively. In a developing country like India, proper planning by using efficient forecasting is of utmost importance for the sustainable growth of the country. Considering these facts, a systematic study is carried out to model and predict the wholesale and retail price of TOP crops in India. Traditional linear statistical models, viz., moving average (MA), autoregressive (AR), autoregressive integrated moving average (ARIMA), autoregressive moving average (ARMA), exponential smoothening, etc., can be used to predict the TOP prices. However, the linear statistical models work under the assumption of linear correlation structure of time series data. Hence, these models are quite competent in capturing the linear patterns existing in time series data. However, these models often fail to handle the nonlinear patterns equally well (Zhang 2003). In order to handle the nonlinear patterns efficiently, nonlinear machine learning models like long short-term memory (LSTM), artificial neural network (ANN), and support vector machine (SVM) have been increasingly considered in the literature. Both linear statistical-based methods and nonlinear machine learning based methods have achieved unparalleled success in their respective linear and nonlinear time series forecasting (TSF) domains. However, the application of nonlinear machine learning methods to linear time series and statistical based methods to nonlinear time series often produces poor forecasts. Additionally, when a series contains both linear and nonlinear patterns, neither the linear statistical models nor the nonlinear machine learning models can provide satisfactory result (Zhang 2003). However, the application of hybrid models by integrating linear statistical models with nonlinear models provide a better chance to capture the underlying combined linear and nonlinear correlation structure of time series. Therefore, a variety of hybrid methods by integrating linear statistical models with nonlinear machine learning models have been developed.

Zhang (2003) introduced the concept of integrating linear and nonlinear models for TSF. Zhang assumed the time series as an addition of two components, namely, linear and nonlinear component. In his approach, at first ARIMA model is applied to a time series to capture the linear component of the time series. Then, the predictions from ARIMA model are obtained, which are subtracted from the original time series to attain the residual series. The obtained residual series is considered as nonlinear, and ANN is used to attain the forecasts on residual series. The final forecasts are calculated by adding ARIMA forecasts and ANN forecasts. Using the same concept, Faruk (2010) proposed a hybrid ARIMA-ANN model, which is trained using an optimized conjugated algorithm and used for forecasting the water quality time series.

Khashei and Mehdi (2011) presented another variant of hybrid ARIMA-ANN model with improved forecasting accuracy. In their model, first the ARIMA model is fitted directly to the given time series and a single data value is forecasted. Then the time series, ARIMA forecasted values, and the forecast error are given as input to ANN to compute the future values. Khashei and Mehdi's (2011) model provided better forecasting accuracy than Zhang's (2003) model, ARIMA and ANN.

Contrasting to the additive hybrid models, Wang et al. (2013) assumed the time series data to be a multiplication of two components, namely, linear and nonlinear component. In their hybrid model, the ARIMA model is applied on the time series data and ARIMA forecasts are calculated. Then, the residual series is computed by dividing ARIMA forecasts from original series. Considering the obtained residual series as nonlinear, ANN is used to obtain the forecasts on nonlinear component. The final forecasts are computed by multiplying ARIMA forecasts with ANN forecasts. Wang et al.'s (2013) model has shown improved forecasting accuracy than Zhang's (2003) model in three time series. However, the method can't be applied on a series when the ARIMA forecasts contain zero values.

In the aforementioned hybrid models, ARIMA model is directly applied on a given time series to model the linear patterns of time series. However, ARIMA model can be suitably applied on a series when the series is stationary and Gaussian in nature. This is because; the ARIMA model parameters are obtained by using GMLE (Yao and Brockwell 2006) procedure, which considers the time series data as Gaussian. Hence, when the time series is non-Gaussian in nature, the direct use of ARIMA to such time series consequences in deprived fitting of ARIMA and hence adversely impacts the performance of aforementioned hybrid models. In order to overcome this, Panigrahi and Behera (2017) developed an additive hybrid ETS-ANN model by assuming the time series has two components. The components may have linear or nonlinear characteristics. The ETS-ANN model provided better forecasting accuracy than Babu and Reddy (2014) and Zhang (2003) hybrid models considering 16 time series datasets.

Despite strong approximation capability and superior performance of SVM and LSTM, the performance of SVM and LSTM models in hybrid forecasting methods have not yet been tested. Therefore, we have introduced and analyzed the performance of Additive-ETS-SVM and Additive-ETS-LSTM and

Multiplicative-ETS-ANN, Multiplicative-ETS-SVM, Multiplicative-ETS-LSTM, Multiplicative-ARIMA-SVM, and Multiplicative-ARIMA-LSTM models in addition to the existing hybrid models to forecast the TOP prices.

The rest of the paper is structured as follows. Section 2 presents the material and methodology used to predict the TOP prices. The simulation results are discussed and presented in Section 3. Section 4 concludes the findings of this paper.

Material and Methodology

Yearly data on yield (MT ha-1) of tomato, onion, and potato from 2013 to 2018 were collected from the report of Horticultural Statistics Division, Department of Agriculture, Cooperation and Farmer Welfares, Govt. of India (available at: http://agricoop.gov.in/, accessed March 12 2020) and Database of National Horticulture Board (NSB) (available at: http://nhb.gov.in/, accessed March 12 2020). The descriptive statistics of the monthly retail price of onion (Onion_Retail), potato (Potato_Retail), and tomato (Tomato_Retail) and monthly wholesale price of onion (Onion_Wholesale), potato (Potato_Wholesale), tomato (Tomato_Wholesale) are presented in Table 1. It can be observed that all the time series data are right asymmetric and platykurtic (non-Gaussian). The non-Gaussian series are highly volatile and are difficult to predict (Babu and Reddy 2014). This demands a systematic study on forecasting methods to efficiently predict these time series datasets.

