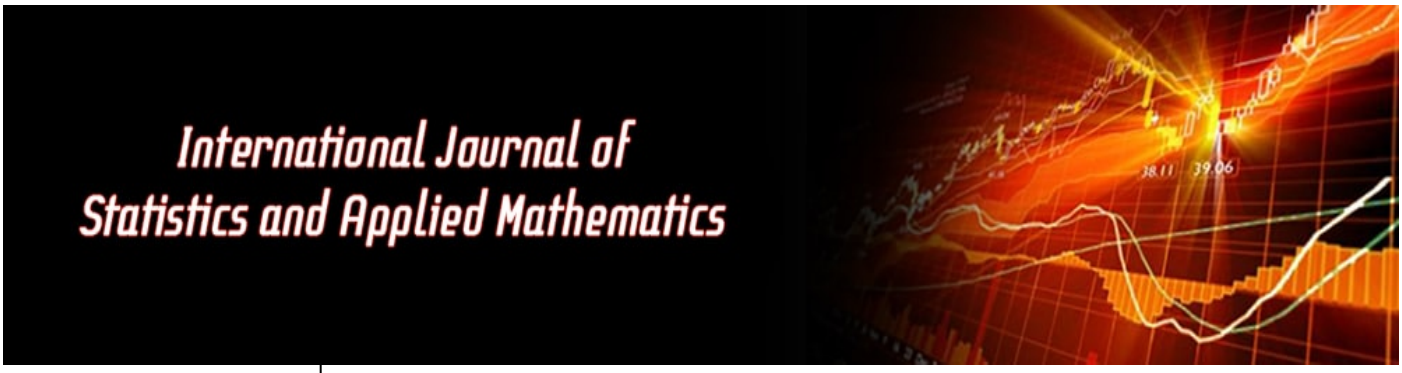
International Journal of Statistics and Applied Mathematics 2023; 8(5): 123-129



ISSN: 2456-1452
Maths 2023; 8(5): 123-129

https://www.mathsjournal.com
Received: 10-07-2023
Accepted: 15-08-2023

G Tejaswini Reddy
Department of Statistics and Computer Applications, Agricultural College, ANGRAU, Bapatla, Andhra Pradesh, India

B Ramana Murthy
Department of Statistics and Computer Applications, SMGR Agricultural College, ANGRAU, Udayagiri, Andhra Pradesh, India

Sk Nafeez Umar
Department of Statistics and Computer Applications, SV Agricultural College, ANGRAU, Tirupati, Andhra Pradesh, India

K Kiran Prakash
Department of Statistics and Computer Applications, Advanced PG Centre, ANGRAU, LAM, Guntur, Andhra Pradesh, India

Corresponding Author:
G Tejaswini Reddy
Department of Statistics and Computer Applications, Agricultural College, ANGRAU, Bapatla, Andhra Pradesh, India

Two stage modelling: Forecasting of onion arrivals and prices

G Tejaswini Reddy, B Ramana Murthy, Sk Nafeez Umar and K Kiran Prakash

Abstract
The purpose of this study was to examine the trend and forecast the future values of onion arrivals and prices in main market (C Camp) at Kurnool district of Andhra Pradesh. Data on daily arrivals and prices of onion was collected for four years seven months (January 2019 to August 2023) from the C-Camp market committee. The trend of onion was fluctuating in case of both arrivals and prices. ARIMA, ANN and Two stage models were used to forecast for next six days. The best fit models were ARIMA (4, 1, 1) model for forecasting onion arrivals whereas ARIMA-ANN (15,4) model for forecasting onion prices at C Camp market. Diebold-Mariano (dm) test was used to compare the models and to choose significant model. Onion arrivals had increasing trend over the next 6 days while the forecasted price was showing decreasing trend. This study will be helpful to farmers and to know the future trend pattern of arrivals and prices regarding onion.

