***Original Research Article***

**A Comparative Study of VAR and ANN Models for the Forecasting of Eggplant Wholesale Prices in Lucknow**

**ABSTRACT**

Accurate forecasting of agricultural commodity prices, particularly vegetables, is essential for comprehending market dynamics and maintaining economic stability among stakeholders. Effective price forecasting fosters agricultural sustainability and economic resilience. This study aims to identify the optimal model for forecasting wholesale eggplant prices in the Lucknow market, comparing the traditional vector autoregressive (VAR) model with the Artificial Neural Network (ANN) model. The most commonly used evaluation metrics were applied to compare and assess model performance. For model building, 204 monthly observations from January 2008 to December 2024, comprising a dataset of wholesale prices and total arrival of eggplant in the Lucknow center, were utilized. In the modeling process, the entire dataset was partitioned into two subsets, training and testing data, with a split of 80:20. To check stationarity, the Augmented Dickey-Fuller test was conducted, and the lag order for the VAR model was selected based on the minimum AIC value, with [5] as the optimal lag order. Similarly, the ANN model was created using [5] lags of the variables, as independent variables. The findings indicate that the ANN [10: 3: 2: 1] model surpasses the VAR (5) model in forecasting wholesale prices within the test dataset. However, both models exhibit overfitting, likely attributable to the VAR model's inadequacy in capturing the heteroscedasticity effect and the ANN model's dependence on a limited dataset and the number of lag variables incorporated as inputs. These methods can be applied to model and forecast prices of other agricultural commodities and have potential in broader agricultural research, as their utilization in this field remains limited.

***Keywords:*** *VAR model, Artificial Neural Network, forecasting, eggplant, wholesale prices, arrival*

1. **INTRODUCTION**

India holds the position of the world's second-largest producer of fruits and vegetables, following China, as reported by the Food and Agriculture Organization (Ibef, 2022). The country's unique climatic and geographical diversity ensures continuous accessibility of fruits and vegetables throughout the year. During the fiscal year 2023, the country’s total vegetable production was estimated at nearly 212 million metric tons, encompassing key crops such as potatoes, tomatoes, onions, eggplants, and cabbage, among others. As a major global producer of cost-effective fruits and vegetables, the country holds a significant position in the international export market. (www.statista.com). With their rapid growth cycle and high productivity, vegetables serve as an essential source of income, contributing to the enhancement of livelihoods (Vanitha *et al.,* 2021). During 2020-21, from an area of 27.74 million hectares, the production of horticulture crops was 334.60 million tonnes (Horticulture Statistics at a Glance, 2021).

Eggplant (*Solanum melongena* L.), commonly referred to as brinjal or aubergine, known edible vegetable of the Solanaceae family. After potatoes, onions, and tomatoes, it is the fourth most significant vegetable cultivated in India, popular for its varied shape, size, and color of fruits (Singh et. al., 2023). Due to its high productivity, adaptability to various growing conditions, and easy availability, this crop is commonly regarded as a poor man’s crop. During the 2021-22 agricultural year, eggplant was cultivated over an area of approximately 19.63 lakh hectares worldwide, achieving a total output of 58.68 lakh tonnes and with an average productivity of 29,893 kg/ha, highlighting its significance in global vegetable production. Globally, eggplant production is dominated by China with 37.45 million tons, accounting for 63.83% of the world's total, followed by India (12.87 million tons; 21.9%). Egypt (1.2 million tons), Turkey (0.83 million tons), and Indonesia (0.67 million tons) also contribute significantly. Within Asia and the Mediterranean region, eggplant is recognized among the top five most critical vegetable crops due to its high consumption and economic value (Rajasekar *et al.,* 2024) (Horticulture Statistics at a Glance, 2021). Higher acreage was covered by the top five producing states, viz. West Bengal (1.62 lakh ha), Odisha (1.26 lakh ha), Gujarat (0.77 lakh ha), Madhya Pradesh (0.61 lakh ha), and Bihar (0.58 lakh ha) (Horticulture Statistics at a glance, 2021).

Accurate forecasting of agricultural commodity prices, especially vegetables, is crucial for understanding market dynamics and ensuring economic stability for all stakeholders, including farmers, traders, consumers, and policymakers. Since vegetable prices are significantly impacted by numerous natural calamities like droughts, floods, and pest infestations, accurate forecasts facilitate informed decision-making, mitigate risks, and guide policy development. Effective price forecasting ultimately contributed to agricultural sustainability and economic resilience (Yashavanth *et al.,* 2017). Foresight of wholesale prices strengthens farmers by improving their bargaining power and increasing market competition. By leveraging this information, farmers can compare price trends across different markets, enabling them to choose the most advantageous selling points. Additionally, this information allows them to make informed decisions regarding the timing of market entry, ultimately enhancing their profitability (Paul *et al.,* 2022).

There are a variety of methods available for forecasting economic variables. The main objective of the present paper is to compare the predictive performance of the traditional multivariate time series model, i.e., VAR (Vector Autoregressive) model, and ANN (Artificial Neural Network) for forecasting wholesale prices of eggplant in the Lucknow market. The primary objective of this study is to determine the optimal forecasting model based on prediction accuracy. Model performance has been systematically evaluated using well-established error metrics, including root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), facilitating a robust comparative analysis.

