Journal of Agricultural Development and Policy, 2022
Volume 32, No. 2, 159-164

Effect of Rainfall in Predicting Tomato Prices in India: An Application of SARIMAX and NARX Model

Sunny Kumar*, Akhilesh Kumar Gupta, Kashish Arora* and Kamal Vatta***

*Department of Economics and Sociology, Punjab Agricultural University, Ludhiana
**Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar, Odisha

Abstract

Although weather variable is irrepressible source for predicting prices, but unfortunately it directly influences the prices in market especially perishable commodities as it has created the demand-supply gap. The present study is an attempt to predicting tomato prices by using seasonal autoregressive moving average with exogenous variable (SARIMAX) and non linear autoregressive exogenous (NARX) model. These models are able to take advantage not only of historical data of tomato prices, but also of the impact of rainfall. It is observed from the results that the NARX model outperformed the SARIMAX model. The forecasting performance has been compared with respect to root mean square error (RMSE) and mean absolute percentage error (MAPE). The study is an effort to predict the tomato prices by taking into account the important weather variable i.e., rainfall so that the stakeholders may make production, marketing and policy decisions well in advance.

Key words: Tomato, forecasting, SARIMAX, NARX

JEL: C53, Q02, Q11, Q12

Introduction

The price forecasting is important for farmers as this process built the base for production and marketing decisions. In agriculture the production and marketing is done by the farmers to fulfill their economic needs but the instability in prices is the greatest source of risk next only to weather in agricultural production system. Another challenge emerged with perishable nature of the commodity that are often random as they are largely influenced by eventualities which are highly unpredictable in case of natural calamities like drought, floods and attacks by pests and diseases. This leads to a considerable risk and uncertainty in the process of price forecasting. In addition to this, the prices of the commodities decide the consumers' access to food as they directly influence finance liabilities. So, the policy planners require accurate predicted prices to manage food security as these estimates are such an essential process to undertake the problem of food insecurity. By the taking care of these issues in mind, there is need to develop statistically sound objective forecasts of prices based on weather variable. For the purpose of price forecasting, the seasonal autorgressive integrated moving average (SARIMA) model has been widely used in past. But this, model cannot incorporate exogenous variable. Hence, seasonal autoregressive integrated moving average with exogenous variables (SARIMAX) model is referred over ARIMA in order to forecast the prices more accurately. Paul and Sinha, 2016 used ARIMAX model with most important weather variable for predicting wheat yield of the Kanpur district of Uttar Pradesh. While the SARIMAX model has not been widely used in terms of forecasting prices in the economic field, it proves good to predict prices that are affected by weather variables. Apart from this, the use of neural network models in forecasting agriculture phenomenon is getting more attention in recent times, as the linear model like SARIMA ignore the nearness of high volatile nature and complex nonlinear structure of the series. In this study, we have used SARIMAX for inclusion of exogenous variable i.e., rainfall. For better comparison of results of SARIMAX, the study used non-linear autoregressive network with exogenous inputs (NARX). It is a very general and powerful black-box model due to its capability of capturing a wide variety of non-linear dynamic behavior. In the present investigation an attempt has been made to apply both SARIMAX model and NARX model for forecasting of prices of tomato in Mulakalacheruvu market of Andhra Pradesh by including important weather variable i.e., rainfall.

Corresponding author email: sunnykumar@pau.edu

Data Sources and Methodology

For the present investigation, we have taken monthly model wholesale price data of tomato from Mulakalacheruvu market of Andhra Pradesh. The state Andhra Pradesh was selected on the basis of the highest production in India. Further the market was chosen based on the highest percentage share of total tomato market arrival. The data for market was gathered from the AGMARKNET portal. The price series spanned a total of 120 months, from January 2013 to December 2021, 96 months utilized as a training set and last 12 data points are used for validation purpose for measuring the performance of SARIMAX and NARX model. For data analyses purpose R-statistical package was used. In addition to this, the historical monthly data of rainfall of the chittor district of Andhra Pradesh has been taken from NASA-Power Data Viewer.