Keywords: Arrivals, prices, trend, ARIMA, ANN, ARIMA-ANN

Introduction
Onion (*Allium cepa*) is one of the world's most important vegetable plant in India and most extensively produced. It is a vital vegetable crop not just for domestic consumption but also as the top earner of foreign exchange among fruits and vegetables. As a vegetable and condiment, it has an indispensable place in every cuisine. The pungency of Indian onions is well-known and they are available all year. Indian onions have two crop cycles the first begins in November and ends in January, and the second begins in January and ends in May. Soups, chutneys, and sauces are among of the culinary applications, and non-culinary applications include science education, breeding and so on. Because of its versatility onion can be boiled, grilled, baked, stuffed, fried, or eaten raw in a salad. Major onion producing states in India are Maharashtra, Madhya Pradesh, Karnataka, Andhra Pradesh, Gujarat etc. World records in area, production and productivity of onion are 5.48 lakh ha, 10.55 million tonnes and 19084.51 Kg/ha respectively. Whereas, India wide records in area, production and productivity are 1.43 lakh ha, 2.67 mt and 18645.70 Kg/ha respectively (FAOSTAT, 2022) [5]. Andhra Pradesh area, production and productivity of onion are 44.60 thousand ha, 722.90 thousand tonne, 16.21 MT/ha (www.indiastat.com, 2021-22).
Ashwini *et al.* (2016) [3] used the ARIMA approach to anticipate onion prices in the Kolhapur market of western Maharashtra for data from 2004 to 2013. The model (1, 1, 1) was best fit, predicting an increase in onion prices and demand in the next years.
Vijay *et al.* (2018) [11] examined the flexibility of ANN in time series forecasting by comparing with classical time series ARIMA model for Production of *Pearl Millet* (Bajra) Crop of Karnataka, India by collecting data from time period 1955-56 to 2014-15. The experiment shows that 2:4S:1L ANN model outperform the ARIMA (0, 1, 1) Models based on RMSE, MAPE and MSE.
Kumar *et al.* (2018) [7] studied ARIMA and ANN to predict the future prices of redgram in Kalaburagi regulated Market of Karnataka using months data for a period of 15 years (2002 to 2016).

On comparing the alternative models, it was observed that among ARIMA models AIC (4468.98) and RMSE (291.74) were the least for ARIMA (3, 1, 2) model and ANN model with minimum RMSE value 244.01 and with highest R^2 (0.98). Therefore, price forecasting using ANN model was considered the most suitable.

Amir *et al.* (2021) [1] forecasted the onion production of Pakistan using ARIMA model by collecting data from 1948-2017 and ARIMA (2, 1, 2) model is best fit showed that onion production is expected to become 1854.33 thousand tons in 2030.

Gholap *et al.* (2021) [8] computed trends in area, production, productivity, prices and arrivals of tomato in the Gultekdi market, which was based on the secondary data from 2009-10 to 2018-19 and revealed that there was declining trend in area and production of tomato.

Areef and Radha (2020) [2] forecasted the prices of potato using Artificial Neural Networks at Bangalore market in Karnataka state using monthly model prices from January 2005 to December 2019 for training and data from January 2020 to August 2020 for finalize the model for forecasting. The study showed that ANN (7-33-1) model outperformed a set of neural networks with least MAPE, MAE and MASE values *i.e.*, 2.0987, 19.4091 and 0.1345 orderly. According to forecasts, high future prices would be in the month of January, 2021 (Rs. 2247 per quintal) and that of lower prices in the month of September, 2020 (Rs. 1024 per quintal).

Reddy *et al.* (2021) [9] forecasted the prices of chilli in Andhra Pradesh using ARIMA and ANN using secondary data from June 2007 to July 2021. ARIMA (3, 1, 2) and (1, 1, 1) had been identified as a best fit model for normal and special varieties respectively, ANN 6-24-1 (6 input nodes, 24 hidden nodes, and 1 output) and 8-20-1 (8 input nodes, 20 hidden nodes and 1 output node) outperformed all other neural networks for normal and special varieties, respectively. The study concluded that ANN model performed well over ARIMA model due to their superior predictive ability in nonlinear and heterogonous data sets & concluded that the forecast price of normal varieties in the months of November 2021 and January 2022 may be Rs.6504 and Rs. 7177, respectively and the forecast price of special varieties in the months of November 2021 and January 2022 might be Rs.13241 and Rs.11861, respectively.

Singh (2007) [10] developed seasonal ARIMA and ANN model of onion arrivals in major Indian markets. Secondary data was collected from NHB, Gurgaon for the period of 1980 to 2005 and found that seasonal ARIMA (1, 1, 1) $(1, 1, 1)_{12}$ model out performed ANN (3:4:1) model in 4 out of the 5 major markets except in Bangalore.

Lack of information on possible markets, as well as onion arrival and price behaviour, exacerbates the situation for vegetable growers. As a result, information on market arrival and price behaviour is critical for producers, as it aids in determining the ideal time for 3 marketing to maximise profit. In light of this, the current study tried to assess onion market arrival/and pricing behaviour in two key market places.

Materials and Methods

Data Collection: The data of study for a period of four years 7 months (January 2019 to August 2023) in Kurnool district of Andhra Pradesh pertaining to daily arrivals and prices of onion crop were collected from market committees, with a view to examine the nature of change and degree of relationship between Arrivals and Prices of Onion crop in Andhra Pradesh by using ARIMA and ANN model in R programming.

Auto Regressive Integrated Moving Average (ARIMA) Most generally used technique for forecasting is the ARIMA model, which has been found to be more versatile in processing various patterns of time series data.

