Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(3) 2023 275–284

Comparative Study of ARIMA, SARIMA and Hybrid (ARIMA + ANN and SARIMA + ANN) Models for Wholesale Monthly Average Price of Tomato and Onion

Sanjeev, Pushpa, Vikram, Preeti and Pawan Kumar

Chaudhary Charan Singh Haryana Agricultural University, Hisar
Maharana Pratap Horticultural University, Karnal

Received 02 February 2023; Revised 22 August 2023; Accepted 18 September 2023

SUMMARY

Time series price forecasting is an important area of forecasting in which past observations of the same variable are collected and analysed to develop a model describing the underlying relationship. In this paper, to compare the forecasting performance of ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal Autoregressive Integrated Moving Average) hybrid (ARIMA + ANN (Artificial Neuron Network) and SARIMA + ANN) techniques for all India wholesale monthly average price time series of tomato and onion crop. The ARIMA and SARIMA techniques are used to capture the linear pattern of data. The ANN technique is used to capture the nonlinear patterns of the residuals obtain from ARIMA and SARIMA techniques. Empirical results indicate that hybrid (SARIMA + ANN) technique is effective way to improve the forecasting performance for price series of tomato and onion crop on the basis of least value of error measure such as RD (%), RMSE, MAPE and MAE.

Keywords: ARIMA; SARIMA; Hybrid; Time series price forecasting; Linear and non-linear patterns.

1. INTRODUCTION

One of the crucial aspects of time series modelling is price forecasting, is a dynamic research area which has been devoted over past few decades. The main aim of time series modelling is to carefully collect and rigorously study the past observations of a variable to develop an appropriate model which describe the inherent structure of the time series. This model is used to forecast the future values of time series. The Box-Jenkins ARIMA (1970) model is one of most significant and often employed in time series forecasting. ARIMA model is incapable to capture seasonality and nonlinearity that are present in price time series. Seasonal ARIMA technique is more than ARIMA suitable for forecasting if the price series contain seasonality.ANN may provide an effective alternative for overcome limitation of ARIMA technique, has used to capture the complex economic relationships with a variety of patterns as they serve as a powerful and flexible computational tool.

Corresponding author: Pushpa
E-mail address: pushpaghiyal@gmail.com

In this paper, hybrid techniques i.e, ARIMA+ANN and SARIMA+ANN techniques used to capture the linear, seasonal and non-linear pattern in all India wholesale monthly average price series of tomato and onion. The ARIMA and SARIMA models can capture the linear and seasonal patterns. The residual obtained from ARIMA and SARIMA technique is contain only the nonlinear patterns. The Artificial Neural Network technique is used to model the nonlinear patterns of the residuals. Hybrid technique can increase the chance to capture different patterns present in time series data and improve the forecasting performance. Several empirical studies have already suggested that by combining several different techniques, forecasting accuracy can often be improved over the individual technique.

Tseng *et al.,* (2002) developed a hybrid model combines SARIMA and Back propagation Neural Network for forecasting seasonal time series data. Zhang (2003) developed ARIMA-ANN hybrid model where hybrid model outperformed ARIMA and

ANN. Temizel and Casey (2005) show that combined forecast based on hybrid model (ARIMA and ANN) underperform significantly compare to its constituents' performance. Chandran and Pandey (2007) forecasted the prices of potato for Delhi market using univariate seasonal ARIMA model.Chen *et al.,* (2007) combined SARIMA and Support Vector Machine, was more effective than SARIMA and SVM. Liang (2009) studies a hybrid forecasting method that combines the seasonal ARIMA model and neural networks with genetic algorithms for predicting the production value of the mechanical industry in Taiwan. Rahman (2010) examined the best fitted ARIMA model for efficient forecast of boro rice production in Bangladesh from 2008-09 to 2012-13. Jha and Sinha (2012) compared ARIMA and TDNN model both in terms of modelling and forecasting using monthly wholesale price data of oilseed crops. Adebiyi *et al.,* (2014) examined the forecasting performance of ARIMA and artificial neural networks model with published stock data obtained from New York Stock Exchange. Khashei and Hajirahimi (2018) evaluated the performance of two types of hybrid models for predicting stock prices in order to introduce the more reliable series hybrid mode. Kambo *et al.* (2018) applied Neural Network approaches for price forecasting of agriculture commodity such as vegetables, fruits, cereals *etc.* in both short term and long terms. Choudhary *et al.* (2019) employed an empirical mode decomposition based neural network model for potato price forecasting. Sivamani *et al.,* (2019) worked Seasonal - Autoregressive Integrated Moving Average (SARIMA) model to forecast the food stock requirement in the livestock barn over a simulated data. Ayub and Jafri (2020) investigated the excellence of hybrid ARIMA-ANN model over ANN-ARIMA in forecasting Karachi stock price.