In this study, we have considered two most popular statistical forecasting models, namely, ARIMA, exponential smoothing with error, trend, and seasonality (ETS); and three most popular machine learning models, namely, support vector machine for regression (SVM), long short-term memory (LSTM), and multilayer perceptron (MLP). Additionally, additive and multiplicative hybrid methods considering the aforementioned statistical and machine learning models are considered. The most parsimonious ARIMA and ETS models for the time series are determined using the Forecast package of R (Hyndma and Khandakar 2008), while the MLP, SVM, and LSTM are implemented by using the corresponding toolboxes of MATLAB. The MLP network is trained using the Levenberg–Marquardt algorithm and early

Table 1. Descriptive statistics of time series datasets.

	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
Onion_Retail	14.07	57.21	25.17	10.78	1.45	4.39
Onion_Wholesale	1014.80	4991.41	2016.18	980.61	1.51	4.60
Potato_ Retail	13.54	29.81	18.20	3.94	1.10	3.68
Potato_ Wholesale	980.15	2452.17	1404.58	346.42	1.18	3.91
Tomato_ Retail	14.67	59.91	26.34	10.15	1.25	4.35
Tomato_ Wholesale	1079.14	4999.49	2048.29	890.64	1.27	4.37

Table 2. Models used in simulations and division of dataset.

	ARIMA	ETS	Lags	Length of Train	Length of Validation	Length of Test
Onion_Retail	ARIMA(0,0,2)	ETS(A,N,N)	12	48	12	12
Onion_Whole	ARIMA(0,0,2)	ETS(A,N,N)	12	48	12	12
Potato_Retail	ARIMA(2,0,0)	ETS(A,N,N)	12	48	12	12
Potato_Whoe	ARIMA(2,0,0)	ETS(A,N,N)	12	48	12	12
Tomato_Reta	ARIMA(2,0,0)	ETS(A,Ad,N)	12	48	12	12
Tomato_We	ARIMA(2,0,0)	ETS(A,Ad,N)	12	48	12	12

stopping is used for model validation. The used ARIMA and ETS models; the significant number of lags for MLP, SVM, and LSTM model; and the length of train, validation, and test are presented in Table 2.

In the additive hybrid methods as in Figure 1, the time series $y = \left[y_1, y_2, \ldots\ldots\ldots\ldots, y_n\right]^T$ is considered as an addition of a linear (L) and a nonlinear (N) component as in Equation (1). First, a linear model is applied on the time series to obtain the forecasts on linear component (L). Then, the residual series ($\boldsymbol{e}$) is computed by subtracting the forecasts on linear component (L) from the original time series y as in Equation (2). The residual series is used by a nonlinear model to obtain the forecasts on nonlinear component N. Then, the final forecasts are obtained by adding the forecasts on linear component with the forecasts on nonlinear component as in Equation (3). In this paper, we have used ARIMA and ETS as the linear model; and MLP, SVM, and LSTM as the nonlinear model. Hence, a total of six different combinations, namely, Additive-ARIMA-ANN, Additive-ARIMA-SVM, Additive-ARIMA-LSTM, Additive-ETS-ANN, Additive-ETS-SVM, and Additive-ETS-LSTM are obtained and used for forecasting.

$$y = L + N \tag{1}$$

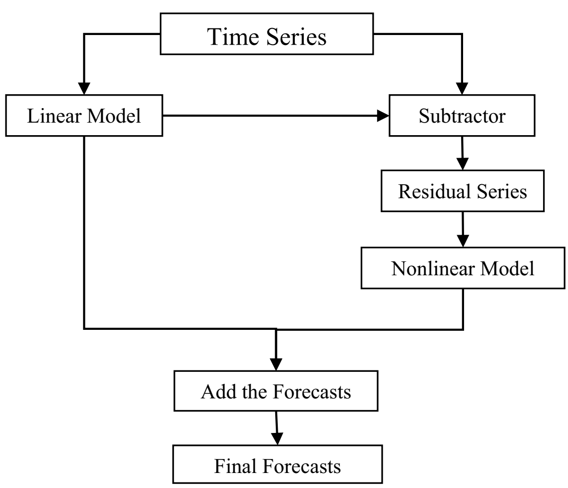

Figure 1. Schematic representation of additive hybrid methods.

$$e = y - \widehat{N} \tag{2}$$

$$\widehat{y} = \widehat{L} + \widehat{N} \tag{3}$$

In the multiplicative hybrid methods as in Figure 2, the time series $y = [y_1, y_2, \ldots\ldots\ldots, y_n]^T$ is considered as a multiplication of a linear (L) and a nonlinear (N) component as in Equation (4). First, a linear model is applied on the time series to obtain the forecasts on linear component (L). Then the residual series ($\boldsymbol{e}$) is computed by dividing the forecasts on linear component (L) from the original time series y as in Equation (5). The residual series is used by a nonlinear model to get the forecasts on nonlinear component N. Then, the final forecasts are computed by multiplying the linear component forecasts with nonlinear component forecasts as in Equation (6). In this paper, we have used ARIMA and ETS as the linear model; and MLP, SVM, and LSTM as the nonlinear model. Hence, a total of six different combinations namely, Multiplicative-ARIMA-ANN, Multiplicative-ARIMA-SVM, Multiplicative-ARIMA-LSTM, Multiplicative-ETS-ANN, Multiplicative-ETS-SVM, and Multiplicative-ETS-LSTM are obtained and used for forecasting. The Multiplicative-ARIMA-ANN (Wang et al. 2013) is an existing method while the others are proposed in this paper. The multiplicative hybrid methods have the problem of division by zero and it occurs when the forecast on linear component ($\hat{L}$) is zero. Hence, to avoid this problem we have set $\hat{L}$ to 0.1 when it has a value 0.