SARIMAX

SARIMA model is an extension of traditional ARIMA introduced by Box and Jenkins (1976), to handle seasonal aspects of data. It is denoted as ARIMA(p,d,q)(P,D,Q)$_{[s]}$, where the small letter parentheses part (p,d,q) indicates the non-seasonal part of model while the capital letter part (P,D,Q)$_{[s]}$ indicates the seasonal part of model, s being the number of periods per season (Barathi 2011; Gupta et al.2019). The general seasonal autoregressive integrated moving average (SARIMA) model written as follows:

$$\Phi_P(B^s)\phi_p(B)\,\nabla_s^D\nabla^d\,Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

where,

$\Phi_P(B^s) = (1-\Phi_1B^s - \ldots - \Phi_PB^{sP})$ is the seasonal AR operator of order P ;

$\phi_p(B) = (1-\phi_1B - \ldots - \phi_pB^p)$ is the regular AR operator of order p;

$\nabla_s^D = (1-\mathrm{B})^d$ represents the seasonal differences and $\nabla_s^D = (1-B^s)^D$ the regular differences;

$\Theta_Q(B^s) = (1-\Theta_1B^s - \ldots - \Theta_QB^{sQ})$ the seasonal moving average operator of order Q;

$\theta_q(B) = (1-\theta_1B - \ldots - \theta_qB^q)$ is the regular moving average operator of order q;

ε_t is a white noise process(i.e., zero mean and iid)

SARIMAX model demonstrates the use of exogenous variables by using the concept of "regression with SARIMA errors". The model can be written as

$$y_t = \beta x_t + z_t$$

This equation is just a linear regression to depict the linear effect of exogenous variables on y_t. The error term z_t follows the SARIMA process and can be described by usual SARIMA equation as

$$\Phi_P(B^s)\phi_p(B)\,\nabla_s^D\nabla^d\,z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

where all the notations and operators have same meaning as above.

NARX model

Neural network is popularly known to be used in forecasting time series with non-linear behavior. In this present study, the data showed the nonlinearity pattern in the monthly wholesale price series for Mulakalacheruvu market of Andhra Pradesh which further recommended neural network model. One of the most promising recurrent neural network for time series analysis is nonlinear autoregressive exogenous (NARX) model. Nonlinear autoregressive with external input is a modified nonlinear autoregressive model by including another relevant time series as extra input to the forecasting model. As far as the usage of the NARX model is concerned, Paul and Sinha, 2016 conducted a study to predict wheat yield. The model can be written as:

$$y_{t+1} = f(x_t, x_{t-1}, x_{t-2} \ldots\ldots\ldots, x_{t-d+1}, y_t\,, y_{t-1}, \ldots\ldots\ldots, y_{t-d+1}$$

which may be written in vector form

$$y_{t+1} = f(x_t, y_t)$$

where,is the external input to the forecasting model with the same number of time delays as . Here, rainfall and price series of tomato crop are used as independent inputs to the hidden layer according to the same number of delay.

The NARX neural network can be expressed as

$$y_t = \sum_i c_i\,\Psi\left(\sum_{j=1}^{d}(a_{ji}x_{t-j} + b_{ji}y_{t-j})\right)$$

where, Ψ is the activation function in the hidden layer ; and a_{ji} and b_{ji} are the input to hidden layer weights at the hidden neuron j. c_i is the hidden to output layer weight, d is number of input nodes.

Criteria of model selection

The various tentative SARIMAX and NARX models were fitted based on the identified p,q, P and Q values. The final model was chosen based on RMSE and MAPE criteria. The model with lowest of these values is considered best among other fitted models and selected for forecasting.

Diebold–Mariano (DM) test

In order to assess whether the observed differences in forecasting power across models are actually significant, the Diebold–Mariano (DM) test for predictive accuracy was performed among the models which present best forecasting power inside each class (Diebold and Mariano 1995). The DM test approach aims to test the null hypothesis of equality of accuracy against the alternative that NARX model is more accurate. The best model was selected on the basis of significance of DM test.

Table 1. HEGY test for stationarity

Statistic	Values	p value
t_1	-3.42	0.0231*
t_2	-2.60	0.0436*
F_3:4	3.69	0.1546
F_5:6	9.28	0.0023**
F_7:8	7.10	0.0123*
F_9:10	9.09	0.0026**
F_11:12	6.91	0.0143*
F_2:12	17.92	0.0000***
F_1:12	17.06	0.000***

*Note: *** Significant at 0.01, **Significant at 0.05, *Significant at 0.10*

Results and Discussion

Testing Seasonal Unit root

Seasonality usually causes the series to be nonstationary and the first step of SARIMAX modelling is to check if the data is stationary. For testing the presence of seasonal unit root in time series data, the approach proposed by Hylleberg et al. (1990) and extended for monthly series by Franses (1990) and Beaulieu and Miron (1993) was used. The null hypothesis for HEGY is that there is presence of seasonal unit roots. The results of HEGY test for testing seasonal roots are presented in Table 1.