2. MATERIAL AND METHODS

2.1 Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) Techniques

The future value of a variable is presumed to be a linear function of a number of prior observations and random errors in an autoregressive integrated moving average model. This means that the procedure used to create the time series has a certain structure. The functional form of ARIMA (p, d, q) model following:

$$\phi(B)(1-B)^d x_t = \theta(B)\varepsilon_t ,$$

$$\phi(B) = 1-\phi_1 B-\phi_2 B-....-\phi_p B^p$$ (AR non seasonal)

$$\theta(B) = 1+\theta_1 B^1+\theta_2 B^2+....+\theta_q B^q$$ (MA non seasonal)

Where, x_t and ε_t are the actual and random error at time period t, respectively; ϕ (i= 1, 2, …, p) and θ (j= 1, 2, …, q) are model parameters. p for order of autoregressive, q for order of moving average and the d is order of differencing transformation. Random errors ε_t are assumed to be independently and identically distributed with a mean zero and a constant variance of σ^2. SARIMA technique is based on traditional ARIMA technique,widely used for modelling of seasonal time series. There are six main parameters for fitting the SARIMA (p, d, q) (P, D, Q)s model: the order of autoregressive (p) and seasonal autoregressive (P), the order of integration (d) and seasonal integration (D), and the order of moving average (q) and seasonal moving average (Q), and s represents the season period length. The SARIMA (p, d, q) (P, D, Q)s model has the following form:

$$\varphi_P B^s \phi(B)(1-B)^d (1-B^s)^D x_t = \theta(B)\Theta_Q\left(B^s\right)\varepsilon_t ,$$

Where:

$$\phi(B) = 1-\phi_1 B-\phi_2 B-....-\phi_p B^p$$ (AR non seasonal)

$$\varphi_P B^s = 1-\varphi_1 B^s-\varphi_2 B^{2s}-.....-\varphi_P B^{Ps}$$ (AR seasonal)

$$\theta(B) = 1+\theta_1 B^1+\theta_2 B^2+....+\theta_q B^q$$ (MA non seasonal)

$$\theta_Q\left(B^s\right) = 1+\theta_1 B^s+\theta_2 B^{2s}+....+\theta_Q B^{Qs}$$ (MA seasonal)

$(1-B)^d$ = non seasonal differencing

$(1-B^s)^D$ = seasonal differencing

Where B is the backward shift operator, ε_t is the estimated residual at time t and x_t denotes the observed price at time t (t = 1, 2… k). The process is called SARIMA (p, q, d) (P, D, Q)s. Once a good model has been chosen, this three-step process is often repeated multiple times. The final model chosen can be applied to prediction.

In the identification step, check the time series is stationary or not, if time series is not stationary then data transformation is often needed to make the time series stationary. ACF and PACF of stationary series are used in identifying the tentative orders of the ARIMA and SARIMA model.

Estimation step, once a tentative model is specified, then model parameters are estimated such that an overall measure of errors is minimized.

The diagnostic evaluation of the model's appropriateness is the third stage of building the model. The main goal of this is to determine whether the model's assumptions on the errors are satisfied.

2.2 Hybrid Technique

This study used a hybrid technique that combines a linear pattern and a nonlinear pattern.

$$x_t = L_t + N_t + \varepsilon_t$$

Where, x_t is the actual series, L_t is linear pattern, and N_t is the nonlinear pattern. We used ARIMA and SARIMA technique to model the linear pattern of x_t. ANN models used to model the nonlinear pattern of residuals from ARIMA and SARIMA models. Let r_t is residuals from ARIMA and SARIMA modelsat time t, then

$$r_t = x_t - \widehat{L_t}$$

$\widehat{L_t}$. is prediction of the linear pattern. ANN are applied to nonlinear patterns.

$$\hat{r}_t = f\left(r_{t-1}, r_{t-2}, \ldots\ldots r_{t-p}\right)$$

Where f is a nonlinear function and p is the number of input lags. Consequently, the final prediction following as,

$$\widehat{x_t} = \widehat{L_t} + \hat{r}_t + \varepsilon_t$$

where ε_t is combined error of model at time t. Since the linear ARIMA and SARIMA models cannot capture the nonlinear pattern of the series. Nonlinearity is present in the residuals from the linear models. ANN model used to capture the nonlinearity of residuals. Consequently, Hybrid technique is anticipated to take advantage of both model's strengths and features in order to enhance forecasting performance as a whole.

