A COMPARATIVE STUDY OF TIME SERIES APPROACHES FOR COTTON PRICE FORECASTING IN ADONI MARKET OF ANDHRA PRADESH

S CHANDANASREE

Faculty, Dept of Seed Science &Technology,

Bharatiya Engineering Science & Technology Innovation University, Andhra Pradesh, India.

Abstract

Cotton is one of the most important fibre and cash crops in India, playing a dominant role in the both industrial and agricultural economy of the country. Present study made an attempt to forecast the cotton market prices in Adoni of Andhra Pradesh using the secondary time series data during January 2010 to April 2025. In this, a comparative analysis was made between various selected univariate times series models (GARCH, ANN, SVR and Ensemble approach) and time series models with exogenous (*x= arrivals* (variable (ARx-GARCH, ANNx and SVRx), as to identify the best fitted model, on the basis of various diagnostic criterion such as RMSE, MAPE, MAE, Box-Pierce test and Diebold-Mariano (DM) test. Finally, it was resulted that ensemble models (EMD-ANN) outperformed among the other selected models for the prices of cotton in Adoni market and for the month of October 2025, it was forecasted at Rs.7682.5 per Quintal.

Keywords: Cotton, price, model and ensemble

Introduction

Agricultural commodity market plays a crucial role in providing price signals and reducing uncertainties for producers, traders and policy makers. Cotton provides the basic raw material (fibre) to textile industry. In India, Cotton provides direct livelihood to 6 million farmers and about 40-50 million people were employed in cotton trade and its processing. India is the largest producer of cotton globally, accounting for 23% of total global cotton production. In India, there are ten major cotton growing states which are divided into three zones namely north zone, central zone and south zone. North zone consists of Punjab, Haryana, and Rajasthan. Central zone includes Madhya Pradesh, Maharashtra and Gujarat. South zone comprises Andhra Pradesh, Telangana, Karnataka and Tamil Nadu. During 2023- 24, India had produced 325 lakh bales of cotton under the cultivable area of 130.6 lakh hectares. In the same year the cultivable area and production of Andhra Pradesh were recorded as 4.22 lakh ha. and 7.37 lakh bales, respectively. According to the Cotton Corporation of India, India’s total cotton exports were stood at 2.8 million bales in 2023- 2024.

An efficient agricultural marketing system is always vital for ensuring remunerative prices to farmers, stabilizing markets and supporting economic development. Price forecasting plays a key role in reducing risks from seasonal fluctuations and guiding policy decisions. Traditional models like ARIMA and GARCH are widely used for capturing volatility (Engle, 2002) [5], while machine learning methods such as Artificial Neural Networks and Support Vector Regression offer robust, nonlinear forecasting capabilities. Ensemble models further enhance predictive accuracy by combining multiple approaches (Kerdprasop *et al*., 2022). Incorporating exogenous variables such as arrivals, climate and macroeconomic indicators improves both accuracy and policy relevance (Pindyck & Rubinfeld, 1998) [11]. By considering the above information, the present investigation had been planned to compare in between and among the univariate time series models (GARCH, ANN, SVR, Ensembling) and time Series models

(GARCH, ANN, SVR) with an exogenous (x=Arrivals) variables, as to obtain the best fitted model as well as to forecast the future Cotton prices for the Adoni market in Andhra Pradesh.

Materials and Methods

In this study, secondary time series data on market prices and arrivals of Cotton in Adoni market was collected from Agmarket (*www.agmarknet.gov.in*) during January 2010 to April 2025. Here various measures such as descriptive statistics and outlier detection techniques were applied initially as to understand the basic behaviour of the series. Later, the following univariate time series models and time series models with exogenous (*x= arrivals*) variable were employed to model and forecast the prices of cotton in Adoni market.

GARCH Model

GARCH is a Generalized Autoregressive Conditional Heteroskedastic model which includes past variances in the explanation of future variances. More specifically, GARCH is a time series technique that allows users to model and forecast the data series by considering conditional variance of the errors. If an ARMA model is assumed for the error variance, the model is called GARCH (Bollerslev, 1986) [1]. The basic GARCH model has two equations; one equation is to describe the behaviour of the mean and another to describe the behaviour of the variance. Here, mean equation (Y_t) is a stationary time series which may be either from a linear regression function that contains a constant or possibly some explanatory variables or it may be from AR model (Muanenda, 2018) [10].