The results from Table 1 indicate that the null hypothesis for F_3:4=3.69 cannot be rejected and it suggests that there is presence of one complex unit root. A seasonal differencing is required to make the data stationary.

Performance Evaluation of the Fitted Models

The study has considered SARIMAX and NARX models for predicting prices of tomato including exogenous variable i.e., rainfall for better forecasting accuracy. As for SARIMAX, the results are illustrated in Table 2. The results revealed that the coefficient of rainfall in SARIMAX model is significant in all the SARIMAX models, which signify that rainfall disrupted the supply of the produce that influenced the tomato prices. Among all the models, the lowest value of RMSE and MAPE suggest that SARIMAX (0, 0, 1)(1, 1, 0)$_{[12]}$ is found to be best for price forecasting of tomato in Mulakalacheruvu market of Andhra Pradesh.

Before proceeding addition model for comparison, it is important to find whether the residuals of the best fitting SARIMAX model of the Mulakalacheruvu market are non-linear or not. If there is nonlinearity, then nonlinear models

Table 2. Parameters estimates and fitting performance of different models by using SARIMAX for the tomato prices of Mulakalacheruvu market of Andhra Pradesh

ARIMA	ar1	ar2	ar3	ma1	sar1	sar2	sma1	xreg	RMSE	MAPE
(3,0,0)(2,1,0)$_{[12]}$	0.657	-0.380	0.209	-	-0.968	-0.532	-	1.553*	1541.8	63.57
	(0.1119)	(0.1283)	(0.11)		(0.095)	(0.0887)		(0.864)	{453.5}	{44.0}
(2,0,1)(1,1,1)$_{[12]}$	0.364	-0.096		0.329	-0.367	-	-1.000	1.539*	1526.6	62.8
	(0.3295)	(0.2067)		(0.317)	(0.106)		(0.199)	(0.921)	{394.2}	{37.4}
(0,0,1) (1,1,1)$_{[12]}$	-	-	-	0.618	-0.339	-	-1.000	2.080***	1473.3	61.8
				(0.086)	(0.106)		(0.208)	(0.862)	{399.7}	{38.1}
(0,0,1) (2,1,0)$_{[12]}$	-	-	-	0.616	-0.958	-0.544	-	1.663**	1529.6	60.61
				(0.090)	(0.095)	(0.0883)		(0.841)	{457.7}	{44.8}
(0,0,1) (0,1,1)$_{[12]}$	-	-	-	0.588		-	-1.000	2.318***	1388.9	50.3
				(0.088)			(0.130)	(0.851)	{440.0}	{40.7}
(0,0,1) (1,1,0)$_{[12]}$	-	-	-	0.606	-0.588	-	-	2.239***	1400.6	63.51
				(0.087)	(0.084)			(0.882)	{564.0}	{53.7}

Notes: Figures in ( ) indicate the standard error; figures in { } indicate the RMSE of training set

Table 3: Brock-Dechert-Scheinkman (BDS) test for nonlinearity for residuals

Markets	Epsilon=0.5		Epsilon=1		Epsilon=1.5		Epsilon=2	
	M=2	M=3	M=2	M=3	M=2	M=3	M=2	M=3
Mulakalacheruvu	6.84 (0.00)	11.17 (0.00)	2.56 (0.01)	2.00 (0.04)	1.96 (0.04)	1.32 (0.08)	3.14 (0.00)	2.56 (0.01)

Note: Figure in parenthesis is the p-value of the respective value

Table 4: Forecasting performance of NARX model of Mulakalacheruvu market of Andhra Pradesh