2.3 Models Performance error Measures

The forecasting performance of the ARIMA, SARIMA and Hybrid (ARIMA + ANN and SARIMA + ANN) model is examined in terms of error measures such as Relative deviation percentage (RD%), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are given below.

$$RD\% = \frac{observed - predicted}{observed} \times 100$$

$$RMSE = \sqrt{\frac{1}{n}\left(observed - predicted\right)^2}$$

$$MPAE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{observed - predicted}{observed}\right|$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|observed - predicted\right|$$

3. RESULTS AND DISCUSSION

In this paper, a total number of observations is 156 in all India wholesale monthly average price of tomato crop, is used from Jan-2010 to Dec-2022, is collected from the website https://agmarknet.gov.in. The data series is divided into two parts: training data set of the first 144 observation and validity data set of the last 12 observations.

3.1 ARIMA and SARIMA techniques result

The foremost step in time series analysis is to plot the data and check the occurrence of a trend as well as seasonality. Figs. 1 and 2 show the decomposition of tomato and onion price time series in three parts, namely, trend, seasonal and random. We can see that there is a positive trend and seasonal effect over time which indicates the nonstationary nature of series. This is also shows in ACF and PACF plot of actual series and stationary series in Figs. 3 and 4.

Friedman (1937) and Kruskal-Wallis (1952) tests used to check the seasonality of data. Table 1 shows the statistic and p-value of the Friedman and Kruskal – Wallis tests. Test statistic of both the test are significant at the 5% level of significance indicate that corresponding time series as seasonal.

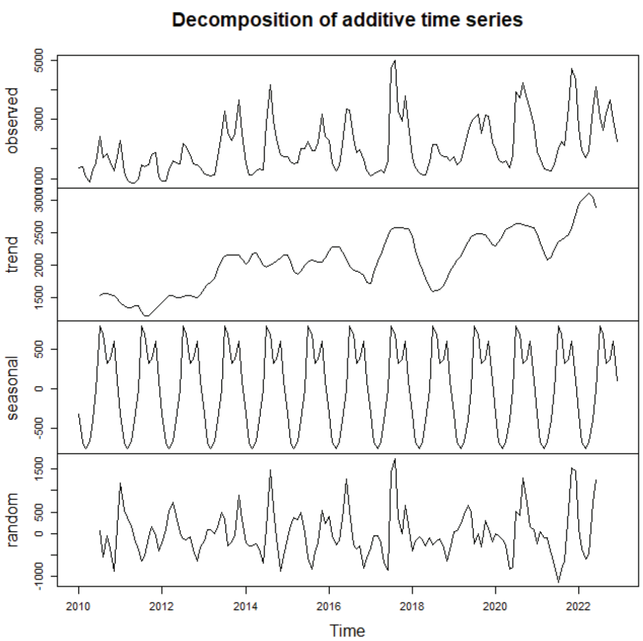

Fig. 1. STL decomposition of wholesale monthly tomato price time series

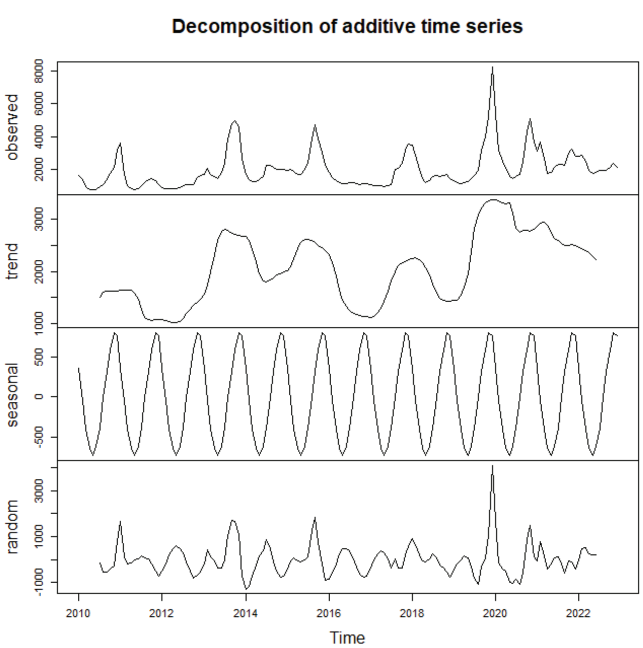

Fig. 2. STL decomposition of wholesale monthly onion price time series

Table 1. Friedman and Kruskal-Wallis seasonality tests for tomato and onion price in India

Series	Friedman		Kruskal-Wallis	
	Statistic	p-value	Statistic	p-value
tomato	52.04	<0.01	50.56	<0.01
onion	52.72	<0.01	57.69	<0.01