GARCH Model with an Exogenous Variable (x)

In many practical applications (*particularly in economics and finance*), it is often desirable to account for the impact of external or exogenous variables on the mean or variance of the process. This extension in the GARCH model is referred to as the *GARCH* with exogenous model (Conrad & Karanasos, 2010) [2].
In the study, mean equation with exogenous variables had been developed by using mean equation of AR (1) model with the selected exogenous variable (x_{jt}); then thoserespective models would be specified as:

The GARCH model with exogenous variables (GARCH-X) extends the traditional GARCH by incorporating external factors (e.g., climate, policy or economic indicators) into the mean equation. Parameters are typically estimated using maximum likelihood estimation (MLE) and the inclusion of exogenous regressors enhances explanatory power and forecasting accuracy, especially in economic and financial series (Conrad & Karanasos, 2010; Francq & Zakoïan, 2010) [2, 6]. Unlike standard GARCH, GARCH-X provides a more comprehensive understanding of both mean and volatility dynamics, though in this study arrivals were included only in the mean equation.

Artificial Neural Network Model (ANN) Model

The ANNs are generally constructed by layers of units i.e., artificial neurons or nodes, hence termed as multilayer ANNs, such that each unit in a layer performs a similar task. The very first layer consists of the input units, which are statistically known as the independent variables. Similarly, the last layer contains the output units, statistically known as the response or dependent variables. The rest of units in the model are known as the hidden units and constitute the hidden layers.
In the present study, single hidden layer with multilayer feed forward network was employed in developing ANN model, which is considered as the most popular for time series modeling and forecasting (Rathod *et al*., 2017) [12]. This model is characterized by a network of three layers of simple processing units. The first layer is input layer, the middle layer is the hidden layer and the last layer is output layer, as shown in the Figure 1.

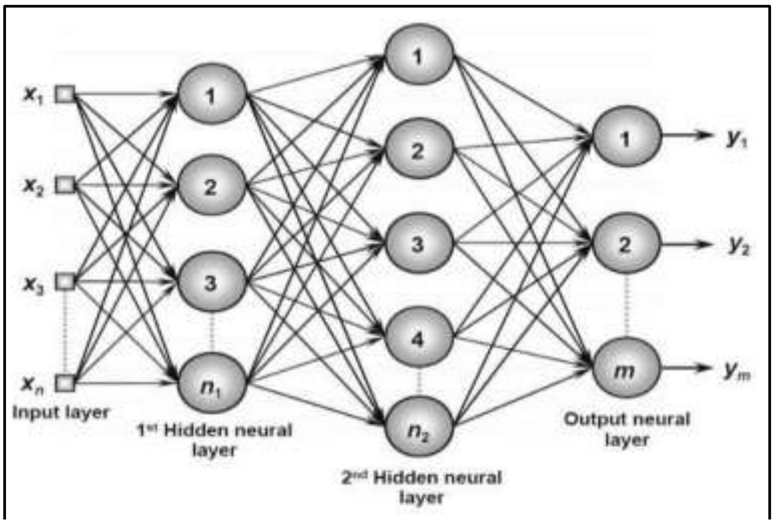


Fig 1: Architecture of Multilayer feed forward networks

Artificial Neural Network Model (ANN) with an Exogenous Variable (x)

Artificial Neural Networks (ANNs) are flexible, data-driven models capable of capturing complex, nonlinear relationships in time series data. When forecasting a time series (Y_t(, exogenous variables X_t can be incorporated as

additional inputs to improve predictive accuracy. When augmented with exogenous variables, the ANN-x model takes both past values of the target series and external regressors as input features, enabling improved forecasting performance (Crone & Kourentzes, 2010) [4].

Model Structure

An ANN model with one hidden layer and exogenous inputs can be represented as:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, X_t, X_{t-1}, \dots, X_{t-q}) + \varepsilon_t,$$

where:

- $f(\cdot)$ is a nonlinear function approximated by the neural network
- Y_{t-i} are lagged endogenous inputs (past values of the target)
- X_{t-j} are current and lagged exogenous inputs
- ε_t is the error term

Support Vector Regression Model (SVR) Model

Support Vector Regression (SVR) is a supervised learning algorithm designed to predict continuous values by identifying a suitable linear function (or hyperplane in higher dimensions). The primary objective of SVR is to determine a function that approximates the mapping from an input domain to real numbers by using training data. The key concept behind SVR is to find a hyperplane that fits the data within a defined threshold (Cristianini and Ricci, 2008) [3]. Unlike traditional regression models that aim to minimize the error between predicted and actual values, SVR instead fits the best line inside a margin of tolerance known as the epsilon (ε) tube.