**Literature Survey:**

Forecasting vegetable prices is vital for stakeholders in the agricultural sector, enabling informed decision-making on crop selection, pricing strategies, and market interventions, ultimately enhancing economic stability and profitability. Numerous classical time series as well as machine learning models have been used by scholars and scientists to forecast agricultural commodity prices. Fouladgar *et al.,* (2013) employed VAR and ANN models for the forecasting of oil prices, findings indicated that the ANN model performed better compared to the VAR model based on the RMSE, MAE, and the coefficient of determination. Singh & Mishra, (2015) utilized the ARIMA and ANN models for the forecasting of groundnut oil prices, and findings indicated that ANN performed better than the ARIMA models. Yashavanth *et al.,*(2017) forecasted the price of coffee seeds using the VAR model and compared it with ARIMA using various accuracy measures and concluded that the VAR (2) model is suitable for the forecasting of coffee seeds prices for centers, namely, Bangalore, Chennai, and Hyderabad. Ramyar & Kianfar (2017) compared the performance of ANN and VAR models to forecast crude oil prices and concluded that a Multi-Layer Perceptron (MLP) neural network can more accurately predict crude oil prices than a VAR model. Doulah (2019) compared the forecasting performance of the VAR, ANN, and SVM models for the prediction of oil prices and concluded that the SVM model yields better results. Ravichandran & Yashavanth (2020) used the ARIMA and VARMA models for the forecasting of India’s cereal production. Rajpoot *et al.,*(2022) analyzed the impact of the lockdown implemented to curb the pandemic of COVID-19 on the prices of potato and onion crops using ARIMA, GARCH (generalized autoregressive conditional heteroscedasticity), and NNAR (Neural Network Autoregressive), Suggested the Complex ES method was more suitable compared to ARIMA and TDNN (Time Delay Neural Network) model. Goyal *et al.,*(2021) used various ARIMA models to forecast pea prices for the Varanasi market. Different ARIMA models were developed by (Kumar *et al.,*2020) for the forecasting of brinjal prices for the seven different markets of eastern Uttar Pradesh including Lucknow, and concluded that for the Lucknow market, ARIMA (1,0,1) with non-zero mean was more appropriate. Kumar *et al.,*(2021) performed vegetable price forecasting for brinjal, including potato, tomato, and onion, using ARIMA modeling. Results indicated that the ARIMA (0,1,1) (1,1,0) [52] model was developed for brinjal price forecasting for the Varanasi market, Uttar Pradesh. Numerous machine learning models, such as Generalized Neural Network (GRNN), Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting Machine (GBM) for forecasting the wholesale price of Eggplant in seventeen major markets of Odisha were explored by Paul *et al.,* (2022). They made a comparison of the predictive accuracies of different models with the ARIMA model, and it was observed that ML techniques, particularly GRNN performed better in most of the cases, particularly in the case of Brinjal. Scientist and researchers studied used multivariate time series model and compared them with univariate and machine learning models for the forecasting performances. Compilation of scientific research on the forecasting performance of multivariate time series models and VAR and machine learning models on commodity price shown in table 1.

**Table 1. Complilation of research on the forecasting performances of multivariate time series models.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author (Year)** | **Commodity** | **Methodology** | **Finding** |
| Aydin and Cavdar (2015) | US Dollar-Turkish Lira (USD/TRY), gold prices and the Borsa Istanbul (BIST) 100 index | VAR, ANN( Multilayered feed forward neural networks) | Study the relationship between US Dollar-Turkish Lira (USD/TRY), gold prices and the Borsa Istanbul (BIST) 100 index. ANN yield better results than VAR with better prediction capability. |
| Kwasi and Sharma (2015) | Groundnut prices in Bikaner district of Rajasthan | ARIMA and VAR | VAR model perform better than Simple ARIMA model and give better forecasting accuracy with multivariate time series approach |
| Pravin and Khanam (2017) | Jute prices in Bangladesh | ARIMA and VAR | ARIMA model outperformed in forecasting jute prices in Bangladesh market. |
| Ramyar and Kianfar(2017) | Crude oil prices | Compare the artificial neural network(ANN) with vector autoregressive models(VAR) | ANN model with appropriate hidden and input layer perform better. |
| Saikia and Singh(2020) | Brinjal whole sale price and arrival data in Assam | Multiplicative time series model, moving average methods, ordinary least square. | The price market of brinjal is affected by arrival in both positive and negative direction due to seasonal volatility. |
| Gopali *et al.* (2024) | Financial time series data encompassing stocks, indices, and volatile crypto currencies. | VAR, LSTM, BI-LSTM,TCN, deep learning model | Machine learning models performed superiorly than conventional VAR model. |
| Jozko and Vergos(2024) | Real estate prices | VAR and ANN | ANN out performed VAR model. This model suitable for forecasting financial market data. |

But in most of the past studies, univariate models have been applied for the forecasting of univariate price series of brinjal or any other vegetable crops, which does not fully reveal the actual behavior of price series, which may be dependent upon some other factors as independent variables apart from their own lags. This avoids the complex and hidden patterns that exist in the multivariate time series data. The main objective of the present paper is to compare the predictive accuracy of the Vector Autoregressive (VAR) model and the ANN model to forecast the wholesale prices, which is based upon the total arrival and their own lags of brinjal for the Lucknow market. Unlike other previous studies, here the wholesale price series and the arrival series of brinjal are considered simultaneously for modeling and forecasting purposes. The performance comparison of both multivariate models has been carried out using RMSE, MSE, and MAPE.