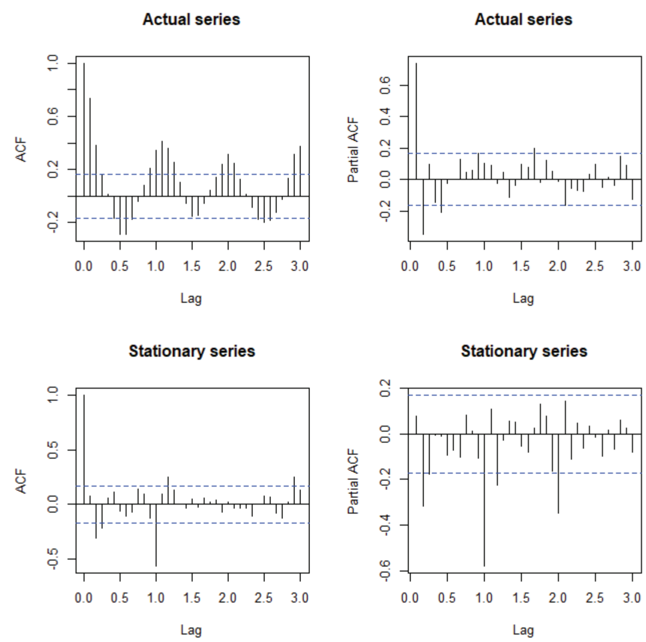

Fig. 3. ACF/PACF plots of actual and stationary series for tomato price series

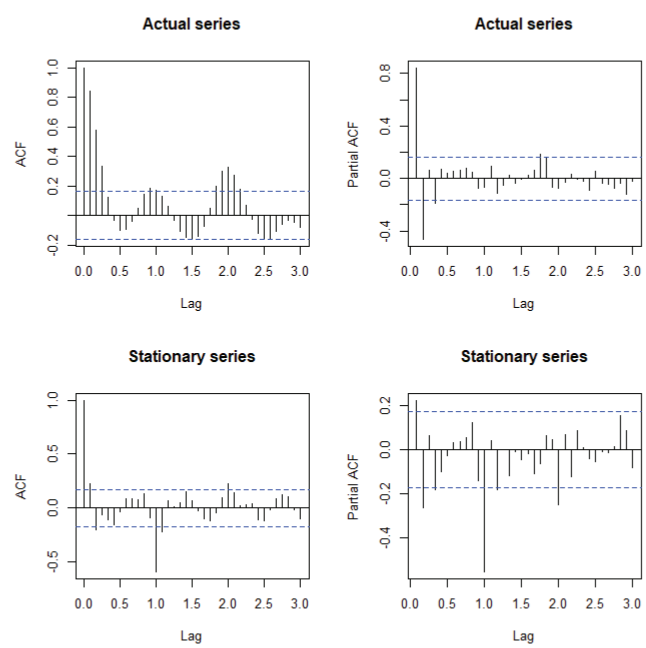

Fig. 4. ACF/PACF plots of actual and stationary series for onion price series

The best fitted ARIMA and SARIMA model for tomato and onion price series based on the smallest AIC information criteria as well as smallest RMSE, MAPE and MAE value. Tables 2 and 3 show the selected ARIMA (1,1,2), SARIMA $(2,1,1)\ (0,1,1)_{12}$ and ARIMA $(0,1,1)$, SARIMA $(1,1,2)\ (0,1,1)_{12}$

models for tomato and onion price series on the basis of least value of AIC, RMSE, MAPE and MAE.

Table 2. Model selection criteria for wholesale price of tomato in India

	AIC	RMSE	MAPE	MAE
ARIMA (p, d, q)				
(0,1,1)	2238.37	595.69	18.03	397.06
(1,1,0)	2242.11	603.64	18.18	398.86
(1,1,1)	2238.65	592.12	18.22	397.49
(2,1,1)	2235.12	581.59	18.63	396.26
(1,1,2)	2217.67	543.64	17.27	366.54
SARIMA (p, d, q) $(P, D, Q)_{12}$				
(1,1,1) $(1,1,0)_{12}$	2068.83	584.22	20.42	414.34
(1,1,1) $(0,1,1)_{12}$	2036.78	472.52	16.28	333.33
(1,1,0) $(1,1,1)_{12}$	2027.65	443.69	15.40	299.73
(0,1,1) $(1,1,1)_{12}$	2024.01	436.41	16.37	304.81
(2,1,1) $(0,1,1)_{12}$	2020.34	431.81	15.09	255.21

Table 3. Model selection criteria for wholesale price of onion in India

	AIC	RMSE	MAPE	MAE
ARIMA (p, d, q)				
(0,1,1)	2236.00	580.40	12.02	304.31
(1,1,0)	2244.86	609.28	12.57	315.77
(1,1,1)	2237.39	589.12	13.09	312.39
(1,1,2)	2242.82	593.01	13.33	312.60
(2,1,0)	2239.18	592.93	13.06	313.70
SARIMA (p, d, q) $(P, D, Q)_{12}$				
(1,1,1) $(1,1,0)_{12}$	2081.95	614.58	17.36	372.84
(1,1,1) $(0,1,1)_{12}$	2073.36	545.95	14.12	309.88
(0,1,1) $(1,1,1)_{12}$	2074.27	579.86	14.23	315.66
(1,1,2) $(1,1,0)_{12}$	2078.16	595.03	16.04	352.07
(1,1,2) $(0,1,1)_{12}$	2062.04	507.57	13.70	298.86