Mathematical Formulation of SVR can be expressed as:

$$f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

To optimize the SVR model, the following objective function is minimized:

$$\min \{\sum^{n} \max [0, 1 - y\, f(x)] + \lambda \sum^{p} \beta^2\}$$

Support Vector Regression Model (SVR) with an exogenous variable (x)

Support Vector Regression with exogenous variables (SVRx) enhances the traditional SVR by including external regressor such as macroeconomic indicators or policy variables which influence the target variable but lie outside its past values. This is particularly useful in time series forecasting when external factors significantly impact the target trend

In SVRx, the feature matrix is augmented by appending exogenous inputs to the lagged values of the target variable. Formally, the SVRx model can be expressed as (Smola & Schölkopf, 2004) [14]:

$$y_t = f(x_t, z_t) + \varepsilon_t$$

where:

- y_t is the target variable (e.g., price) at time t,
- x_t represents lagged values of y_t,
- z_t is a vector of exogenous variables at time t,
- ε_t is the error term.

The SVRx model is trained similarly to standard SVR, but with an expanded input space. The svm() function from the e1071 package in R was also used here, incorporating both endogenous (lagged values of the target) and exogenous variables in the training matrix. As with the basic SVR, hyperparameter tuning was conducted via grid search using k-fold cross-validation.

Empirical Mode Decomposition (EMD)

The empirical mode decomposition is a self-adaptive time series decomposition technique to decompose non-linear and non-stationary time series data into several oscillatory functions (intrinsic mode functions) along with a trend component (residue). However, each of these IMFs should$i$=1i i

follow two conditions. Firstly, the number of extrema (sum of the number of maxima and minima) and the number of zero crossings should differ by at most one. Secondly, for an

This formulation includes a regularization term to avoid overfitting by penalizing large coefficients and a loss function that allows for deviations within an acceptable range (ε).

In this study, the SVR model was trained using the svm() function from the e1071 package in R software. Here, the radial basis function (RBF) kernel was selected to capture nonlinear patterns in price movements. In the present study, grid search approach was employed to identify the optimal SVR hyper-parameters. The model was tuned using 10-fold cross-validation over the following parameter space:

$C \in \{2^{-2}, 2^{-1}, 2^0, 2^1, \ldots, 2^9\}\gamma\epsilon \in \{0.1, 0.5, 1\}\{0.001, 0.01, 0.1, 0.2\}$

IMF, the mean value of the envelope defined by local maxima and the envelope defined by local minima must be zero at all points (Ghose *et al.*, 2024) [7].

The IMFs are extracted through a sifting procedure as follows.

- The local maxima and minima of the time series data (y_t) are identified.
- All the local maxima points are connected by a cubic spine function to create the upper envelope y_{up}. Similarly, the local minima points are utilised to form the lower envelope y_{low}.
- The mean envelope m_{11} (t) is formed by computing the mean values of the lower and upper envelopes.2
- **Cost (C):** The values {0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512} were tested across exponential ranges as this controls the penalty for misclassification (the regularization parameter).
- **Gamma (γ):** Gamma defines the shape of the RBF kernel — low values mean "far" influence, high values mean "close" influence.
- **Epsilon (ε):** Epsilon defines the margin of tolerance where no penalty is given for errors.In the next step, the mean envelope is subtracted from the actual data series.
  $h_{11}(t) = y_t - m_{11}(t)$
- The series $h_{11}(t)$ is then checked whether it is fulfilling all the necessary conditions of an IMF or not. If not, the sifting process is again followed on $h_{11}(t)$ until the necessary conditions are statisfied. The

process of obtaining the first IMF after the k^{th} iteration can be expressed as:
$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t) = c_t(t)$

- To ensure enough physical sense of both amplitude and frequency modulations, Huang *et al.* (1998) have
h^2
(t)
The value of SC lies between a predetermined limit of
0.2 to 0.3.
- After obtaining the first IMF, it is subtracted from the actual series,
$y_t - c_1(t) = r_1(t)$
Now, if $r_1(t)$ is not a monotonic function, then it is treated as a new series and the same sifting process is followed again to extract the second IMF.
- This sifting process is continued until the residue becomes a monotonic function from which no more IMF can further be extracted. The final residue after the extraction of the n^{th} IMF can be given as
$r_{n-1} - c_n(t) = r_n(t)$
Therefore, the actual data series is finally decomposed into the following form:

was preferred over others, for forecasting purposes. In addition to this diagnostics, Box-Pierce test was also employed to verify the residual of series were independently distributed or not.