$$y = L \times N \tag{4}$$

$$e = y \div \widehat{N} \tag{5}$$

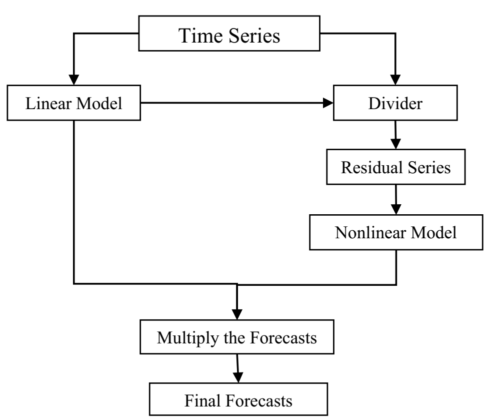

Figure 2. Schematic representation of multiplicative hybrid methods.

$$\widehat{y} = \widehat{L} \times \widehat{N} \tag{6}$$

Results and Discussion

This section presents the simulation results using mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), and root mean square error (RMSE). Since the stochastic models like MLP and LSTM are used in the hybrid methods, we have repeated the simulations fifty times for each dataset and method separately and measured the forecasting accuracy. To draw decisive conclusions, we have conducted statistical tests such as Wilcoxon signed-rank test (Hollander, Wolfe, and Chicken 1999) with 95% confidence level. The results are presented in three subsections such as i) Results for onion time series, ii) Results for potato time series and iii) Results for tomato time series.

Results for Onion Time Series

Table 3 presents the mean RMSE, SMAPE and MAE over fifty independent simulations on monthly onion retail price time series. It can be observed from Table 3 that the Additive-ARIMA-ANN method provides the best RMSE, SMAPE, and MAE in Onion retail price time series. However, the Wilcoxon signed-rank test (Hollander, Wolfe, and Chicken 1999) is applied on the obtained results to check the statistical superiority of the Additive-ARIMA-ANN method over other methods. Table 4 presents the test results on monthly onion retail price time series. It can be observed from Table 4 that the Additive-ARIMA-ANN method provides statistically superior RMSE, SMAPE, and MAE than other methods considered in this study. To show

Table 3. Mean RMSE, SMAPE, and MAE of different methods by considering monthly onion retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	2.7783	7.7029	2.0578
ETS	3.9129	11.8140	3.0765
SVM	2.7982	10.3260	2.3043
LSTM	5.1692	17.2880	4.3021
MLP	3.5318	10.4960	2.6794
Additive-ARIMA-ANN	**1.9925**	**6.6665**	**1.5696**
Additive-ARIMA-LSTM	2.1539	7.1808	1.7176
Additive-ARIMA-SVM	2.3458	7.9974	1.9239
Multiplicative-ARIMA-ANN	2.5378	7.7604	1.9591
Multiplicative-ARIMA-LSTM	3.1057	8.4623	2.2754
Multiplicative-ARIMA-SVM	2.6647	8.0782	2.0497
Additive-ETS-ANN	3.9761	12.231	3.1693
Additive-ETS-LSTM	3.9402	11.748	3.0764
Additive-ETS-SVM	2.7423	8.6255	2.1834
Multiplicative-ETS-ANN	4.005	12.799	3.2268
Multiplicative-ETS-LSTM	4.3881	12.996	3.4454
Multiplicative-ETS-SVM	2.88	10.491	2.3967

Table 4. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈),inferior (–) or superior (+) method with respect to the Additive-ARIMA-ANN method by considering monthly onion retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	–	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-LSTM	–	–	–
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Multiplicative-ARIMA-SVM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–
Multiplicative-ETS-SVM	–	–	–

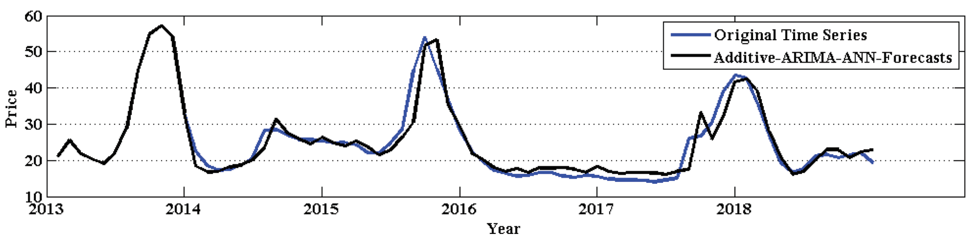

Figure 3. Prediction of onion retail price time series.

the closeness of original time series and Additive-ARIMA-ANN forecasts, we have plotted the comparison graph for onion retail price time series and presented it in Figure 3. Furthermore, the correlation plot as in Figure 4 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.98.

Table 5 presents the mean RMSE, SMAPE, and MAE over fifty independent simulations on monthly onion wholesale price time series. It can be observed from Table 5 that the Additive-ARIMA-ANN method provides the best RMSE and MAE in onion wholesale price time series. Although the Additive-ARIMA -LSTM method provides the best SMAPE, it can be observed from Table 6 (showing the Wilcoxon signed-rank test results) that the SMAPE of Additive-ARIMA-LSTM is statistically equivalent to Additive-ARIMA-ANN method. It can also be seen from Table 6 that the MAE obtained by Additive-ARIMA-ANN and Additive-ARIMA-LSTM are statistically equivalent to each other. However, the Additive-ARIMA-ANN method provides statistically superior RMSE than Additive-ARIMA-LSTM method. Hence, it is concluded that the Additive-ARIMA-ANN method is the best method to forecast the monthly

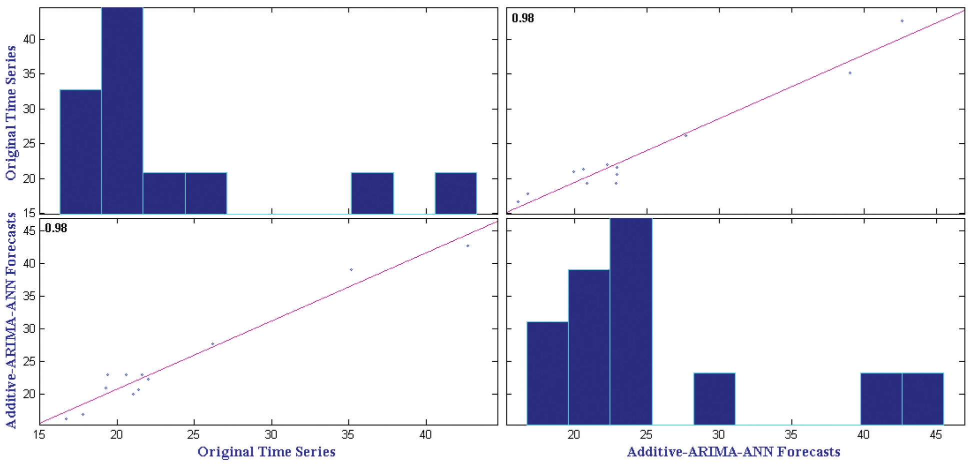

Figure 4. Correlation plot for onion retail price time series.