In the estimation stage, the models with lowest values of AIC, RMSE, MAPE and MAE are concluded to be the better estimation model. Parameters estimate of selected ARIMA and SARIMA models for tomato and onion price series are shown in Tables4 and 5 respectively. Because the coefficients were significant at less than 1% level of significance during the parameter estimation step, the Ljung-Box test was used to examine the residuals. Table 6 shows the test statistic of Ljung-Box was significant at 5% level of significance indicating that residuals were white noise. Hence, it can be conclude that the ARIMA (1, 1, 2) and SARIMA (1, 1, 2) $(1, 1, 1)_{12}$, ARIMA (0, 1, 1) and SARIMA (1, 1, 2) (0, 1, 1) models were appropriate for forecasting the tomato and onion respectively.

Table 4. Parameter estimate of the selected SARIMA Model for in sample data set of tomato

	Estimate	S.E.	z value	p value
ARIMA (1, 1, 2)				
AR1	0.57	0.09	6.60	<0.01
MA1	-0.52	0.08	-6.19	<0.01
MA2	-0.44	0.07	-5.63	<0.01
SARIMA (2, 1, 1) $(0, 1, 1)_{12}$				
AR1	0.92	0.08	11.11	<0.01
AR2	-0.33	0.08	-3.85	<0.01
MA1	-0.98	0.03	-26.06	<0.01
SMA1	-0.95	0.11	-9.36	<0.01

Table 5. Parameter estimate of the selected SARIMA Model for in sample data set of onion

	Estimate	S.E.	z value	p value
ARIMA (0, 1, 1)				
MA1	0.49	0.08	6.42	<0.01
SARIMA (1, 1, 2) $(0, 1, 1)_{12}$				
AR1	0.67	0.07	8.94	<0.01
MA1	-0.47	0.09	-4.88	<0.01
MA2	-0.52	0.08	-5.93	<0.01
SMA1	-0.95	0.23	-4.25	<0.01

Table 6. Ljung-Box tests for residuals

Model	Q Statistic	df	P value
tomato			
ARIMA (1, 1, 2)	16.61	21	0.25
SARIMA (1, 1, 2) $(1, 1, 1)_{12}$	30.02	22	0.31
onion			
ARIMA (0, 1, 1)	18.23	23	0.41
SARIMA (1, 1, 2) (0, 1, 1)	25.59	20	0.74

Brock-Dechert-Scheinkman (BDS) test (Brock *et al.,* 1996) is used to determine whether data are nonlinear. The results of BDS test shown in Tables 7 and 8 indicated that some test statistic is significantat 5% level of significance while other are not. As a result, the BDS test results showed that some portion of residuals series are nonlinear.

Table 7. Brock- Dechert-Scheinkman (BDS) test for residuals from ARIMA and SARIMA models for tomato

Embedding dimension			
2		3	
Statistic	p value	Statistic	p value
ARIMA			
2.84	<0.01	4.75	<0.01
2.89	<0.01	3.44	<0.01
2.19	0.03	1.32	0.18
0.18	0.31	0.37	0.27
SARIMA			
3.18	<0.01	3.35	<0.01
3.30	<0.01	4.49	<0.01
2.29	0.02	3.06	<0.01
0.36	0.71	1.31	0.18

Table 8. Brock-Dechert-Scheinkman (BDS) test for residuals from ARIMA and SARIMA models for onion

Embedding dimension			
2		3	
Statistic	p value	Statistic	p value
ARIMA			
4.14	<0.01	6.28	<0.01
1.11	0.26	2.96	<0.01
1.26	0.20	2.59	<0.01
1.95	0.04	2.52	<0.01
SARIMA			
7.79	<0.01	9.82	<0.01
5.17	<0.01	6.15	<0.01
2.75	<0.01	3.49	<0.01
3.29	<0.01	3.40	<0.01

3.2 ANN technique for residual series

ANN model is used to deal with non-linearity pattern are found in residual series. Before applying ANN model, residual series is splitting into three data set, namely, training, testing and validity set. ANN model used here is a three-layered feed forward neural network, trained with training data set, using the gradient descent back propagation algorithm with a learning rate of 0.001 and threshold of 0.01. Tables 9 and 10 showed RMSE and MAPE values of fitted ANN models. ANN (12-4-1) and ANN (12-7-1), ANN (12-5-1) and ANN (12-3-1) models are selected as appropriate models on the based minimum value of RMSE and MAPE of testing set to capture nonlinearity of residuals from ARIMA and SARIMA for price of tomato and onion respectively.