The ARIMA (p, d, q) process is given by

$$y_t=\theta_0+\emptyset_1 y_{t-1}+\emptyset_2 y_{t-2}+\ldots\ldots+\emptyset_p\, y_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots\ldots\ldots\theta_q\varepsilon_{t-q}$$

Where, y_t and ε_t are the actual value and random error at time period t, respectively.

$\emptyset_p$ (i=1, 2,….,p) and θ_j (j=1, 2,….,q) are model parameters. p and q are integers and often referred to as orders of the model.

Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2. The ARIMA model incorporates non-seasonal factors in a multiplicative model and is denoted as: ARIMA (p, d, q)

Where,

p = order of non- seasonal Auto Regressive (AR)

d = order of non- seasonal difference

q = order of non- seasonal Moving Average (MA)

Artificial Neural Networks (ANN): Techniques have been the focus of forecasting, owing to their vast range of applicability and ease of usage in solving complex issues. The nnetar () function fits an NNAR (p, k) model. If the values of p are not specified, it is selected automatically. For non-seasonal time series, the default is the optimal number of lags for a linear AR (p) model. To determine the appropriate configuration of the feed-forward network, several parameters have been varied. The number of neurons in the hidden layer was determined automatically by adopting network complexity.

Two stage Methodology: The two-stage method considers the time series Y_t as a combination of original time series and significant residual components. This approach follows the Zhang's hybrid approach, accordingly the relationship between original time series and significant residual components can be expressed as follows: $Y_t = S_1 + S_2$

Where S_1 and S_2 represents the linear and nonlinear component respectively

Comparison criteria

Diebold Mariano test

The Diebold-Mariano (DM) test is employed to determine the statistical significance difference among the models used, based on the residuals of the models (Diebold and Mariano, 1995). Consider the residuals of two models as r_1 and r_2, d_i is the absolute difference between residuals $d_i=|r_1|-|r_2|$ and the auto covariance function γ_k is expressed as,

$$\gamma_k=\frac{1}{n}\sum_{i=k+1}^{n}(d_i-\bar{d})(d_{i-k}-\bar{d})$$

The Diebold-Mariano test statistic is expressed as;

$$DM=\frac{\bar{d}}{\sqrt{\left[\gamma_0+2\sum_{k=1}^{h-1}\gamma_k\right]/n}}$$

Where, $h=n^{1/3}+1$. For testing of hypothesis, the null hypothesis (H_0) and alternative hypothesis (H_1) are defined as; $H_0=E(d)=0$ or the forecast accuracy is similar for two models and $H_1=E(d)\neq 0$, or the forecast accuracy is different for two models.

Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE) will be used as comparison criteria for the model performance.

Results and Discussion

Examination of Trend

C Camp market–onion arrivals: The fourth-degree polynomial was used to examine the trend in onion arrivals on the basis of R^2 value 0.2152. The graph showed gradually increasing trend from January 2019 to August 2023 as well as slight increasing trend in arrivals for the future time *i.e.*, the next six days.

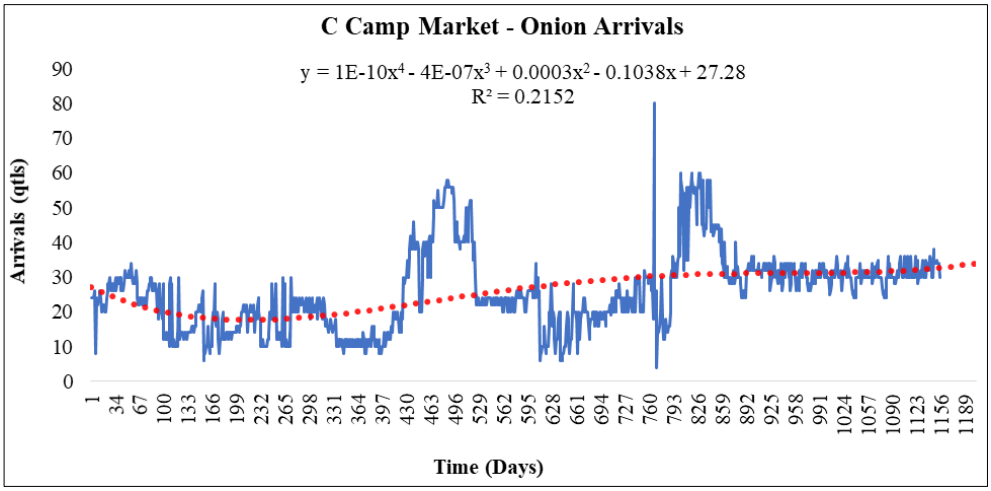


Fig 1: Trend of onion arrivals in C Camp market

3.1.2. C Camp market – onion prices

Among all trend lines fitted fourth-degree polynomial was best fit for analysing trend in onion prices. For the given polynomial, the R^2 value was found to be 0.1973. The graph showed fluctuations in trend for the data period (January 2019 to August 2023) and a decreasing trend in prices for the predicted time of next six days.