Table 5. Mean RMSE, SMAPE, and MAE of different methods by considering monthly onion wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	249.2	8.6006	182.32
ETS	365.25	13.546	279.17
SVM	273.93	13.144	231.93
LSTM	515.55	21.677	428.41
MLP	344.46	12.122	254.28
Additive-ARIMA-ANN	**186.64**	8.1184	**148.32**
Additive-ARIMA-LSTM	194.91	**8.0366**	151.03
Additive-ARIMA-SVM	203.13	8.6259	159.78
Multiplicative-ARIMA-ANN	228.65	8.779	173.9
Multiplicative-ARIMA-LSTM	286.19	9.6737	207.05
Multiplicative-ARIMA-SVM	233.63	9.1107	181.4
Additive-ETS-ANN	357.13	13.799	278.99
Additive-ETS-LSTM	352.29	12.932	267.1
Additive-ETS-SVM	244.96	9.5184	187.82
Multiplicative-ETS-ANN	399.41	15.3	312.09
Multiplicative-ETS-LSTM	389.62	14.431	298.81
Multiplicative-ETS-SVM	288.04	14.032	253.82

onion wholesale price time series data. To show the closeness of original time series and Additive-ARIMA-ANN forecasts, we have plotted the comparison graph for onion wholesale price time series and presented it in Figure 5. Furthermore, the correlation plot as in Figure 6 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.98.

Results for Potato Time Series

Table 7 presents the mean RMSE, SMAPE, and MAE over fifty independent simulations on monthly potato retail price time series. It can be observed from Table 7 that the Multiplicative-ETS-SVM method provides the best RMSE,

Table 6. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈),inferior (–) or superior (+) method with respect to the Additive-ARIMA-ANN method by considering monthly onion wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	–	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-LSTM	–	≈	≈
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Multiplicative-ARIMA-SVM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–
Multiplicative-ETS-SVM	–	–	–

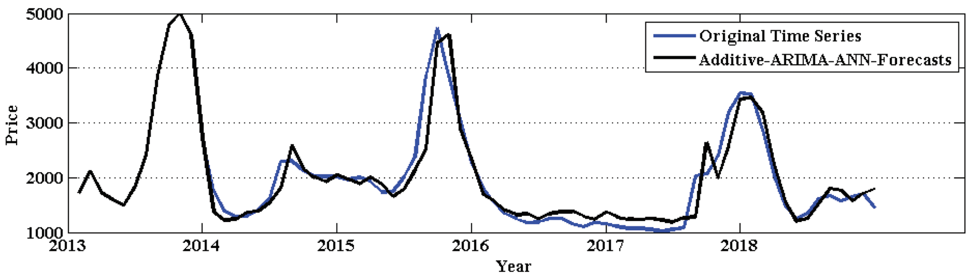

Figure 5. Prediction of onion wholesale price time series.

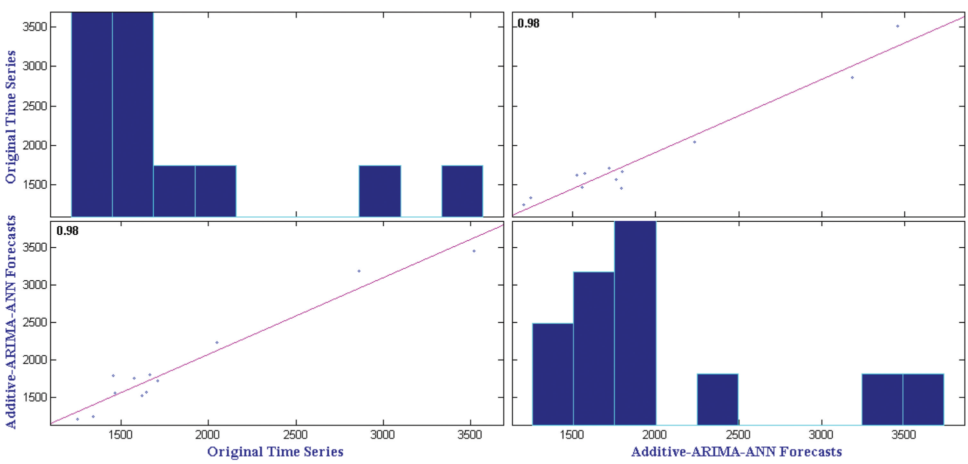

Figure 6. Correlation plot for onion wholesale price time series.