Table 9. RMSE and MAPE values of the ANN model on tomato

Input Layer	Number of Artificial neurons	Output Layer	Training set		Testing set	
			RMSE	MAPE	RMSE	MAPE
Residuals from ARIMA						
12	3	1	103.45	42.52	49.56	30.71
12	4	1	65.17	11.09	16.03	8.29
12	5	1	90.81	15.70	37.87	11.84
Residuals from SARIMA						
12	5	1	30.85	28.65	23.17	9.19
12	6	1	44.23	22.20	25.99	9.61
12	7	1	22.89	19.25	20.26	9.10

Table 10. RMSE and MAPE values of the ANN model on onion

Input Layer	Number of Artificial neurons	Output Layer	Training set		Testing set	
			RMSE	MAPE	RMSE	MAPE
Residuals from ARIMA						
12	3	1	94.60	153.65	42.57	13.27
12	4	1	82.57	76.38	27.74	10.73
12	5	1	55.43	16.78	21.08	6.75
Residuals from SARIMA						
12	3	1	16.25	71.07	5.14	2.73
12	4	1	18.74	75.22	8.71	3.82
12	5	1	36.25	81.17	7.14	5.73

These four statistical error measures such as RD (%), RMSE, MAPE and MAE are used to compare the forecasting performance of selected best fitted ARIMA, SARIMA and Hybrid (ARIMA+ANN and SARIMA+ANN) models given in Tables 11, 12 and 13. Tables 11 and 12 show the performance of corresponding selected ARIMA, SARIMA and hybrid (ARIMA+ANN and SARIMA+ANN) models in terms of relative deviation (RD (%)) for price of tomato and onion respectively. Table 13 shows that a hybrid SARIMA+ANN technique out performed over ARIMA, SARIMA and Hybrid ARIMA+ANN techniques for all India wholesale monthly average price of tomato and onion crop on the basis of least value of RMSE, MAPE and MAE of validity set.

Table 11. Actual, predicted and RD (%)valuesfrom best fitted different techniques for validity set of tomato price series

Month	Actual	ARIMA		SARIMA		ARIMA+ANN		SARIMA+ANN	
		Predicted	RD(%)	Predicted	RD(%)	Predicted	RD(%)	Predicted	RD(%)
Jan-22	2714.39	3559.26	-31.13	3618.15	-33.30	3024.97	-11.44	3014.97	-11.07
Feb-22	1949.69	3083.79	-58.17	2716.03	-39.31	2207.31	-13.21	2137.31	-9.62
Mar-22	1689.83	2810.52	-66.32	2455.36	-45.30	1989.59	-17.74	1980.59	-17.21
Apr-22	1951.86	2653.46	-35.95	2397.00	-22.81	2302.61	-17.97	2202.61	-12.85
May-22	3372.95	2563.20	24.01	2609.11	22.65	2586.82	23.31	2786.82	17.38
Jun-22	4115.73	2511.32	38.98	2957.21	28.15	2884.70	29.91	3284.70	20.19
Jul-22	3084.5	2481.50	19.55	4021.73	-30.39	2508.28	18.68	2808.28	8.95
Aug-22	2631.38	2464.37	6.35	3793.69	-44.17	2385.75	9.33	2285.75	13.13
Sep-22	3211.89	2454.52	23.58	3323.76	-3.48	2440.22	24.03	2490.22	22.47
Oct-22	3644.46	2448.86	32.81	2893.75	20.60	2592.00	28.88	2622.00	28.06
Nov-22	2983.54	2445.60	18.03	2709.76	9.18	2608.73	12.56	2618.73	12.23
Dec-22	2247.1	2443.73	-8.75	2167.96	3.52	2116.66	5.80	2016.66	10.25

Table 12. Actual, predicted and RD (%) values from best fitted different techniques for validity set of onion price series

Month	Actual	ARIMA		SARIMA		ARIMA+ANN		SARIMA+ANN	
		Predicted	RD(%)	Predicted	RD(%)	Predicted	RD(%)	Predicted	RD(%)
Jan-22	2714.39	2721.05	2.86	2414.18	13.81	2811.51	1.63	3056.52	-9.12
Feb-22	1949.69	2675.34	6.40	2204.95	22.85	2694.46	-6.00	2925.40	-2.35
Mar-22	1689.83	2659.35	-4.61	1934.62	23.90	2311.53	-16.21	2671.92	-5.11
Apr-22	1951.86	2553.75	-28.38	1427.15	28.25	2234.26	-25.28	2389.37	-20.12
May-22	3372.95	2651.79	-48.69	1353.95	24.08	2141.86	-16.67	2049.66	-14.93
Jun-22	4115.73	2656.10	-44.69	1474.36	19.69	2015.04	-3.21	2137.81	-16.45
Jul-22	3084.5	2550.86	-30.66	1718.00	12.00	2747.33	-39.93	2514.22	-28.78
Aug-22	2631.38	2656.78	-35.32	2114.70	-7.71	2616.97	-32.51	2209.17	-12.52
Sep-22	3211.89	2652.75	-34.32	2445.62	-23.84	2682.59	-23.59	2322.33	-17.59
Oct-22	3644.46	2750.74	-26.73	2726.73	-25.62	2848.14	-17.72	2417.23	-11.36
Nov-22	2983.54	2869.74	-18.62	2948.93	-21.89	2822.89	-32.48	2468.84	-2.05
Dec-22	2247.1	2620.74	-22.99	2910.99	-36.61	2256.83	-2.50	2157.22	-1.24