Results and Discussion

From table 1, it was confirmed that there were no outliers detected by the Grubbs test during the study period. It was also observed that the prices of Cotton had varied from 2992.90 to 9931.90 (Rs./Q.) with an average of 5593.86 (Rs./Q.). Standard Deviation was recorded as 1592.69, which indicated that prices were dispersed highly over the months for Adoni market. Further, it was confirmed that the price series had no significant outliers and seasonal effects during the study period, as detected by Grubbs test and QS test respectively.

Table 1: Descriptive Statistics for the prices of Cotton in Adoni market

Cotton in Adoni Market	Prices (Rs./Quintal)
Mean	5593.86
Maximum	9931.90
Minimum	2992.90
Standard deviation	1592.69
Outlier detected (Grubbs test)	No
QS test (*Prob.*)	0.08

In order to employ the selected time series model (GARCH), stationarity of the data has to be examined first. For this, Augmented Dickey Fuller (ADF) test was applied to the market prices of Adoni. From table 2, it was also

concluded that the data series was non stationary and became stationary at first difference as the null hypothesis was not accepted at 5% LOS as p-value was 0.01 (<0.05).

Table 2: Result of ADF test for the prices of Cotton in Adoni market

Cotton	Data type	ADF statistic	Critical value (P value)	Decision
Adoni	ADF at level	-3.23	0.08	Data Non-Stationary
	ADF at 1st difference	-5.78	0.01	Data became Stationary

GARCH Model

Later, the GARCH model was developed by using mean equation (Y_t), as a stationary time series of Autoregressive (AR) model. Before employing the GARCH model, residuals of AR (1) model was verified for existence of ARCH effect using ARCH-LM test. It was found that the residuals of mean equation model had heteroskedastic nature, as due to significant prob. value (0.045) at 5% LOS. Hence GARCH (1,1) model was developed by using AR (1) model [as mean equation model ($Y_t = 218.30+0.96\ Y_{t-1} + e_t$] and their model fit statistics were depicted in Table 3.

Table 3: Estimated parameters GARCH model for the market prices of Cotton

Model form	Mean Equation		Variance Equation		
	Constant	AR1	Constant	ARCH effect (α1)	GARCH effect (β1)
AR(1)-GARCH (1,1)	218.3	0.96**	2783.00*	0.70**	0.11*

From Table 3 & Table 6, estimated parameters and diagnostics of the model were estimated as RMSE (379.15), MAPE (4.75) and MAE (253.68). Later, residual analysis was also carried out to check the adequacy of the selected model through ARCH-LM test, which further confirmed that the residuals had no heteroskedastic nature, as due to non-significant prob. value (0.68) at 5% LOS.

Artificial Neural Network (ANN) model:

Among the different Neural Network model combinations, the network structure with 3-8-1 (3 input nodes, 8 hidden

nodes with 1 output layer) was outperformed among all other neural networks with lower a RMSE (297.26), MAE (211.22) and MAPE (3.98) values, as per Table 4.

Further, residual analysis was also done as to check the adequacy of the selected ANN model and it was discovered that none of the lags of residual Auto Correlation Function chart were found to be significant as per Figure 2 and the Box-Pierce test was also informed the same as its p-value was 0.08 (>0.05), hence null hypothesis was accepted as residuals were independently distributed, which indicated good fit of the selected model i.e., NNAR (3-8-1).

Table 4: Performance of different Neural Network models

Network structure (Cotton in Adoni market)	RMSE	MAE	MAPE
3-1-1	362.22	253.23	4.69
3-2-1	355.86	253.49	4.69
3-3-1	341.46	244.02	4.54
3-4-1	333.86	236.13	4.39
3-5-1	318.27	228.22	4.27
3-6-1	312.16	225.96	4.24
3-7-1	306.76	219.80	4.12
3-8-1	297.26	211.22	3.98
3-9-1	299.59	212.74	4.01
3-10-1	299.87	214.57	4.04