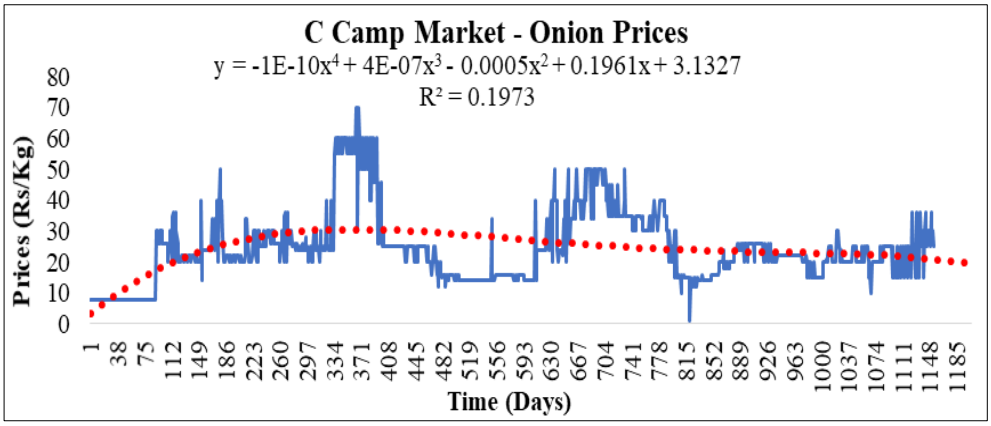


Fig 2: Trend of onion prices in C Camp market

Forecasting using ARIMA Model

Model identification: The possible models are initially found using R programme auto. arima () function, which employs a Hyndman-Khandakar method version. Table 1. Shows the values for all of the models that were tentatively recognised. ARIMA (4, 1, 1) and (1, 1, 2) were the best models for forecasting onion arrivals and prices in C Camp market based on criteria such as highest R^2 value and lowest RMSE & MAPE values.

Table 1: The tentative models for onion on arrivals and prices in C Camp market

Crop	Type	Model	R^2 Value	RMSE	MAPE
Onion	Arrivals	(4, 1, 1)	0.897	4.316	13.096
		(0, 1, 4)	0.886	4.328	13.126
		(3, 1, 5)	0.885	4.384	13.116
		(1, 1, 4)	0.884	4.385	13.106
		(2, 1, 3)	0.881	4.328	13.117
	Price	(1, 1, 2)	0.854	4.462	10.530
		(1, 1, 3)	0.854	4.474	10.637
		(3, 1, 4)	0.853	4.482	10.621
		(2, 1, 3)	0.852	4.475	10.625
		(2, 1, 2)	0.851	4.485	10.751

Estimation of parameters

Table 2. shows the Least Squares Estimation for onion arrivals and price. The parameters and were calculated using the ordinary least square algorithm with the goal of minimising the error sum of squares. The residual for each model is obtained by the iterative estimating procedure, which involves a lot of calculations.

Table 2: Conditional least square estimates of onion on arrivals and prices in C Camp market

Crop	Type	ARIMA (p, d, q)	Components	Estimate
Onion	Arrivals	ARIMA (4, 1, 1)	Constant	28.021
			AR1	0.150
			AR2	0.129
			AR3	0.093
			AR4	0.080
			MA1	0.975
	Prices	ARIMA (1, 1, 2)	Constant	21.016
			AR1	-0.863
			MA1	-0.179
			MA2	0.464

Diagnostic checking of parameters

Residual analysis can be used to determine the models' suitability. The residuals should be independent according to one of ARIMA's basic assumptions and represented in Figure 3. The optimal model selection criteria were the highest R^2 value and the lowest RMSE, MAPE, and AIC.After fitting of the model, diagnostic checking of residuals by the Box-Pierce non-correlation test revealed that the residuals were auto correlated or non-random (p<0.0001) in nature for onion prices, whereas the residuals 0.34 > 0.05 are random for onion arrivals.

Forecasting

Tables 4. and Figures 3 and 4, respectively, show the anticipated figures for onion arrivals and prices. The arrivals and prices of onion were anticipated for the next six days. Because of the high swings in arrivals, the actual and anticipated values were frequently similar. The real and forecasted price values were nearly identical.