Table 7. Mean RMSE, SMAPE, and MAE of different methods by considering monthly potato retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	**0.8400**	3.5161	0.6538
ETS	1.1386	4.4751	0.80872
SVM	1.3303	6.3573	1.1819
LSTM	1.4690	6.8677	1.2507
MLP	1.0891	4.7420	0.86817
Additive-ARIMA-ANN	0.9411	4.2589	0.79655
Additive-ARIMA-LSTM	0.8759	3.5410	0.66502
Additive-ARIMA-SVM	0.8586	3.5706	0.65439
Multiplicative-ARIMA-ANN	0.9068	4.0389	0.75795
Multiplicative-ARIMA-LSTM	0.8675	3.5770	0.66989
Multiplicative-ARIMA-SVM	0.8692	3.6460	0.66631
Additive-ETS-ANN	0.9880	4.3372	0.79689
Additive-ETS-LSTM	1.0898	4.7722	0.87243
Additive-ETS-SVM	0.8986	3.5601	0.65508
Multiplicative-ETS-ANN	0.9173	4.1297	0.75947
Multiplicative-ETS-LSTM	1.0048	4.4252	0.80979
Multiplicative-ETS-SVM	**0.8400**	**3.3057**	**0.60497**

Table 8. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈), inferior (–) or superior (+) method with respect to the multiplicative-ETS-SVM method by considering monthly potato retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	≈	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-LSTM	–	–	–
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Multiplicative-ARIMA-SVM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–
Multiplicative-ETS-SVM	–	–	–

SMAPE, and MAE in potato retail price time series. However, the Wilcoxon signed-rank test (Hollander, Wolfe, and Chicken 1999) is applied on the obtained results to check the statistical superiority of the Multiplicative-ETS-SVM method over other methods. Table 8 presents the test results on monthly potato retail price time series. It can be observed from Table 8 that the ARIMA method provides statistically equivalent RMSE to Multiplicative-ETS-SVM method, and in all other cases, the Multiplicative-ETS-SVM method provides statistically superior forecasting accuracy. To show the closeness of original time series and Multiplicative-ETS-SVM forecasts, we have plotted the

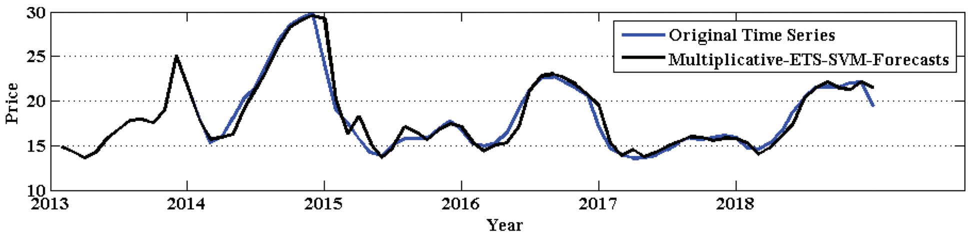

Figure 7. Prediction of potato retail price time series.

comparison graph for potato retail price time series and presented it in

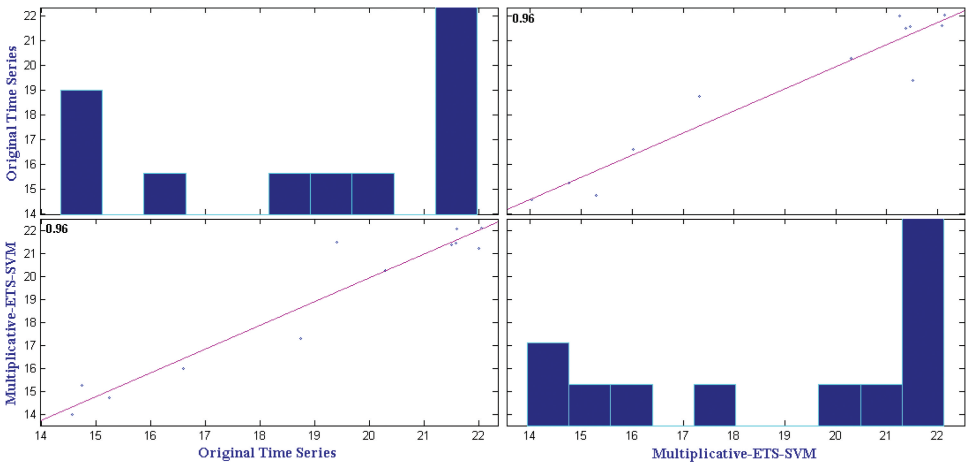

Figure 8. Correlation plot for potato retail price time series.

Table 9. Mean RMSE, SMAPE, and MAE of different methods by considering monthly potato wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	**78.5980**	4.2666	59.2300
ETS	103.7900	5.2409	71.79100
SVM	154.8200	10.0470	141.2400
LSTM	136.2600	8.3813	114.5200
MLP	151.5600	9.1835	127.2600
Additive-ARIMA-ANN	86.9030	4.8943	68.9760
Additive-ARIMA-LSTM	83.0390	**4.2100**	58.8610
Additive-ARIMA-SVM	84.4520	4.4228	59.9850
Multiplicative-ARIMA-ANN	91.0920	5.0700	72.1550
Multiplicative-ARIMA-LSTM	83.2980	4.2490	59.5020
Multiplicative-ARIMA-SVM	83.7010	4.4715	60.7520
Additive-ETS-ANN	97.9700	5.6350	77.3250
Additive-ETS-LSTM	98.8600	5.6232	77.2200
Additive-ETS-SVM	86.2830	4.3337	59.2610
Multiplicative-ETS-ANN	93.8260	5.4763	75.4570
Multiplicative-ETS-LSTM	93.8920	5.4303	74.4250
Multiplicative-ETS-SVM	81.7310	**4.2100**	**56.9820**

Table 10. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈), inferior (–) or superior (+) method with respect to the multiplicative-ETS-SVM method by considering monthly potato wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	+	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-ANN	–	–	–
Additive-ARIMA-LSTM	–	≈	–
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Multiplicative-ARIMA-SVM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–

Figure 7. Furthermore, the correlation plot as in Figure 8 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.96.