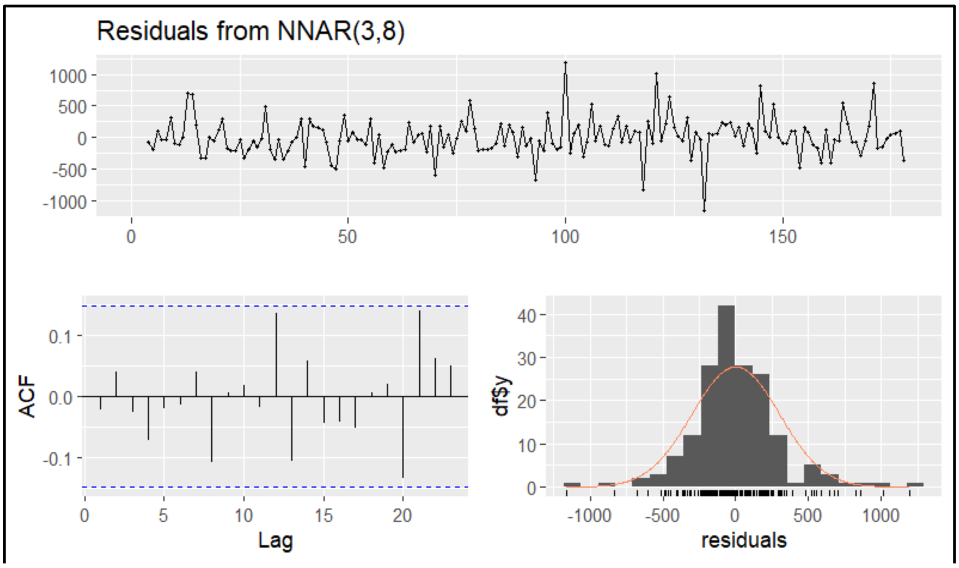


Fig 2: Residual plots for the prices of Cotton in Adoni market

Support Vector Regression (SVR) model

Later, the Radial Basis Function (RBF) kernel was used to fit the Support Vector Regression (SVR) model, as it is well-suited for modeling non-linear problems and can effectively handle complex data distributions. The parameter specifications of the SVR model were presented in Table 5 for the prices of Adoni market during the study period. According to this, the model employed an RBF kernel with 175 support vectors. Based on minimum cross validation error, the optimal parameters were determined to be a regularization parameter (C) of 4, a gamma value of 0.1 and an epsilon (ε) of 0.001. The model was found to perform best with an optimum lag length of 3. Furthermore, a residual diagnostic test using the Box-Pierce method yielded a p-value of 0.22 ($p>0.05$), indicated that the residuals were not autocorrelated. From table 6, the diagnostic performance metrics of the model were also computed, with Root Mean Square Error (RMSE) of 372.81, Mean Absolute Error (MAE) of 244.08 and Mean Absolute Percentage Error (MAPE) of 4.52.

Table 5: Model specification of SVR for the prices of Cotton in Adonimarket

Cotton Prices (Rs./Quintal)	**Model Specifications**
Kernel function	Radial Basis Function
No of support vectors	175
Cost (C)	4
Gamma (γ)	0.1
Epsilon (ε)	0.001
No of lags used	3

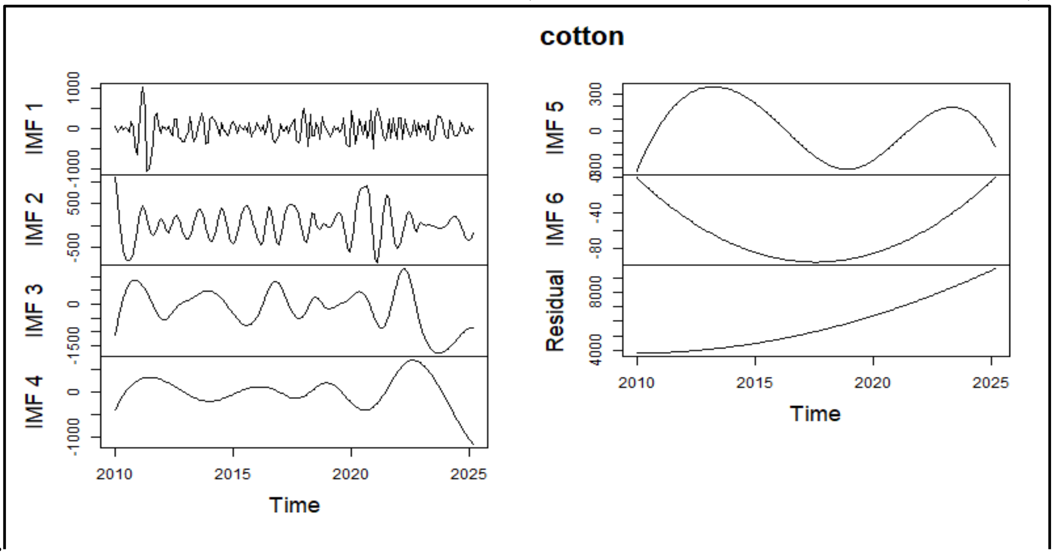
Table 6: Model fit statistics among the selected univariate time series models

Criterion\Model	**GARCH**	**ANN**	**SVR**
RMSE	379.15	297.26	372.81
MAPE	4.75	3.98	4.52
MAE	253.68	211.22	244.08

Ensemble Appraoch

Initially, Cotton price series was decomposed by EMD method into six IMFs (Intrinsic Mode Functions) and one residual component as illustrated in Figure 3. Subsequently, time series models such as GARCH, ANN and SVR were individually fitted to each IMF and the residual component using the usual procedures (Wu and Huang, 2009) [16]. The forecasted value from all IMFs and the residue were summed up to obtain the final forecasted value.