Table 3: BDS test for residuals of ARIMA model for C-Camp Market onion prices

eps(1)		eps(2)		eps(3)		eps(4)	
M=2	M=3	M=2	M=3	M=2	M=3	M=2	M=3
21.6	22.4	27.5	28.6	24.6	25.6	20.7	21.1
(p<0.001)	(p<0.001)	(p<0.01)	(p<0.001)	(p<0.001)	(p<0.01)	(p<0.001)	(p<0.01)

Forecasting through ANN model

For the development and application of artificial neural network (ANN) models, secondary data on daily arrivals & price of onion for four years seven months (Jan 2019 – Aug 2023) in selected markets were used. The network was trained using the feed forward back propagation algorithm. Single hidden layer network was trained for a total of 100 epochs. The optimal model was chosen based on the lowest Root Mean Squared Error (RMSE) and Mean Absolute Percent Error (MAPE) values. The creation of ANN models was done using R programming.

Total thirty input combinations were considered for selecting the best onion arrivals in the C Camp market - prediction model, namely current day, previous one day, two days, and so on up to thirty days, which are represented as Y_{t-1}, Y_{t-2},..... Y_{t-30}, respectively; and 100 combinations were explored, and for onion prices in the C Camp market - prediction model, total fifteen input combinations were considered, namely current day, previous one day, two days, and so on up to fifteen days, which are represented as Y_{t-1}, Y_{t-2},..... Y_{t-15}, Finally, Table 3. shows the optimal lagged input variables and hidden nodes in hidden layer for the C Camp market.

Table 4: ANN Models for onion on arrivals and prices in C Camp market

Market	Crop	Type	Model	R^2 Value	RMSE	MAPE
	Onion	Arrivals	NNAR (30,18)	0.831	4.466	14.096
		Price	NNAR (15,4)	0.889	3.662	10.437

Table 5: Forecasted values of onion arrivals and prices in C Camp market of using ARIMA & ARIMA-ANN technique

	Days	Forecasted values using ARIMA model	Forecasted values using ARIMA-ANN model
Arrivals		Onion (qtls)	Onion(qtls)
	1-09-2023	24	28
	2-09-2023	20	26
	3-09-2023	24	30
	4-09-2023	35	25
	5-09-2023	26	28
	6-09-2023	30	30
Prices		Onion (Rs/Kg)	Onion (Rs/Kg)
	1-09-2023	30	35
	2-09-2023	25	30
	3-09-2023	25	25
	4-09-2023	25	22
	5-09-2023	25	22
	6-09-2023	20	20