1. **METHODOLOGY**
2. **Data**

The present study was carried out using monthly data of eggplant’s wholesale prices (Rs. / Qtl.) and the arrivals (tonnes) in the Lucknow market, Uttar Pradesh, India, available at the official website of the National Horticulture Board (https://nhb.gov.in) from January 2008 to December 2024. In the procedure of time series modeling, part of the data is utilized to train the model, known as training data, and the remaining part is used for validation of the model, often known as testing data. Therefore, as an 80:20 ratio, the first 163 data points (January 2008 to June2021) were used to train the model, and the remaining 41 data points were utilized for validation purposes. The analysis part was performed using Eviews and the R programming language.

1. **VAR Approach**

A univariate autoregressive model analyzes a single variable at a time. Specifically, in an autoregressive model of order *p*, the current value of the variable is estimated using its *p* prior lagged terms. On the contrary, a multivariate auto regression such as Vector Autoregression (VAR) considers *k* number of variables. A VAR model of order *p* with *k* number of variables involves the estimation of k equations separately. In each equation, the response variable is regressed on its own *p* prior lagged terms as well as *p* lags of all the remaining variables in the system. VAR models capture the linear interdependencies among multiple time series. The VAR model provides a framework to accommodate multivariate time series, facilitating the dynamic interactions between multiple series and their lagged observations. Each variable is assumed to influence or be related not only to its own past but also to the past of all other variables in the system, which makes direct interpretation of the estimated coefficients very problematic.

Let *Yt* = (*y1t,y2t,.….., ykt*) denotes an (k$×$1) vector of k time series variables, for t = 0,1,2,3,…,n, at equally spaced time intervals. A Vector Autoregressive model with order p denoted by VAR (p) can be expressed in mathematical form as:

$ y\_{t}= β\_{0}+ β\_{1}y\_{t-1}+ β\_{2}y\_{t-2}+ β\_{3}y\_{t-3}+ ………+ β\_{p}y\_{k-p}+ ε\_{t}$… (1)

Where $y\_{t}$ denotes the m$×$1 vector of dependent variables, $β\_{0}$ is an n-dimensional vector of intercept, $β\_{i}$’s (i=1,2,3,….) are k$×k $matrices of parameters,$ y\_{t-i}$represent the ith lag of y, $ε\_{t}$is a sequence of serially uncorrelated random vectors with 0 mean (E($ε\_{t}$) =0)and constant dispersion matrix Σ.

 For a given vector time series, to choose optimal lag order (p) for VAR model implementation, there are different types of information criteria (IC) such as AIC(p), HQ(p), SC(p), and FPE(p), where AIC stands for Akaike Information Criteria, HQ Stands Hannan–Quinn information criterion, SC is Schwarz information criterion and FPE denotes Final prediction error, for such that IC(p) = min {1 ≤ i ≤ p, IC(i)}, where p is the positive integer(Hossain *et al.,* 2015). This implies that the lag order with a minimum value of all these information criteria, especially AIC, should be selected as the optimal lag order for the VAR model (Ramyar & Kianfar, 2017). Before fitting a VAR model, it is necessary to ensure that all time series must be stationary at the level. If the non-stationarity exists, perform differencingto achieve stationarity before employing the VAR framework (Alexandru, et. al., 2013). Another fundamental assumption in the application of the VAR model is the lack of cointegration between the two time series under investigation. Cointegration between the series is possible only when both series are integrated and the orderof integration is also the same. To verify the existence of cointegration, some statistical tests such as the Granger causality test, Johansen cointegration test, etc., can be employed, but only if the given time series are integrated and have identical orders of integration. To estimate the parameters of the VAR model, the Ordinary Least Square and maximum likelihood method can be used simply, since both the methods are asymptotically equivalent (Hossain *et al.,*2015).

1. **Artificial Neural Networks (ANNs)**

Artificial Neural Networks (ANNs) are advanced computational systems characterized by a network of numerous fundamental processing units that communicate by transmitting signals througha dense matrix of weighted connections to facilitate signal transmission and pattern recognition. ANNs were originally developed as computational models inspired by the structure and functionality of the biological brain. Similar to the biological brain, neural networks also consist of processing elements (artificial neurons) and connections (weights) between them. ANNs are non-linear, data-driven, self-adaptive, and nonparametric techniques (Zhang *et al.,* 1998; Khashei & Bijari, 2009) contrary to the conventional models based methods. ANN model does not require prior assumptions of the underlying data, rather it is mostly dependent on the data attributes.