The decomposed cotton price series (six IMFs and one residual) was first tested for ARCH effects and after confirming the heteroskedasticity, accordingly GARCH (1,1) model was applied to each component and the fitted values were combined to form the EMD- GARCH model. Further, diagnostic checks confirmed that the residuals were non- autocorrelated. From table 7, the model performance was recorded as RMSE = 565.23, MAE = 449.86, and MAP



.

Fig 3: Decomposition plot for Cotton price in Adoni market

For, the EMD - ANN model, a multilayer feed - forward neural network with back-propagation was applied to each IMF and residual. Different network structures were trained, using a sigmoid activation function in the hidden layer and a linear identity function in the output layer. The best performing networks were selected based on error measures and their fitted values were combined to obtain the EMD- ANN model. It was also informed as the residuals were non- autocorrelated and from table 7, the model showed strong

erformance with RMSE = 211.53, MAE = 166.35 and MAPE=3.11.
Later, the EMD-SVR model was also developed by applying Support Vector Regression with a Radial Basis Function (RBF) kernel to each decomposed component. The combined fitted values produced the ensemble model. Residual analysis confirmed the absence of autocorrelation. From table 7, the model achieved RMSE = 253.13, MAE =
189.87 and MAPE = 3.64, demonstrating good predictive accuracy.

Table 7: Model fit statistics among the Ensemble univariate time series models to the prices of Cotton in Adoni market

Criterion\ Model	EMD-GARCH	EMD-ANN	EMD-SVR
RMSE	565.23	211.53	253.13
MAPE	7.61	3.11	3.64
MAE	449.86	166.35	189.87

GARCH with an Exogenous Variable (x=*Arrivals*) GARCH model was extended by including Arrivals (x) as an exogenous variable in the mean equation (Y_t), using a stationary time series of the Autoregressive (AR) model for the Cotton prices in Adoni market. Prior to applying the GARCH model, the residuals of the ARx(1) model were tested for the presence of ARCH effects and a GARCH(1,1) model was then constructed. From table 8, the final mean

equation was $Y_t = 3101.50+0.99\ Y_{t-1} +0.08\ x_t + e_t$. The inclusion of arrivals as an exogenous variable in the GARCH model improved its performance in capturing cotton price volatility. From table 11, the modified ARx- GARCH model achieved RMSE = 382.98, MAE = 260.47 and MAPE = 4.85. Further, the ARCH-LM test ($p = 0.56 > 0.05$) confirmed the absence of heteroskedasticity, validating the adequacy and robustness of the model.

Table 8: Estimated GARCH parameters with arrivals for the Cotton prices

Model form	Mean Equation			Variance Equation		
	Constant	AR1	x	Constant	ARCH effect (α1)	GARCH effect (β1)
ARx (1)-GARCH (1,1)	3101.50**	0.99**	0.08*	2635.01	0.01*	0.98**
** Significant at 1% level, * Significant at 5% level						

ANN with an exogenous variable (x=*Arrivals)*

A multilayer feed-forward neural network with backpropagation was employed to model cotton prices in the Adoni market by incorporating arrivals as an exogenous input. The optimal network structure was identified with three lags of the price series and the current arrival value, forming a 4-input design. From table 9, the network architecture 3-8-1 outperformed all others, achieving the lowest error values with RMSE = 275.36, MAE = 184.99, and MAPE = 3.46. Further, residual diagnostics were performed to evaluate the adequacy of the selected neural network model with exogenous input. The analysis of the residual ACF plot (Figure 4) revealed that none of the lag values were statistically significant, indicating the absence of autocorrelation in the residuals. This was further supported by the Box-Pierce test, which yielded a p-value of

0.37 (> 0.05), leading to the acceptance of the null hypothesis that residuals were independently distributed. These findings further confirmed the adequacy and robustness of the selected NNAR (3-8-1) model with Arrivals as an exogenous variable, indicating a good fit to the data.