Model Structure	Training		Testing	
	RMSE	MAPE	RMSE	MAPE
3:1:1	448.2	60.6	1269.3	64.7
3:2:1	334.0	43.1	725.0	57.6
3:3:1	223.9	30.7	701.5	58.7
3:4:1	180.9	21.8	1071.1	77.1
3:5:1	153.5	15.1	876.3	44.4
3:6:1	129.2	11.2	1045.6	78.8
4:1:1	422.3	54.5	1231.8	67.5
4:2:1	291.2	37.8	1001.8	76.0
4:3:1	205.0	24.5	636.4	54.8
4:4:1	145.6	14.0	404.2	31.5
4:5:1	118.6	7.8	709.3	57.8
4:6:1	122.6	6.4	797.5	56.8
5:1:1	408.5	54.0	1179.7	67.6
5:2:1	244.1	31.6	598.2	46.8
5:3:1	153.1	18.6	800.0	63.8
5:4:1	79.5	7.3	692.3	59.2
5:5:1	38.4	3.8	800.3	39.0
5:6:1	31.8	2.7	624.9	53.6

must be used to test for nonlinearity in the data. The study used BDS non-linearity model to test the residuals of the datasets.

The results of nonlinearity test presented in Table 3, reveal strong rejection of linearity in the case of residuals of the price series. In other words, the analysis has indicated the existence of some hidden structure left unaccounted in the residuals of linear model in selected tomato market. The test recommended the nonlinear autoregressive with exogenous variable (NARX) for better price forecasting of tomato. NARX model is a modified nonlinear autoregressive model by including another relevant time series as extra input to the forecasting model. Apart from it, the study go ahead one step by using NARX model, and pursued an iterative approach to select the output node, and we eventually chose one output node for better forecasting. We go through the different input nodes from 3 to 5 and the number of hidden nodes 2 to 6 (Table 4). The results are summarized in Table 4. The results revealed that, the best fit NARX model market was identified with 4 perceptron input nodes and 4 perceptron hidden nodes with one output (4:04:01) among all the models. The study also compared both SARIMAX and NARX models for better forecasting, results indicates that RMSE and MAPE are lower in NARX model as compared to SARIMAX model. The value of RMSE ranged between 404 to 1269 (Table 4). However, the value of the same accuracy parameter lies in between 1440 to 1541in SARIMAX model (Table 2). The lower values of all the models reflect the superiority of NARX model over SARIMAX model for forecasting purposes. Fig.1 depicts the structure of fitted NARX model. It indicates that there are 3 inputs nodes at lag 1, 2 and 12 along with an exogenous variable and 11 dummies for monthly variation.

Figure 2 reveals that NARX model is good fit for the considered seasonal time series data. To sum up, the results

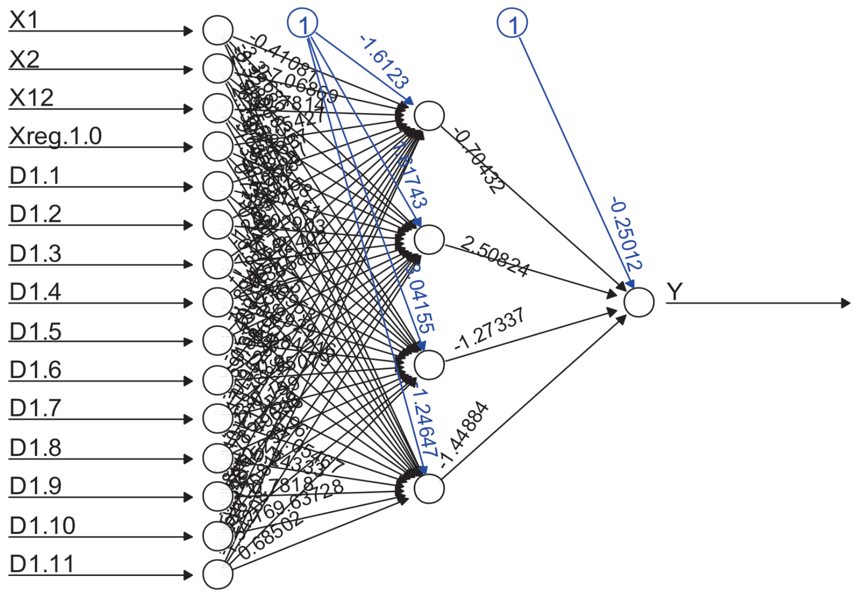

Fig. 1: Architecture of best fit NARX model for the tomoto prices of Mulakalacheruvu market