A well knows ANN architecture is a Multilayer Feed Forward neural network that comprises an interconnected network of three layers, viz. input, hidden, and output layer, and a set of connected cells known as neurons, joined by acyclic links i.e., there is no cycle or loop in the network (Adhikari & Agrawal, 2013). The three-layered architecture of ANN is diagrammatically presented in Figure 1. The neuron receives impulses from either input cells or other neurons executes some type of transformation of the input and transmits the outcome to the other neuron or output cells. A neuron is a real function of the input vector (yt) (t = 1, 2,….,n). The output of the model can be mathematically expressed as:

$y\_{t}= α\_{0}+ \sum\_{j=1}^{q}α\_{j}g\left(β\_{0j}+ \sum\_{i=1}^{p}β\_{ij}y\_{t-p}\right)+ ε\_{t} , ∀ t $…. (2)

Where$ α\_{j} (j=01,2,…,q)$ and $β\_{ij}(i=0,1,2,..,p;j=01,2,…,q)$ are the model parameters, also known as connection weights, and p & q are the number of input nodes and hidden nodes, respectively, $ε\_{t}$ the disturbance term,$α\_{0}$ and $β\_{0j}$ are the bias terms, and g is the sigmoid (logistic) activation function$\left(g\left(x\right)= \frac{1}{1+ e^{-x}}\right)$. The most utilized ANNs in forecasting are Multilayer Perceptron (MLP) neural networks, which are trained under a back propagation algorithm, which utilizes data to adjust network weights and thresholds, minimizing prediction errors on the training set.



**Fig. 1. Architecture of the ANN model with a single hidden layer**

1. **Model Evaluation Criteria**

To make the comparison of the predictive performance of the particular models, their relative performance on the testing set is considered. Therefore, in this study to check the forecasting ability of different models, the most frequently used model evaluation criteria such as RMSE, MAE, and MAPE were used. These model evaluation measures with their respective statistic are presented in Table 2. In Table 2, $y\_{t }$representsthe actual value, $\hat{y}\_{t}$ denotes the predicted value, and n is the size of the dataset.

**Table 2. Model Evaluation Criteria**

|  |  |
| --- | --- |
| **Criteria** | **Statistic** |
| Root Mean Squared Error (RMSE) | $$RMSE= \sqrt{\frac{1}{n}\sum\_{t=1}^{n}\left(y\_{t}-\hat{y}\_{t}\right)^{2}}$$ |
| Mean Absolute Error (MAE) | $$MAE= \frac{1}{n}\sum\_{t=1}^{n}\left|y\_{t}-\hat{y}\_{t}\right|$$ |
| Mean Absolute Percentage Error (MAPE) | $$MAPE= \frac{1}{n}\sum\_{t=1}^{n}\left|\frac{y\_{t}-\hat{y}\_{t}}{y\_{t}}\right|×100$$ |

1. **Results and Discussion**
2. **Descriptive Analysis**

The overall summary statistics of the wholesale prices and arrival data are presented in Table 3. The maximum and minimum wholesale prices were observed in February 2024(Rs 3306/quintal) and in Dec 2009 (Rs. 303/quintal), respectively. Similarly, total arrivals of Eggplant in the Lucknow market were maximum in December 2016 (940 tonnes) and minimum in Feb 2012 (82 tonnes). The kurtosis is much less (<3) in almost both data series, leading to the platykurtic nature of the distribution. The Jacque-Bera statistic of normality indicated that both wholesale prices and arrival series are not normally distributed. To highlight the dynamics of demand and supply of eggplant in Lucknow markets in the state of Uttar Pradesh, the combined time series plot of wholesale prices and arrivals is represented in Figure 2.

**Table 3. Descriptive Statistics of Eggplant’s wholesale prices and arrival in Lucknow Market**

|  |  |  |
| --- | --- | --- |
| **Statistic** | **Wholesale Price** | **Total Arrival** |
| (in Rs. /quintal) | (in tonnes) |
| Minimum | 303 | 82 |
| Maximum | 3306 | 940 |
| Mean | 1301.21 | 393.79 |
| Standard Deviation | 576.38 | 150.87 |
| Median | 1246.5 | 367 |
| Skewness | 0.94 | 0.81 |
| Kurtosis | 1.00 | 0.48 |
| Standard Error | 40.35 | 10.56 |
| Jarque- Bera | 40.05 | 20.08 |
| Probability | < 0.0001 | < 0.0001 |

****

**Fig. 2. Time Series Visualization plot of monthly Wholesale Price and Total Arrival of eggplant in Lucknow Market, Uttar Pradesh**

 To confirm the stationarity of the series, the most widely used unit root test, namely the Augmented Dickey-Fuller (ADF) test, along with the Phillips-Perron test, were applied, and the results obtained from both of the tests were presented in the tabular form using the Table 4. The p-value obtained for both monthly time series was less than 0.05 at a 5% level of significance, so here we have sufficient evidence to reject the null hypothesis of non-stationarity, and we can conclude that both the time series are originally stationary at the level. Hence, there is no need to take the differencing of the series and we can use the original series for modeling purposes. Thus, here both the time series under investigation are stationary at level and not integrated, so it is evident that the assumption of the absence of cointegration between the series is fully satisfied for VAR modeling.