Table 9 presents the mean RMSE, SMAPE, and MAE over fifty independent simulations on monthly potato wholesale price time series. It can be observed from Table 9 that the Multiplicative-ETS-SVM method provides the best SMAPE and MAE in potato wholesale price time series. The ARIMA method provides the best RMSE and the Additive-ARIMA-LSTM method provides same SMAPE to that of Multiplicative-ETS-SVM method. For a statistical evaluation of the obtained results, we have applied Wilcoxon signed-rank test and presented the test results in Table 10. It can be observed from Table 10 that the ARIMA method provides statistically superior RMSE than the Multiplicative-ETS-SVM method and the Additive-ARIMA-LSTM method provides statistically equivalent SMAPE to Multiplicative-ETS-SVM method. In all other cases, the Multiplicative-ETS-SVM method provides statistically superior forecasting accuracy than other methods considered in this study. Hence, it is concluded that the Multiplicative-ETS-SVM method is the best method to forecast the monthly potato wholesale price time series data. To

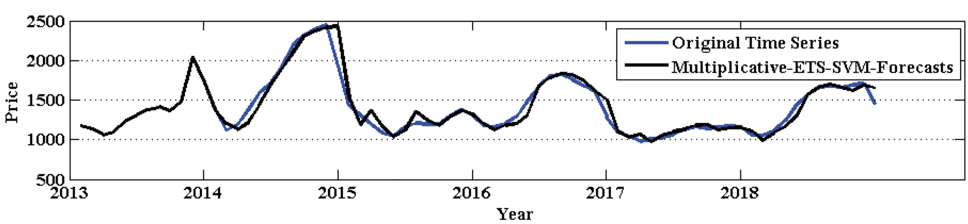

Figure 9. Prediction of potato wholesale price time series.

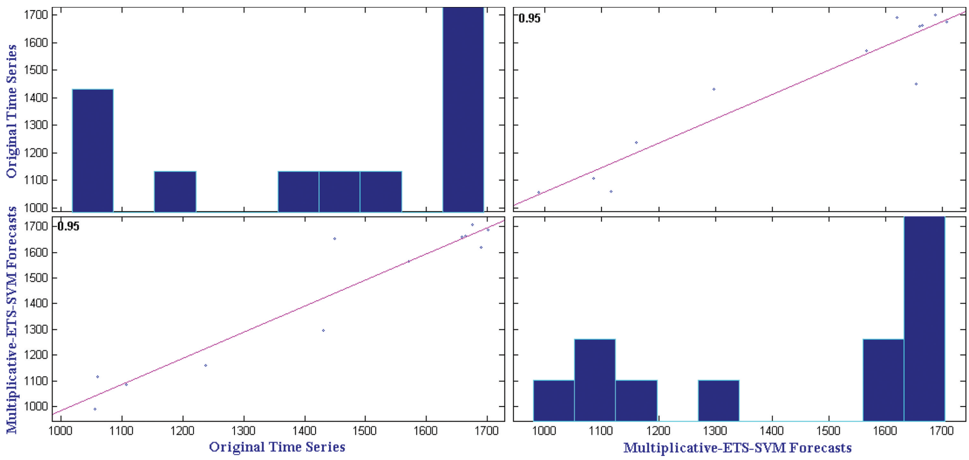

Figure 10. Correlation plot for potato wholesale price time series.

show the closeness of original time series and Multiplicative-ETS-SVM forecasts, we have plotted the comparison graph for potato wholesale price time series and presented it in Figure 9. Furthermore, the correlation plot as in Figure 10 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.95.

Results for Tomato Time Series

Table 11 presents the mean RMSE, SMAPE, and MAE over fifty independent simulations on monthly tomato retail price time series. It can be observed from Table 11 that the Multiplicative-ARIMA-SVM method provides the best RMSE and MAE in tomato retail price time series. It is also observed that the

Table 11. Mean RMSE, SMAPE, and MAE of different methods by considering monthly tomato retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	4.5145	14.1620	3.6818
ETS	5.9770	15.1690	4.3196
SVM	5.1230	20.0820	4.5354
LSTM	8.2514	24.6500	6.6708
MLP	5.7228	21.9840	4.8663
Additive-ARIMA-ANN	4.0699	15.4800	3.4187
Additive-ARIMA-LSTM	3.6198	15.2230	3.2199
Additive-ARIMA-SVM	3.7337	15.9530	3.0920
Multiplicative-ARIMA-ANN	6.0732	15.0760	3.9029
Multiplicative-ARIMA-LSTM	4.0178	14.1080	3.3911
Multiplicative-ARIMA-SVM	**3.0255**	12.6740	**2.6281**
Additive-ETS-ANN	6.7453	24.6690	5.7098
Additive-ETS-LSTM	5.3224	17.4290	4.0341
Additive-ETS-SVM	3.7989	13.1890	3.0637
Multiplicative-ETS-ANN	7.0075	17.6070	4.6684
Multiplicative-ETS-LSTM	5.0255	16.4910	4.0180
Multiplicative-ETS-SVM	3.5524	**12.6210**	2.8360

Table 12. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈), inferior (–) or superior (+) method with respect to the multiplicative-ARIMA-SVM method by considering monthly potato retail price time series data.

	RMSE	SMAPE	MAE
ARIMA	–	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-ANN	–	–	–
Additive-ARIMA-LSTM	–	–	–
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–
Multiplicative-ETS-SVM	–	+	–

Multiplicative-ETS-SVM method provides the best SMAPE in tomato retail price time series. Additionally, the Wilcoxon signed-rank test (Hollander, Wolfe, and Chicken 1999) is applied on the obtained results to check the statistical superiority of the Multiplicative-ARIMA-SVM method over other methods. Table 12 presents the test results on monthly tomato retail price time series. It can be observed from Table 12 that the Multiplicative-ARIMA-SVM method provides statistically superior RMSE, SMAPE, and MAE than other methods except Multiplicative-ETS-SVM method in SMAPE measure. Hence, it is concluded that the Multiplicative-ARIMA-SVM method is the best method to forecast the monthly tomato wholesale price time series data. To show the closeness of original time series and Multiplicative-ARIMA-SVM forecasts, we have plotted the comparison graph for tomato retail price time series and presented it in Figure 11. Furthermore, the correlation plot as in Figure 12 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.83.