Table 9: Performance of different Neural Network models (ANNx) for the prices of Cotton in Adoni market

Network structure (Cotton)	RMSE	MAE	MAPE
3-1-1	372.19	253.46	4.70
3-2-1	359.24	246.61	4.57
3-3-1	331.52	229.34	4.27
3-4-1	308.66	216.57	4.02
3-5-1	301.75	208.80	3.88
3-6-1	281.15	198.05	3.70
3-7-1	278.00	188.06	3.52
3-8-1	275.36	184.99	3.46
3-9-1	277.02	189.01	3.56
3-10-1	284.90	197.65	3.64

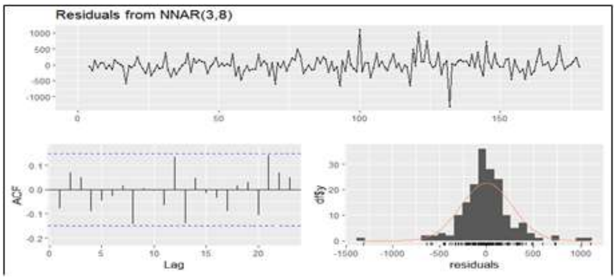


Fig 4: Residual plots of ANNx for the prices of Cotton in Adoni market

SVR with an exogenous variable (x=Arrivals)

The Support Vector Regression model with a Radial Basis Function (RBF) kernel was applied by incorporating cotton arrivals as an exogenous variable. From table 10, optimal parameters (Cost = 64, Gamma = 0.1, Epsilon = 0.01) with three lags were selected based on cross-validation, and the model employed 165 support vectors. From table 11, the SVR model demonstrated strong predictive accuracy with RMSE = 338.39, MAE = 203.83 and MAPE = 3.79.

Furthermore, the Box-Pierce test yielded a Chi-square statistic (χ^2) with a high p-value of 0.93, indicating that the residuals were free from autocorrelation and the model was well specified.

Table 10: Model specification of SVRx for the prices of Cotton in Adoni market

Cotton Prices (Rs./Quintal)	**Model Specifications**
Kernel function	Radial Basis Function
No of support vectors	165
Cost	64
Gamma	0.1
Epsilon	0.01
No of lags used	3

Table 11: Model fit statistics among the selected time series models with an exogeneous variable (x=*Arrivals*) to the Cotton market prices

Criterion\Model	**ARX-GARCH**	**ANNX**	**SVRX**
RMSE	382.98	275.36	338.39
MAPE	4.85	3.46	3.79
MAE	260.47	184.99	203.83

From Table-7 & Table-11, finally it was resulted that ensemble models (EMD-ANN) outperformed among the other selected models for the prices of cotton in Adoni market. In addition to this, Diebold-Mariano (DM) test also confirmed the same, as its test statistic was found to be significant over other model (p-value of 0.004 (< 0.05)), which indicating that the selected model had better forecasting accuracy. A similar kind of result was obtained by Wang *et al*. (2012) [15], Singh *et al*. (2020) [13] and Jain & Gupta (2023) [8], which reported that the EMD-ANN model was generally superior to ANN with exogenous inputs (such as arrival data) in the context of agricultural price forecasting due to its enhanced ability to model nonlinearity and non-stationary, robustness to data noise and irregularities and reduced dependency on external inputs.

Hence the EMD-ANN model was considered to forecasting the prices and from Figure 5, it was revealed that actual and forecasted values were close to each other, which also confirmed the appropriateness of the selected model. It was concluded as the there would be steady increase in the future prices and the prices of Adoni market for the period of Oct 2025, was forecasted as Rs. 7682.5 per Quintal as per Table 12.

Table 12: Predicted values of EMD-ANN for the prices of Cotton in Adoni market

Period	Actual	Forecasted	Forecast Error (%)
Nov-24	7124.40	7180.92	0.79
Dec-24	7083.80	7177.21	1.32
Jan-25	7269.40	7258.77	0.15
Feb-25	7223.20	7435.99	2.95
Mar-25	7358.40	7610.77	3.43
Apr-25	7527.00	7625.43	1.31
May-25		7655.07	
Jun-25		7684.53	
Jul-25		7725.39	
Aug-25		7609.23	
Sep-25		7681.23	
Oct-25		7682.51	

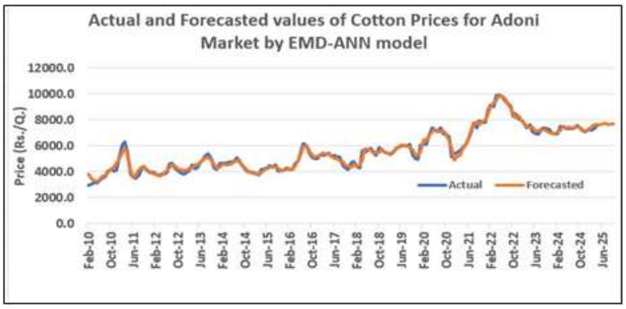


Fig 5: Actual and forecastedvalues of EMD-ANN for the prices of Cotton in Adoni market