**Table 4. Results of the Unit Root Test opted for the testing of Stationarity**

|  |  |  |
| --- | --- | --- |
| **Monthly Time Series** | **Augmented Dickey-Fuller Test**  | **Phillip Peron (PP) unit root test** |
| **Test Statistic** | **p Value****(< 0.05)** | **Test Statistic** | **p Value****(<0.05)** |
| **Wholesale Price (WP)** | -5.36 | 0.01 | -116.46 | 0.01 |
| **Total Arrival (TA)** | -3.72 | 0.02 | -113.04 | 0.01 |

1. **VAR Estimation**

The Vector Autoregression (VAR) model of order up to p=5 was evaluated, and the VAR (5) model was identified to be the most appropriate as it exhibited the minimum information criteria (IC) as AIC, HQ, and FPE value, however, SCIC is minimum for lag order 1. As the majority of the ICs have a minimum value at lag order 5, so in VAR modeling, the optimal lag order is five. Table 5 represents the values of IC with their indication of selected lag order for each criterion by an asterisk “\*”.This analysis reveals that the minimum value of information criteria has been obtained at the lag length of order five than that of any other lag lengths of orders.

**Table 5. Lag order selection criteria and Selected VAR order (p)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lag** | **AIC** | **HQ** | **SC** | **FPE** |
| 1 | 21.37916 | 21.42639 | **21.49546\*** | 1.926880e+09 |
| 2 | 21.35495 | 21.43367 | 21.54878 | 1.880850e+09 |
| 3 | 21.35088 | 21.46109 | 21.62225 | 1.873358e+09 |
| 4 | 21.35208 | 21.493777 | 21.70098 | 1.875846e+09 |
| **5** | **21.34699\*** | **21.41017\*** | 21.77373 | **1.866706e+09\*** |

The ordinary Least Square Estimation method was used for the estimation of parameters. Parameter estimation results for wholesale prices based upon arrivals lags and lags of itself using the VAR (5) model are given in Table 6. Here we obtained a total of eleven parameters, including constants, from wholesale prices as an output variable and up to 5 lags of both the variables (wholesale prices and arrival) as independent variables. Out of these, six parameters were statistically significant at 90%, 95%, and 99% confidence levels. The models identified suggest that both present and future values of a series are influenced not only by its previous values but also by the past values of other series within the system.

**Table 6. Model Estimation results from the VAR (5) model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameters** | **Estimate** | **Std. Error** | **t value** | ***Pr(>|t|)*** |
| Wholesale price (-1) | 0.46526 | 0.09273 | 5.017 | 1.49e-06 \*\*\* |
| Total Arrival (-1) | -0.12104  | 0.27865  | -0.434  | .66 |
| Wholesale price (-2) | 0.19431  | 0.10726  | 1.812  | .07**.** |
| Total Arrival (-2) | 0.63234  | 0.30072  | 2.103  | .03\* |
| Wholesale price (-3) | -0.16618  | 0.10840  | -1.533  | .12 |
| Total Arrival (-3) | -0.08064  | 0.30371  | -0.266  | .79 |
| Wholesale price (-4) | 0.04284  | 0.10813  | 0.396  | .69 |
| Total Arrival (-4) | -0.57078  | 0.30288  | - 1.884  | .06**.** |
| Wholesale price (-5) | 0.09914  | 0.09567  | 1.036  | .30 |
| Total Arrival (-5) | 0.61908  | 0.27149  | 2.280  | .02\* |
| Const | 240.49010  | 126.12231 | 1.907 | .05**.** |
| Log Likelihood: -2112.797  |

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1: Significant at <1%, 1%, 5% and 10% Level of Significance level

Before diagnostic checking, the stability condition of the VAR (5) model is checked through the unit circle of the inverse root of the AR characteristics polynomial (Figure 3). It indicates clearly that no root lies outside the circle that satisfies the stability condition. Also, Table 7 represents the roots of the characteristic polynomial, in which all the roots are less than one, which implies that the estimated VAR (5) model is stable.

**Table 7. Roots of the characteristic polynomial**

|  |
| --- |
| **Roots of the characteristic polynomial** |
| 0.9389 | 0.7437 | 0.6937 | 0.6937 | 0.692 | 0.6474 | 0.6476 | 0.6008 | 0.6008 |



**Fig. 3. Inverse roots of the AR characteristics polynomial of the estimated VAR (3) model**

Model diagnostics is concerned with the checking the autocorrelation and heteroscedasticity effect, or ARCH effect. Table 8 represents how each univariate equation fits the data. In the testing of autocorrelation, the Durbin-Watson (D-W) test and the Ljung Box test were applied. The D-W test statistic value is close to two, and the p value from the Ljung Box test was more than 0.05for both of the variables, indicating the lack of autocorrelation between the residuals. In addition, in the testing of the heteroscedasticity effect, the p value is significant at a 5% level of significance, indicating the presence of the ARCH effect, which means, the VAR (5) model is seriously affected by the heteroscedasticity effect. Table 9 represents the comparison of accuracy measures for the training and testing sets. From Table 9, the RMSE, MAE, and MAPE values are higher for the testing set compared to the training set, this implies that the VAR (5) model causes overfitting. The main reason behind the overfitting of the model is the presence of heteroscedasticity.