Table 13 presents the mean RMSE, SMAPE, and MAE over fifty independent simulations on monthly onion wholesale price time series. It can be observed from Table 13 that the Multiplicative-ETS-SVM method

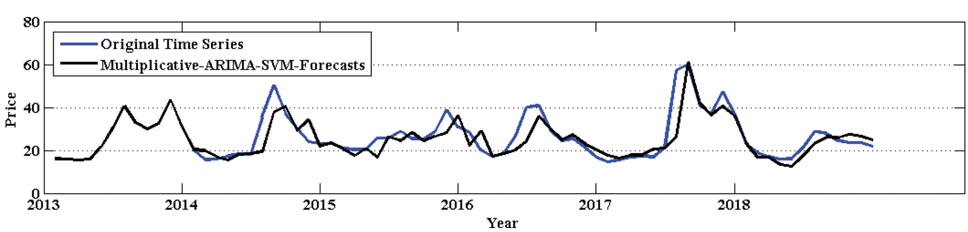

Figure 11. Prediction of tomato retail price time series.

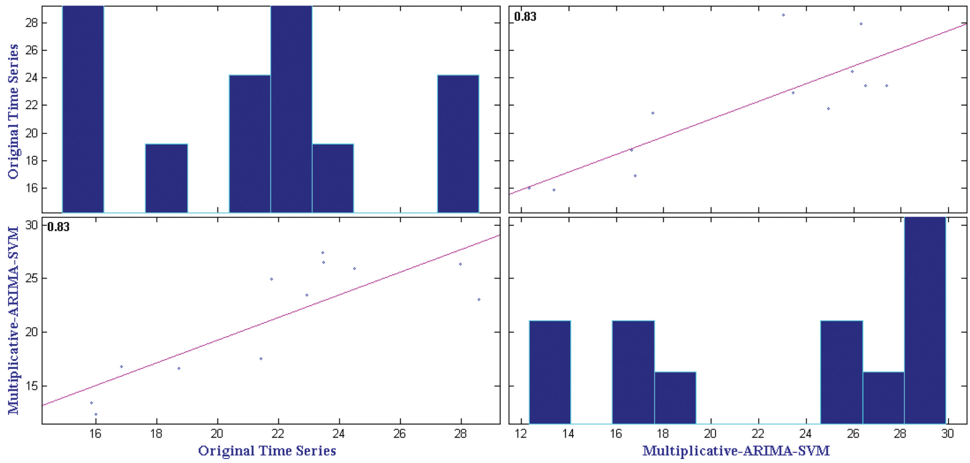

Figure 12. Correlation plot for tomato retail price time series.

Table 13. Mean RMSE, SMAPE, and MAE of different methods by considering monthly tomato wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	400.0200	16.8790	328.9500
ETS	507.7100	16.9180	365.9000
SVM	438.9900	23.1870	390.4100
LSTM	781.1400	29.9020	629.2300
MLP	566.8500	29.3460	481.0300
Additive-ARIMA-ANN	392.0400	21.0450	321.7500
Additive-ARIMA-LSTM	322.2400	16.3530	279.2300
Additive-ARIMA-SVM	345.5600	22.8080	298.7200
Multiplicative-ARIMA-ANN	504.7200	19.5470	355.5100
Multiplicative-ARIMA-LSTM	368.0300	16.9300	307.4300
Multiplicative-ARIMA-SVM	273.8500	16.3120	250.9600
Additive-ETS-ANN	476.2300	24.5200	396.5200
Additive-ETS-LSTM	449.4000	20.0960	340.1400
Additive-ETS-SVM	296.0000	14.1520	236.6200
Multiplicative-ETS-ANN	544.0500	19.4550	377.9800
Multiplicative-ETS-LSTM	457.0800	20.3600	372.5100
Multiplicative-ETS-SVM	**266.6700**	**13.6090**	**221.1500**

provides the best RMSE, SMAPE, and MAE in tomato wholesale price time series. However, for a statistical significance test of the obtained results, we have employed Wilcoxon signed-rank test (Hollander, Wolfe, and Chicken 1999) and presented the test results in Table 14. It can also be observed from Table 14 that the Multiplicative-ETS-SVM method provides statistically superior RMSE, SMAPE, and MAE than all other methods considered in this study. Hence, it is concluded that the Multiplicative-ETS-SVM method is the best method to forecast the monthly tomato wholesale price. To show the closeness of original time series and Multiplicative-ETS-SVM forecasts, we have plotted the comparison graph for tomato

Table 14. Wilcoxon signed-rank test results using forecasting accuracies indicating the equivalent (≈), inferior (–) or superior (+) method with respect to the Multiplicative-ETS-SVM method by considering monthly tomato wholesale price time series data.

	RMSE	SMAPE	MAE
ARIMA	–	–	–
ETS	–	–	–
SVM	–	–	–
LSTM	–	–	–
MLP	–	–	–
Additive-ARIMA-ANN	–	–	–
Additive-ARIMA-LSTM	–	≈	≈
Additive-ARIMA-SVM	–	–	–
Multiplicative-ARIMA-ANN	–	–	–
Multiplicative-ARIMA-LSTM	–	–	–
Multiplicative-ARIMA-SVM	–	–	–
Additive-ETS-ANN	–	–	–
Additive-ETS-LSTM	–	–	–
Additive-ETS-SVM	–	–	–
Multiplicative-ETS-ANN	–	–	–
Multiplicative-ETS-LSTM	–	–	–

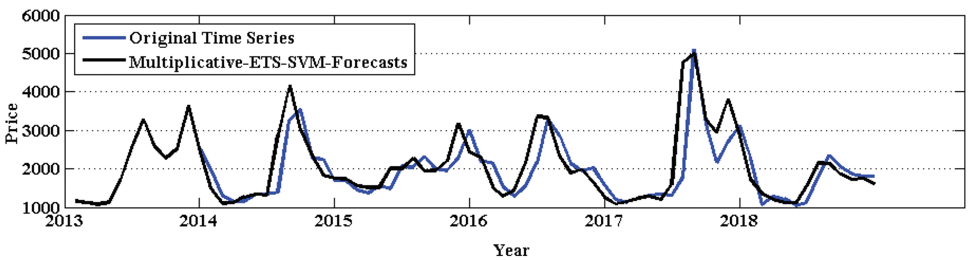

Figure 13. Prediction of tomato wholesale price time series.