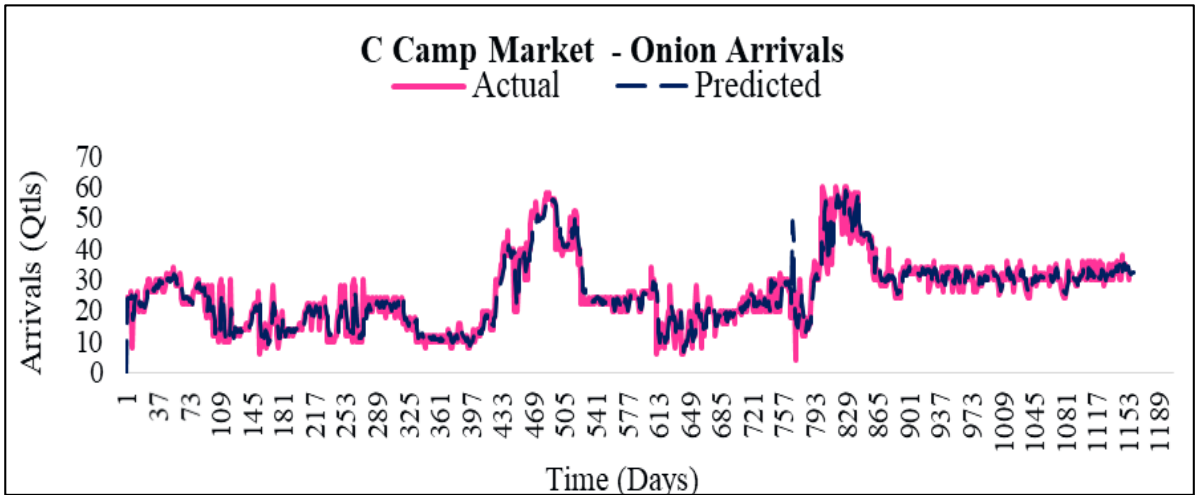


Fig 3: Ex Post Forecast of Onion Arrivals in C Camp Market

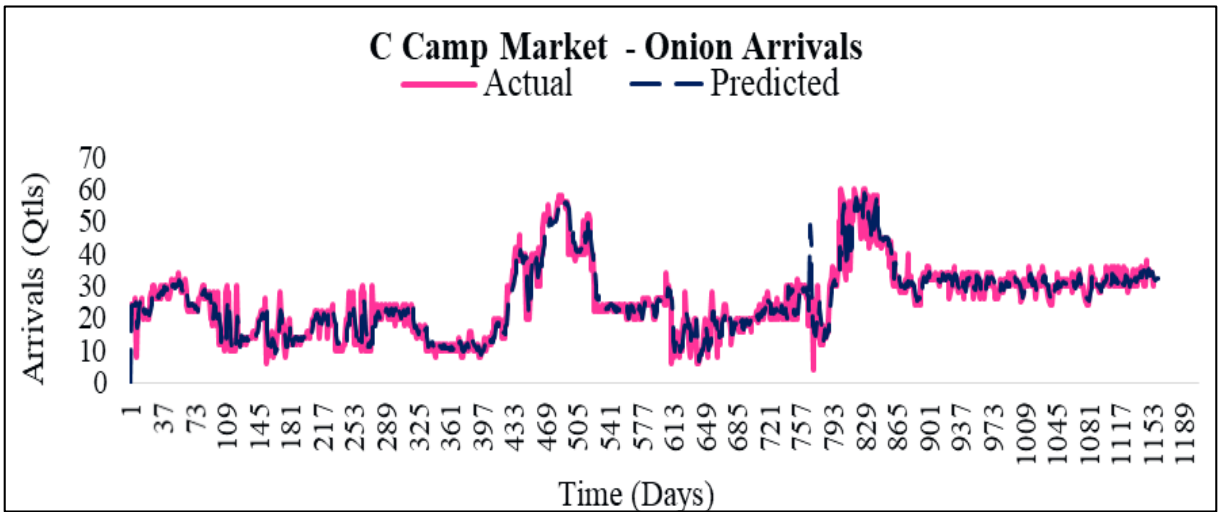
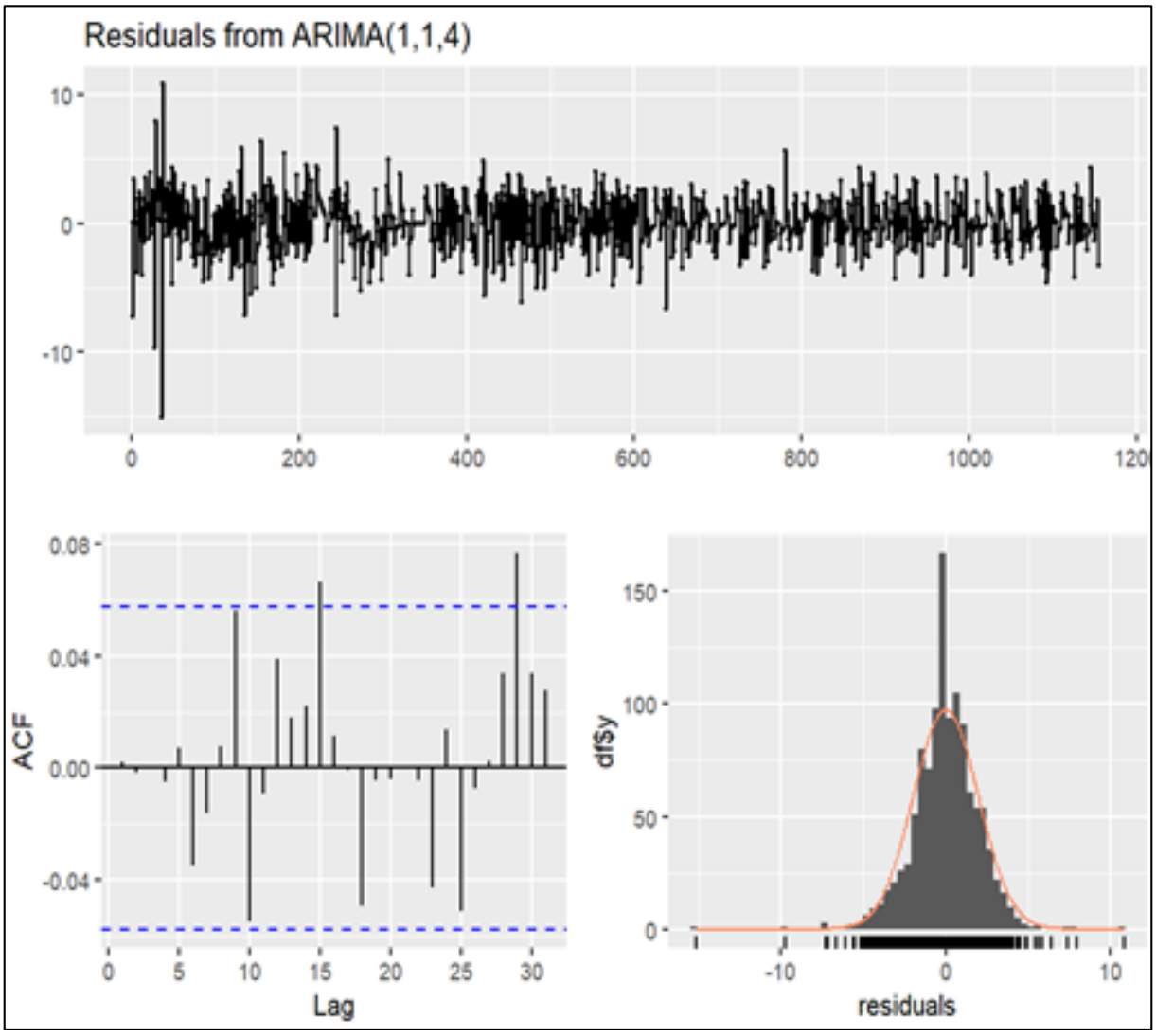