**Table 8. Model Diagnostics test**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** |  | **Autocorrelation** | **ARCH effect** |
|  | **Durbin-Watson test** **(D - W test statistic)** | **Ljung Box test** **(*P* value)** | **Chi-squared** | **df** | ***P* value** |
| **Wholesale price (WP)** |  | 2.00 | 0.94 | 78.271 | 45 | *<0.001* |
| **Total Arrival (TA)** |  | 2.06 | 0.99 |

**Table 9. Model Evaluation measures for the VAR (5) model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **RMSE** | **MAE** | **MAPE (%)** |
| **Training Set** | 486.9259 | 377.6685 | 30.21515 |
| **Testing set** | 864.8721 | 666.6018 | 52.92745 |

* 1. **ANN Model Implementation**

This study employs a multi-layer feed forward neural network model, trained using a supervised machine learning algorithm, to iteratively adjust network parameters and improve prediction accuracy. Eggplant wholesale price prediction is facilitated through the back-propagation algorithm, integrating a sigmoid activation function in the hidden layer, and for the output layer, a linear activation function was used to enhance model efficiency. Prior to ANN model implementation, all input variables undergo the process of normalization to the [0,1] interval to achieve standardization and optimize network training stability. Throughout the modeling process, coefficients are optimized through iterative comparisons between predicted outputs and actual values, ensuring model accuracy. The weight adjustment process continues until the model achieves predefined performance metrics, ensuring stability and accuracy. From a theoretical perspective, neural networks may be constructed with varying numbers of hidden layers, but the universal approximation theorem establishes that a network with one or two hidden layers, given an adequately large number of neurons, is capable of approximating any arbitrary input-output function with high precision. Hence, the proposed neural network has been structured with two hidden layers. The optimal number of neurons in the hidden layer is identified by training the artificial neural network multiple times with varying neuron counts. The structure that achieves the lowest root mean square error (RMSE) on the testing dataset is considered the most effective structure, with a sufficient number of neurons. The number of neurons in the input layer is determined by the number of independent variables, which is ten because we considered five lags of both variables, similar to the VAR (5) model. The number of candidate models were tried with different structures and compared using various accuracy measures like RMSE, MAE, and MAPE for both training and testing datasets, and the results are presented in Table 10. The best fitted network based on the most proper performance with the test dataset contains ten neurons in the input layer, three neurons in the first hidden layer & two neurons in the second hidden layer, with one in the output [ANN [10: 3: 2: 1]]. The architecture of the ANN [10: 3: 2: 1] model is presented through Figure 4. However, the obtained best candidate ANN model is also showing overfitting. The reason behind that could be the selection of the number of lag values as independent variables and the availability of limited data to train a machine learning model.

**Table 10. Forecasting the performance of candidate ANN models**

|  |  |  |
| --- | --- | --- |
| **Candidate ANN models** | **Training Set** | **Testing Set** |
| **RMSE** | **MAE** | **MAPE** | **RMSE** | **MAE** | **MAPE** |
| 10:5:1 | 505.96 | 403.42 | 40.85 | 975.16 | 747.0761 | 36.10 |
| 10:6:1 | 575.97 | 471.86 | 48.39 | 1028.94 | 763.80 | 43.25 |
| 10:10:1 | 623.65 | 452.93 | 46.57 | 871.67 | 674.61 | 38.64 |
| 10:12:1 | 623.32 | 474.78 | 45.54 | 834.21 | 670.98 | 35.71 |
| 10:15:1 | 726.17 | 573.17 | 57.61 | 1017.63 | 831.91 | 46.98 |
| 10:2:2:1 | 397.64 | 308.20 | 31.83 | 730.23 | 542.63 | 25.58 |
| 10:2:3:1 | 402.19 | 309.21 | 31.28 | 747.21 | 547.38 | 24.98 |
| **10:3:2:1** | **394.54** | **306.15** | **30.05** | **718.93** | **532.41** | **24.63** |
| 10:3:3:1 | 646.74 | 392.23 | 38.32 | 1975.44 | 1049.49 | 50.65 |

****

**Fig. 4. Architecture of obtained best ANN [10: 3: 2: 1] model with two hidden layers**

* 1. **Comparison of the VAR (5) model and ANN [10: 3: 2: 1] for wholesale prices**

For the prediction of wholesale prices, VAR (5) and ANN [10: 3: 2: 1] models were used. To compare the predictive performance of both models, Table 11 represents, RMSE, MAE, and MAPE values obtained in the testing set for wholesale prices. The results show that for wholesale prices, the series ANN [10: 3: 2: 1] model has minimum RMSE, MAE, and MAPE compared to VAR (5) in the testing dataset, but both the models cause overfitting, which may be the reason behind this is the presence of ARCH effect (Table 8), which does not full captured by VAR and the reason for ANN model could be the selection of numbers of lag values as independent variables and the availability of limited data to train a machine learning model.