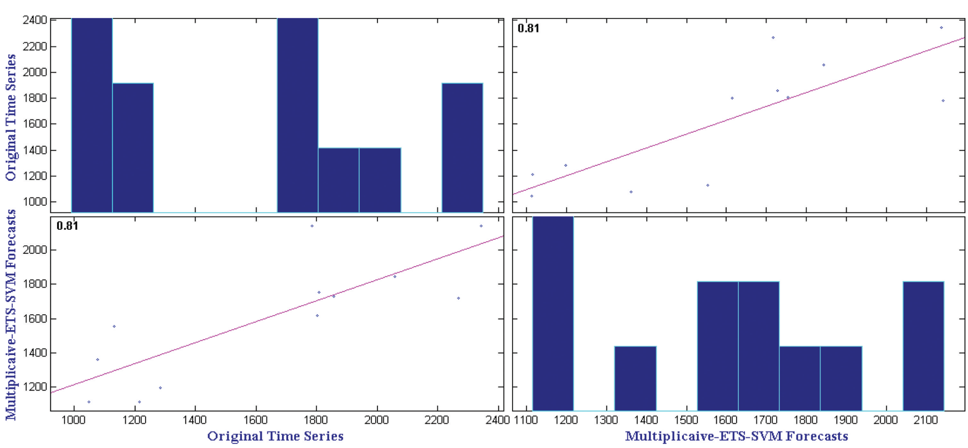

Figure 14. Correlation plot for tomato wholesale price time series.

wholesale price time series and presented it in Figure 13. Furthermore, the correlation plot as in Figure 14 shows a strong correlation between predicted and observed values with a correlation coefficient of 0.81.

Conclusion

In this paper, we have considered three most prominent horticultural commodities of India, namely, tomato, potato, and onion, and tried to efficiently predict their prices by using the most popular statistical (ARIMA, ETS), machine learning (MLP, SVM, LSTM), and hybrid methods. In addition to the existing hybrid methods (Additive-ARIMA-ANN [5], Multiplicative-ARIMA-ANN [8], Additive-ETS-ANN [10], Additive-ARIMA-SVM), we have proposed and considered Additive-ETS-SVM, Additive-ETS-LSTM, Multiplicative-ARIMA-SVM, Multiplicative-ARIMA-LSTM, Multiplicative-ETS-ANN, Multiplicative-ETS-SVM, and Multiplicative-ETS-LSTM methods to forecast the monthly retail and wholesale TOP prices. It is observed that none of the methods provide the best forecasts in all the considered crop price time series data. Hence, different methods need to be considered for different crop price time series data. Simulation results reveal that the hybrid methods provide better result than the individual models in predicting the crop prices. The Additive-ARIMA-ANN [5] method provides the best forecasts in monthly retail and wholesale onion price; the proposed Multiplicative-ETS-SVM method provides the best forecasts in monthly retail and wholesale potato price; the proposed Multiplicative-ARIMA-SVM method provides the best forecasts in monthly retail tomato price; the proposed Multiplicative-ETS-SVM method provides the best forecasts in monthly wholesale tomato price. Additionally, the proposed hybrid methods can be applied to forecast other time series data.

Acknowledgments

The authors would like to thank Dr. Prabhat Kumar Sahu, Reader, Sambalpur University, Odisha, India for providing the computational resources.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Sourav Kumar Purohit http://orcid.org/0000-0002-3343-6818
Sibarama Panigrahi http://orcid.org/0000-0002-1409-3856
Prabira Kumar Sethy http://orcid.org/0000-0003-3477-6715

Santi Kumari Behera http://orcid.org/0000-0003-4857-7821

References

Babu, C. N., and B. E. Reddy. 2014. A moving-average filter-based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing* 23:27–38. doi:10.1016/j.asoc.2014.05.028.

Faruk, D. Ö. 2010. A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence* 23 (4):586–94. doi:10.1016/j.engappai.2009.09.015.

Hollander, M., D. A. Wolfe, and E. Chicken. 1999. *Nonparametric statistical methods*. Hoboken: Wiley. doi:10.1002/9781119196037.biblio.

Hyndma, R. J., and Y. Khandakar. 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27 (3). doi: 10.18637/jss.v027.i03.

Khashei, K., and B. Mehdi. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing* 11 (2):2664–75. doi:10.1016/j.asoc.2010.10.015.

Kumar, A., C. Bantilan, P. Kumar, S. Kumar, and S. Jee. 2012. Food security in India: Trends, patterns and determinants. *Indian Journal of Agricultural Economics* 67:445–63. doi:10.22004/ag.econ.204827.

Kumar, P., and P. K. Joshi. 2016. Food demand and supply projections to 2030: India. In *International trade and food security: The future of Indian agriculture*, eds. Floor Brouwer and P.K. Joshi. Part 2, Chapter 4: 29–63.

Panigrahi, S., and H. S. Behera. 2017. A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence* 66:49–59. doi:10.1016/j.engappai.2017.07.007.

Wang, L., H. Zou, J. Su, L. Li, and S. Chaudhry. 2013. An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioral Science* 30 (3):244–59. doi:10.1002/sres.2179.

Yao, Q., and P. J. Brockwell. 2006. Gaussian maximum likelihood estimation for ARMA Models. I. Time series. *Journal of Time Series Analysis* 27 (6):857. doi:10.1111/j.1467-9892.2006.00492.x.

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–75. doi:10.1016/S0925-2312(01)00702-0.