Table 13. The error measures of prediction

Error Measures	ARIMA		SARIMA		ARIMA+ANN		SARIMA+ANN	
	Training	Validity	Training	Validity	Training	Validity	Training	Validity
Tomato								
RMSE	543	899.10	391.74	763.78	325.02	629.95	252.25	520.84
MAPE	0.17	0.30	0.13	0.25	0.11	0.17	0.10	0.15
MAE	366.54	806.08	269.74	676.51	251.05	532.26	201.02	450.68
Onion								
RMSE	490.40	568.43	451.18	506.14	398.23	445.83	295.96	284.27
MAPE	0.13	0.25	0.12	0.22	0.10	0.18	0.09	0.11
MAE	304.31	501.31	298.86	476.90	254.21	371.27	201.05	241.57

REFERENCES

Adebiyi, A.A., Adewumi, A.O. and Ayo, C.K. 2014. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics,* 7.

Ayub, S. and Jafri, Y.Z. 2020. Comparative Study of an ANN-ARIMA hybrid model for predicting Karachi stock price. *AmericanJournalofMathematicsandStatistics*, **10**, 1-9. Doi: 10.5923/j.ajms.20201001.01.

Box, G.E.P. and Jenkins, G.M. 1970. Time series analysis: forecasting and control. *San Francisco: Holden-Day.*

Brock, W. A., Scheinkman. J. A., Dechert, W. D. and LeBaron, B. 1996. A Test for Independence Based on the Correlation Dimension. Econometric Reviews, **15**, 197-235.

Chandran, K.P and Pandey, N.K. 2007. Potato price forecasting using seasonal ARIMA approach. *Potato Journal*, **34**(1), 1-2.

Chen, K.Y. and Wang, C.H. 2007. A Hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, **32**, 254-264.

Choudhary, K., Jha, G.K., Das, P. and Chaturvedi, K.K. 2019. Forecasting potato price using ensemble artificial neural networks. *Indian Journal of Extension Education,* **55**, 73-77.

Friedman, M. 1937. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association,* **32**(200), 675-701.

Jha, Girish. K. and Sinha, K. 2012. Time delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Computing and Application,* Springer.

Kambo, R., Pillai, J. and Soni, S. 2018. Agriculture commodity price forecasting using soft computing techniques: a survey. *International Journal of Advanced in Management, Technology and Engineering Sciences,* **8**, 1580-1583.

Khashei, M. and Hajirahimi, Z. 2018. A Comparative Study of Series ARIMA/MLP Hybrid Models for Stock Price Forecasting. *Communications in Statistics - Simulation and Computation,* pp 1-16. https://doi.org/10.1080/03610918.2018.1458138.

Kruskal, W. H. and W. A. Wallis 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association,* **47**(260), 583–621.

Liang Y. H. 2009. Combining seasonal time series ARIMA method and neural networks with genetic algorithms for predicting the production value of the mechanical industry in Taiwan. *Neural Computational and Application,* **18**, 833-841.

Rahman, N.M.F. 2010. Forecasting of boro rice production in Bangladesh: an ARIMA Approach. *Journal of Bangladesh Agricultural University*, **8**(1), 103–112.

Sivamani, S., Venkatesan, S.K., Lee, M., Park, J., Cho, Y. and Shin, C. 2019. A study of seasonal ARIMA model-based forecasting method for intelligent food control in a livestock environment. *International Journal of Engineering & Technology*, **8**, 555-564.

Temizel, Tugba T., and Casey, M. C. 2005. A comparative study of autoregressive neural network hybrids. *Neural Networks,* **18**, 781-789.

Tseng, F., Yu, H.C. and Tzeng, G.H. 2002. Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting & Social Change,* **66**, 71-87.

Zhang, G. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* **50**, 159-175.