**Table 11. Model Evaluation measures for the VAR (5) model in the testing dataset**

|  |  |  |
| --- | --- | --- |
|  | **Training Dataset** | **Testing Dataset** |
|  | **RMSE** | **MAE** | **MAPE (%)** | **RMSE** | **MAE** | **MAPE (%)** |
| VAR (5) | 486.92 | 377.66 | 30.21 | 864.87 | 666.60 | 52.92 |
| **ANN [10: 3: 2: 1]** | **394.54** | **306.15** | **30.05** | **718.93** | **532.41** | **24.63** |

Hence, in future investigations, the VARMA or multivariate ARCH or GARCH model can be studied to analyze the ARCH effect present in the price series of brinjal. In the case of artificial neural networks (ANNs) or machine learning models can be improved by carefully selecting input variables and ensuring a sufficiently large number of data points. So that the performance accuracy of the developed model would improve and show better performance compared to existing results.

1. **Conclusion**

Accurate forecasting of agricultural commodity prices, especially vegetables, is crucial for understanding market dynamics and ensuring economic stability for all stakeholders, including farmers, traders, consumers, and policymakers. Hence, in this study, for the prediction of wholesale prices of eggplant, numerous VAR (p) and ANN models were developed, and to obtain the most suitable model, the results were compared concerning testing data using RMSE, MAE, and MAPE. The VAR model is suitable for stationary time series only, but because of the presence of the ARCH effect in the model, VAR (5) model causes overfitting. Similarly, ANN model can be performed better, but it also causes overfitting, and the reason could be the selection of the number of lag values as independent variables and the availability of limited data to train a machine learning model. But overall, the ANN [10: 3: 2: 1] model performed better compared to the VAR (5) model in both the training and testing datasets to forecast the wholesale price of brinjal for Lucknow market.

1. **Future Scope:**

In the future investigations, for a traditional multivariate time series model, the VMA, VARMA or multivariate ARCH or GARCH model can be studied to analyze the ARCH effect present in the price series of brinjal. In the case of artificial neural networks (ANNs) or machine learning models can be improved by carefully selecting input variables and ensuring a sufficiently large number of data points. This allows the model to be effectively trained, enhancing its performance and accuracy. So that the performance accuracy of the developed model would improve and show better performance compared to existing results. In addition to this, numerous advanced machine learning models like random forest, support vector regression etc. can be used for future perspective. These methodologies can be extended for modeling and forecasting of prices of other agricultural commodities and can also be applied in other areas of agricultural research, as these techniques have still not been much utilized in agriculture and allied disciplines, specifically for multivariate time series.

**Disclaimer (Artificial intelligence)**

Option 1:

Author(s) hereby declared that No generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**References:**

Ibef. (2022, October 28). India’s fruits and vegetables industry. *India Brand Equity Foundation*. https://www.ibef.org/blogs/india-s-fruits-and-vegetables-industry

https://www.statista.com/statistics/621210/vegetable-production-in-india/#statisticContainer

Vanitha SM, Roy S, Singh N, Singh J. (2021). Growth trend in vegetable production time series analysis. *Journal of Applied Horticulture*.;23(3):294-298.

Horticulture Statistics at a glance, (2021). Horticulture Statistics Division Department of Agriculture & Farmers Welfare Ministry of Agriculture & Farmers Welfare Government of India

Singh, P. L., Singh, P., Singh, O., & Kumar, D. (2023). A study on trend analysis of area, production, and productivity of brinjal in Uttar Pradesh. *The Pharma Innovation*, 12(7), 322–325.

Rajasekar, B., Nagaraju, K., Nirosha, K., Harshitha, Dr. D. N., Mishra, Dr. P., Laxmi, Dr. K. V., Swetha, T. N., Hudge, Dr. B. V., & Bindu, S. H. (2024). Pesticide use and residue management in brinjal crop Solanum melongena (Lin.). *International Journal of Advanced Biochemistry Research*, 8(3), 312–319. https://doi.org/10.33545/26174693.2024.v8.i3d.734

Yashavanth, B. S., Singh, K. N., Paul, A. K., & Paul, R. K. (2017). Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model. *The Indian Journal of Agricultural Sciences*, *87*(6). https://doi.org/10.56093/ijas.v87i6.70960

Paul, R. K., Yeasin, Md., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., & Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLOS ONE*, 17(7), e0270553. <https://doi.org/10.1371/journal.pone.0270553>

Fouladgar, M. M.; Yazdani, M.; Khazaee, S.; Zavadskas, E. K.; Fouladgar, V. (2013). Comparison of vector time series and ANN techniques for forecasting WTI oil price. *Economic Computation and Economic Cybernetics Studies and Research* (ECECSR) 47(4): 19–35.