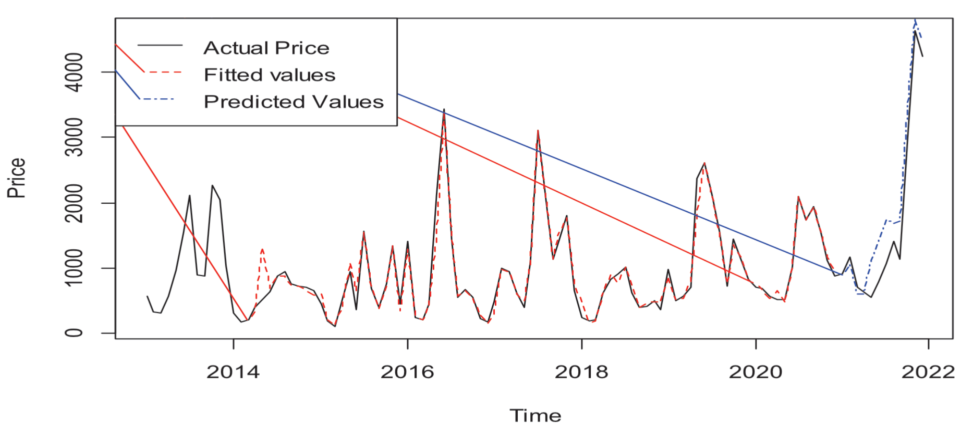

Fig 2. Actual vs. fitted price series of Mulakalacheruvu market of Andhra Pradesh by using NARX model

Table 5. Diebold–Mariano (DM) test to see the forecasting power across models

Model Pair	Training	Testing
SARIMAX(0,0,1)(0,1,1)$_{[12]}$- 4:4:1	DM=7.66 p-value=0.000	DM=1.844 P-value=0.046

indicated that the NARX model in general provided a better forecast accuracy in terms of RMSE and MAPE values as compared to SARIMAX.

To this end, Diebold–Mariano test (Diebold and Mariano, 1995) was applied for statistical comparison of forecasting performance among the SARIMAX and NARX models. It is found that the predictive accuracy with NARX are significantly better than that of SARIMAX models of the selected market (Table 5).

Conclusion and Policy Implications

The study is an effort to predict the tomato prices by taking into account the important weather variable i.e., rainfall. For evaluation in this study, two models namely SARIMAX and NARX have been applied. Empirical comparison has been carried out for the forecast performance of these two models with respect to RMSE and MAPE. From the findings, it may conclude that NARX model performs better than SARIMAX, as it has more accuracy in providing the forecasts than that of SARIMAX model. The study suggests that the rainfall as an exogenous variable in predicting prices of perishable commodities, plays an important role, and can be used for modeling and forecasting prices. It would help the policy makers for better crop planning.

References

Barathi R, Havaldar Y N, Meregi S N, Patil G M and Patil B L 2011. A study on market integration of Ramanagaram and Siddlaghatta markets and forecasting of their prices and arrivals. *Karnataka Journal of Agricultural Sciences* **24**:347-49. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.924.5902&rep=rep1&type=pdf

Beaulieu J J and Miron J A 1993. Seasonal Unit roots in aggregate U.S. data. *Journal of Econometrics* **55**: 305-28. https://doi.org/10.1016/0304-4076(93)90018-Z

Box G E P and G M Jenkins 1976. *Time series analysis: forecasting and control*, *Fifth Edition,* Holden Day, St. San Francisco, California, United States.http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015.pdf

Diebold F X and Mariano R S 1995. Comparing predictive accuracy. *Journal of Business Economics and Statistics* **13**:253–63 https://doi.org/10.1198/073500102753410444

Franses P H 1990. Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of Forecasting* **7**:199-208. https://doi.org/10.1016/0169-2070(91)90054-Y

Gupta A K, Patra C and Singh A K 2019. Forecasting market prices of soybean in Ujjain market (Madhya Pradesh). *Green Farming* **10**: 752-55. https://www.researchgate.net/publication/349288025_Forecasting_market_prices_of_soybean_in_Ujjain_market_Madhya_Pradesh

Hylleberg S, Engle R F, Granger C W J and Yoo B S 1990. Seasonal integration and cointegration. *Journal of Economics* **44**:215–38 https://doi.org/10.1016/0304-4076(90)90080-D

Paul R K and Sinha K 2016. Forecasting crop yield: a comparative assessment of ARIMAX and NARX model. *RASHI* **1**:77–85 https://www.sasaa.org/complete_journal/vol1__12.pdf

Received: January 20, 2022 Accepted: March 16, 2022