APPENDIX

```
Library(tseries)
Library(forecast)
Library(astsa)
Library(sarima)
Library(seastests)
Library (metrics)
Tomato<-read.csv(file.choose())
train_set<-tomato[1:144]
validity_set<-tomato[145:156]
ARIMA<-arima (train_set, c(p,d,q))
Coeftest(ARIMA)
Checkresiduals(ARIMA)
Forecast_ARIMA<-forecast (ARIMA, h=12)
### training set ###
RMSE_ARIMA<-rmse (train_set, Forecast_ARIMA$fitted)
MAPE_ARIMA<-mape (train_set, Forecast_ARIMA$fitted)
MAE_ARIMA<- mae (train_set, Forecast_ARIMA$fitted)
###validityset###
RMSE_ARIMA<-rmse (validity_set, Forecast_ARIMA$mean)
MAPE_ARIMA<-mape (validity_set, Forecast_ARIMA$mean)
MAE_ARIMA<- mae (validity_set, Forecast_ARIMA$mean)
## SARIMA model
kw(tomato)
fried(tomato)
SARIMA<-arima (train_set, order = c(p,d,q), seasonal = list(order = c(P,D,Q), period = 12))
Coeftest(SARIMA)
Checkresiduals(SARIMA)
Forecast_SARIMA<-forecast (SARIMA, h=12)
### training set ###
```

```
RMSE_SARIMA <- rmse (train_set, Forecast_SARIMA$mean)
MAPE_SARIMA <- mape (train_set, Forecast_SARIMA$mean)
MAE_ARIMA <- mae (train_set, Forecast_SARIMA$mean)
###validityset ###
RMSE_SARIMA<-rmse (validity_setForecast_SARIMA$mean)
MAPE_SARIMA<-mape(validity_set, Forecast_SARIMA$mean)
MAE_ARIMA<- mae (validity_set, Forecast_SARIMA$mean)
### Hybrid model ###
Res_ARIMA<-residuals (ARIMA)
Bds.test (Res_ARIMA)
x1 <- Lag(Res_ARIMA, k = 1);class(x1)
x2 = Lag(Res_ARIMA, k = 2)
x3 = Lag (Res_ARIMA, k = 3)
x4 = Lag (Res_ARIMA, k = 4)
x5 = Lag (Res_ARIMA, k = 5)
x6 = Lag (Res_ARIMA, k = 6)
x7 = Lag (Res_ARIMA, k = 7)
x8 = Lag (Res_ARIMA, k = 8)
x9 = Lag (Res_ARIMA, k = 9)
x10 = Lag (Res_ARIMA, k = 10)
x11 = Lag (Res_ARIMA, k = 11)
x12 = Lag (Res_ARIMA, k = 12)
y <- cbind( x1 , x2 , x3, x4 ,x5, x6, x7, x8, x9, x10, x11, x12, Res_ARIMA)
range_data<- function ( y ) {( y - min( y ) ) /( max( y )
                                        -min( y ) ) }
data.matrix<- data.matrix (y )
min_data<-min( y )
max_data<-max( y )
b <- range_data( y )
### training, testing and validity sets ###
trainset<-data.frame(b[1:132,])
testset<-data.frame(b[133:144,])
validityset<-data.frame(b[145:156,])
###train the model###
Output<-trainset[,13]
input<-trainset[,-13]
d<-cbind(output,input)
set.seed(100)
fit=neuralnet(output~., data=d, hidden =6, act.fct = “tanh”, linear.output = FALSE, lifesign = ‘full’, rep = 5, algorithm = “rprop+”, err.fct = “sse”,stepmax = 10000 )
plot(fit,rep=2)
predict_p=compute(fit,trainset, rep=2)
predict=compute(fit,testset,rep=2)
###testing set ###
predict_p=compute(fit,trainset, rep=2)
predict=compute(fit,testset,rep=2)
# resulting output
results_p<- data.frame(actual = trainset$z, prediction = predict_p$net.result)
results <- data.frame(actual = testset$z, prediction = predict$net.result)
### training set ###
predicted_p=results_p$prediction * abs(diff(range(a))) + min(a)
actual_p=results_p$actual * abs(diff(range(a))) + min(a)
RMSE = (sum((actual_p - predicted_p)^2) / 132) ^ 0.5
MAPE = (sum(abs((actual_p - predicted_p)/actual_p)))*(100/132)
### testing set###
predicted=results$prediction * abs(diff(range(a))) + min(a)
actual=results$actual * abs(diff(range(a))) + min(a)
RMSE = (sum((actual - predicted)^2) / 12) ^ 0.5
MAPE =(sum(abs((actual - predicted)/actual)))*(100/12)
###validityset###
```

predict_v=compute(fit,validityset, rep=2)

results_v<- data.frame(actual = validityset$z, prediction = predict_v$net.result)

predicted_v=results_v$prediction * abs(diff(range(a))) + min(a)

actual_v=results_v$actual * abs(diff(range(a))) + min(a)

RMSE = (sum((actual_v - predicted_v)^2) / 12) ^ 0.5

MAPE =(sum(abs((actual_v - predicted_v)/ actual_V)))*(100/12)