Singh, A., & Mishra, G. C. (2015). Application of Box-Jenkins Method and Artificial Neural Network Procedure for Time Series Forecasting of Prices. *Statistics in Transition New Series*, 16(1), 83–96. https://doi.org/10.21307/stattrans-2015-005

Ramyar, S., & Kianfar, F. (2017). Forecasting crude oil Prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, *53*(2), 743–761. https://doi.org/10.1007/s10614-017-9764-7

Doulah, M.S.U. (2019). Time Series Forecasting: A Comparative Study of VAR ANN and SVM Models, *Journal of Statistical and Econometric Methods*, 8(3), 21-34.

Ravichandran, S., & Yashavanth, B. S. (2020). ARIMA Vs VARMA-Modelling and Forecasting of India’s Cereal Production. *Journal of the Indian Society of Agricultural Statistics,* 74(2), 121–128

Rajpoot, K., Singla, S., Singh, A., & Shekhar, S. (2022). Impact of COVID-19 lockdown on prices of potato and onion in metropolitan cities of India. *Journal of Agribusiness in Developing and Emerging Economies*, 12(3), 386–399. https://doi.org/10.1108/JADEE-04-2021-0099

Goyal, A., Badal, P.S., Kamalvanshi, V., Kumar, P. and Mondal, B. (2022). Forecasting of Pea Prices of Varanasi Market Uttar Pradesh, India through ARIMA Model. *Agro Economist - An International Journal*, 09(01): 49-54.

Kumar, P., Badal, P. S., Paul, R. K., Jha, G. K., Venkatesh, P., Kingsly, I. T., et. al., (2020). Empowering farmers through future price information: a case study of price forecasting of Brinjal in eastern Uttar Pradesh. *Indian Journal of Economics and Development*, 16(4), 479-488

Kumar, P., Badal, P. S., Jha, G. K., Paul, R. K., Venkatesh, P., Kamalvanshi, V., et. al., (2021). Enabling informed resource allocation decision by vegetable growers of Varanasi Uttar Pradesh: price forecasting using ARIMA. *Agricultural Situation in India*, 77, 16-24.

Aydin, A. D., & Cavdar, S. C. (2015). Comparison of prediction performances of artificial neural network (ANN) and vector autoregressive (VAR) models By using the macroeconomic variables of gold prices, Borsa Istanbul (BIST) 100 index and US Dollar-Turkish lira (USD/TRY) exchange rates. *Procedia Economics and Finance*, *30*, 3–14. https://doi.org/10.1016/s2212-5671(15)01249-6

Kwasi, B. R., & Sharma, A. (2015). Comparing the forecasting power of multivariate VAR and univariate Arima models: A case of groundnut prices in Bikaner District of Rajasthan. *Asian Journal of Research in Business Economics and Management*, *5*(3), 51. https://doi.org/10.5958/2249-7307.2015.00057.2

Pravin, S. and Khanam, M. (2018). Comparison between ARIMA and VAR Model Regarding the Forecasting of the Price of Jute Goods in Bangladesh. *Dhaka University Journal Science*. 66(2): 91-94.

*Saikia S. and Singh R.(2020).* BEHAVIOUR OF PRICE AND MARKET ARRIVAL OF BRINJAL IN THE WHOLESALE VEGETABLE MARKETS OF ASSAM. *International Journal of Management (IJM).4327-4334.*

DOI: <https://doi.org/10.17605/OSF.IO/YUZ4W>

Gopali, S., Siami-Namini, S., Abri, F., & Namin, A. S. (2024). A comparative multivariate analysis of VAR and Deep Learning-Based models for forecasting volatile time series data. *IEEE Access*, *12*, 155423–155436. https://doi.org/10.1109/access.2024.3447354

Jozko, M., & Vergos, K. (2024). Comparing Var and Artificial Neural Network Model in Forecasting Real Estate Prices during Good and Bad Times. *preprint.org*, *doi: 10.20944/preprints202407.2185.v1*. https://doi.org/10.20944/preprints202407.2185.v1

Hossain, N. A., Kamruzzaman, N. M., & Ali, N. M. A. (2015). Vector Autoregressive (VAR) modeling and projection of DSE. *Chinese Business Review*, *14*(6). https://doi.org/10.17265/1537-1506/2015.06.001

Alexandru, C., Caragea, N., & Dobre, A. M. (2013). Vector Autoregressive Models Using “R”. *SEA-Practical Application of Science*, *1*(1), 1, *DOAJ (DOAJ: Directory of Open Access Journals)*. https://doaj.org/article/88531fc98bd6482fb3bef2eda8806ccc

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: *International Journal of Forecasting*, *14*(1), 35–62. <https://doi.org/10.1016/s0169-2070> (97)00044-7

Khashei, M., &Bijari, M. (2009). An artificial neural network (p,d,q) model for time series forecasting. *Expert Systems with Applications*, *37*(1), 479–489. https://doi.org/10.1016/j.eswa.2009.05.044

Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613. (https://arxiv.org/pdf/1302.6613)