Fig 4: Ex Post Forecast of Onion prices in C Camp Market



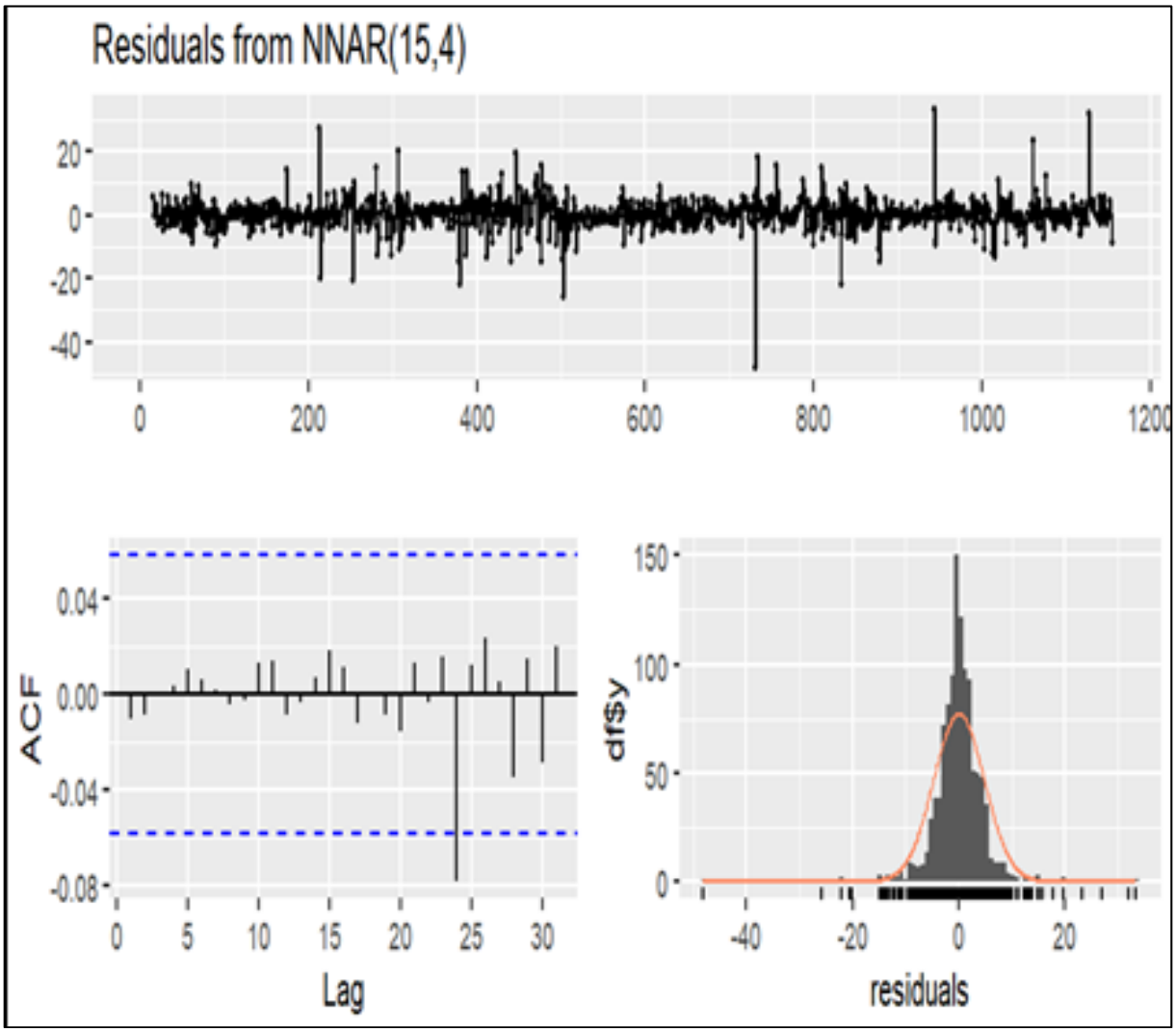


Fig 5: Residuals of Onion Arrivals & Prices in C Camp Market

Results of ARIMA-ANN model for onion prices in C CAMP market

After the confirmation of auto correlation by Box-Pierce test and nonlinearity of ARIMA residuals by BDS test (Table 3), the same residuals were modelled and forecasted using ANN model. Further the predicted residuals were combined with the forecasts obtained from original ARIMA model. This modelling procedure is called as ARIMA-ANN Two stage time series methodology. The ANN model structure (15:4S:1L) was chosen for modelling and forecasting of residual series based on lowest MSE and RMSE values. After fitting of the model, the diagnostic checking of the residuals by Box- Pierce non-correlation test showed that residuals were non-correlated in nature as probability value was 0.44.

Comparison of ARIMA Model with ANN Model in C Camp market

Comparison of onion arrivals and prices in the C Camp market, with respect to criteria of R^2, RMSE & MAPE values through ARIMA and ANN model and oresented in Table 5.

Table 6: Comparison of forecast techniques used in onion through ARIMA Model with ANN Model

Market	Crop	Type	Model	R^2 Value	RMSE	MAPE
	Onion	Arrivals	ARIMA (4, 1, 1)	0.847	4.316	13.096
			NNAR (30,18)	0.831	4.466	14.096
		Price	ARIMA (1, 1, 2)	0.854	4.462	10.530
			ARIMA-ANN (15,4)	0.889	3.662	10.437

Conclusion

After comparing onion arrivals and prices in the C Camp market, the results showed that the ARIMA model was better forecast model for arrivals whereas ANN model was better forecast model in case of onion prices with respect to maximum R^2, minimum RMSE & MAPE criteria ARIMA (4, 1, 1) model for forecasting onion arrivals whereas ARIMA-ANN (15,4) model for forecasting onion prices in the C Camp market. Diebold Mariano test was conducted in R programming and *p*-values with respect to arrivals & prices of onion ($0.012 < 0.05$) ($0.041 < 0.05$) respectively were noted and inferred that there was significance difference between ARIMA and ARIMA-ANN model for forecasting in C Camp market at 5 percent level of significance. The forecasted values of onion arrivals had increasing trend over the next 6 days while the forecasted price was showing decreasing trend. Due to perishable nature and lack of cold storage facilities farmers were selling onion with low cost.