Conclusion

From The study, it was revealed that the average price of Cotton in Adoni market during the study period (Jan 2010 to Apr 2025) was obtained as 5593.86 (Rs./Q.) with a simple growth rate (SGR) of 0.75% per month and Standard Deviation was recorded as 1592.69, which indicated that the prices were dispersed highly over the months. Here, among the selected univariate time series models (GARCH, ANN, SVR, EMD-GARCH, EMD-ANN, EMD-SVR), the optimal
model was recognized as EMD-ANN model due to the better diagnostic criterion i.e., RMSE (211.53), MAPE (3.11), MAE (166.35). Among the selected times series model (GARCH, ANN and SVR) with exogeneous variable (x=arrivals), the optimal model was considered as ANNx model due to the better diagnostic criterion i.e., RMSE (275.36), MAPE (3.46), MAE (184.99). As per comparative study among the two optimal models (EMD-ANN and ANNx), the most plausible model was recognised as EMD- ANN model due to its better diagnostics with lowest RMSE, MAPE, MAE values (*for both training and testing data* (significance of Diebold-Mariano (DM) test and non- significance of Box-pierce test. Finally, through use of the model, it was concluded as there would be steady increase in the future prices and the prices of Adoni market for the period of Oct 2025, was forecasted as Rs. 7682.5 per Quintal.

References

1. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. 1986;31(3):307-327.
2. Conrad C, Karanasos M. Modeling the persistence of conditional variances: A GARCH-X approach. Journal of Financial Econometrics. 2010;8(2):174-192.
3. Cristianini N, Ricci E. Support vector machines. In: Encyclopedia of Algorithms. Springer; 2008. p. 928- 932.
4. Crone SF, Kourentzes N. Input-variable selection for time series prediction: An empirical comparison of automatic and semi-automatic approaches. Neurocomputing. 2010;73(10-12):1920-1932.
5. Engle RF. New frontiers for ARCH models. Journal of Applied Econometrics. 2002;17(5):425-446.
6. Francq C, Zakoïan J-M. GARCH Models: Structure, Statistical Inference and Financial Applications. John Wiley & Sons; 2010.
7. Ghose B, Pandit P, Mazumder C, Sinha K, Sahu P. Comparative study of EMD-based modelling techniques for improved agricultural price forecasting. Journal of the Indian Society of Agricultural Statistics. 2024;78(1):53-62.
8. Jain A, Gupta SC. Peak load forecasting using machine learning algorithms. IEEE Renewable Energy and Sustainable E-Mobility Conference. 2023;14:1-4.
9. Kerdprasop K, Chuaybamroong P, Kerdprasop N. Food production forecasting with time series and ensemble modeling methods. In: Proceedings of the 5th Artificial Intelligence and Cloud Computing Conference (AICCC). Osaka, Japan; 2022. p. 59-65.
10. Nandigama, N. C. (2016). Teradata-Driven Big Data Analytics For Suspicious Activity Detection With Real-Time Tableau Dashboards. International Journal For Innovative Engineering and Management Research, 5(1), 73–78.
11. Pindyck RS, Rubinfeld DL. Econometric Models and Economic Forecasts. 4th ed. McGraw-Hill; 1998.
12. Todupunuri, A. . (2024). Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency in AI Systems to Ensure Responsible Deployment and Use of AI Technologies. International Journal of Innovative Research in Science,Engineering and Technology, 13(09), 1–14. https://doi.org/10.15680/ijirset.2024.1309002
13. Rathod S, Singh KN, Paul RK, Meher SK, Mishra GC, Gurung B, Ray M, Sinha K. An improved ARFIMA model using Maximum Overlap Discrete Wavelet Transform (MODWT) and ANN for forecasting agricultural commodity price. Journal of the Indian Society of Agricultural Statistics. 2017;71(2):103-111.
14. Charan Nandigama, N. (2020). An Integrated Data Engineering and Data Science Architecture for Scalable Analytical Warehousing and Real-Time Decision Systems. International Journal of Business Analytics and Research (IJBAR).
15. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004;14(3):199- 222.
16. Wang T, Zhang M, Yu Q, Zhang H. Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal. Journal of Applied Geophysics. 2012;83:29-34.
17. Wu Z, Huang NE. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009;1:1-41.
18. Zhang G, Eddy PB, Hu MY. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting. 1998;14(1):35-62.