Acknowledgement
The authors are highly thankful to the Chief Editor and anonymous referee for their valuable comments and suggestions which led to much improvement of this article.

References

1. Amir MW, Amin M, Nazir HZ. Statistical modelling and forecasting for onion production of Pakistan. Journal of Agriculture Research. 2021;59(2):171-176.
2. Areef M, Radha Y. Application of GARCH and ANN models for potato price forecasting: a case study of Bangalore market, Karnataka state. Indian Journal of Agricultural Marketing. 2020;34(3):44-52.
3. Ashwini S, Pokharkar VG, Snehal BD. Onion price forecasting in Kolhapur market of western Maharashtra using ARIMA Technique. International Journal of Information Research and Review. 2016;3(12):3364-3368.
4. Box GEP, Jenkin GM. Time series of analysis, Forecasting and Control. Sam Franscico, Holden Day, California, USA; c1976.
5. FAOSTAT. Food and Agriculture Organization of the Un ited Nations; c2022. https://www.fao.org/statistics/en/.
6. Indiastat. Socio Economic Statistical Information about India; c2021. https://www.indiastat.com/table/agriculture/selected - state - wise – area – production -productivity -o/1526854.
7. Kumar K, Bhat ARS, Pavithra NL, Megha J. Comparative performance of ANN and ARIMA models in redgram price forecasting-Kalaburagi market. Journal of Pharmacognosy and Phytochemistry. 2018;7(5):1630-1632.
8. Gholap VB, Patil SN, Benke SR. Economic analysis of arrival and price behaviour of tomato in Gultekdi market Pune. Journal of Pharmacognosy and Phytochemistry. 2021;10(2):416-419.
9. Reddy GR, Gita BM, Ketali SS, Reddy MC. Price-behaviour and forecast of chilli prices in Andhra Pradesh: a comparative analysis using ARIMA and ANN models. Angrau. 2021;9(4):95-112.
10. Singh S. Seasonal ARIMA vs. ANN model for arrival of onion in India. Ph.D. Thesis. Chaudhary Charan Singh Haryana Agricultural University, Hisar; c2007.
11. Vijay N, Mishra GC. Time series forecasting using ARIMA and ANN models for production of bajra crop in Karnataka, India. International Journal Current Microbiology Application Science. 2018;7(12):880-889.