International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(7): 163-171

https://www.mathsjournal.com
Received: 19-06-2025
Accepted: 22-07-2025

Ajay Sharma
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Joginder Kumar
Department of Statistics, Central University of Haryana, Mahendergarh, Haryana, India

Corresponding Author:
Ajay Sharma
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Statistical modelling and analysis of volatility of potato prices in wholesale markets of Northern India

Ajay Sharma and Joginder Kumar

Abstract
Potato (*Solanum tuberosum* L.) price volatility affects consumers, producers, and policymakers to support the needs of stakeholders. Therefore, the present study is planned to fit the statistical models for the analysis of volatility in potato prices of selected markets of Northern India using advanced time-series models such as Standard Generalized Autoregressive Conditional Heteroskedasticity (SGARCH), Glosten, Jagannathan, and Runkle Generalized Autoregressive Conditional Heteroskedasticity (GJR-GARCH) and Artificial Neural Network (ANN). For this, the time series data of monthly wholesale potato prices from January 2010 to December 2023 have been used for fitting and analysing the statistical models. The accuracy of the fitted models has been evaluated using various statistical measures, including the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC), the Root Mean Square Error and the Mean Absolute Percentage Error (MAPE). The findings of the present study revealed that ARMA (1,1)-GJR-GARCH (1,1), ARMA (1,1)-GJR-GARCH (1,1), ARMA (1,1) -SGARCH (1,1), ANN (12-6-1), ANN (12-8-1) and ANN (12-6-1) models have been found better among various fitted models for estimating the potato price volatility in selected monthly wholesale markets of Northern India. However, the effectiveness of the fitted models has also been evaluated using Relative Deviation (RD %), and found that the GJR-GARCH model has low RD (%) values as compared to other fitted models. This modelling approach can provide valuable insights to aid decision-making for farmers, traders, policymakers, and researchers involved in agricultural economics and market planning.

Keywords: Statistical modelling, volatility, potato prices, wholesale markets and time series

Introduction

The Potato (*Solanum tuberosum*) is a globally significant food crop, ranking fourth after rice, wheat, and maize. Native to the Andes Mountains in South America, it was introduced to Europe in the 16th century and later to Asia and Africa. Its adaptability to different climates and high yield per hectare make it crucial for ensuring food and nutritional security, in developing countries. According to the Food and Agriculture Organization**,** global potato production exceeds 370 million metric tons annually (FAO, 2023) [8]. Major producers include China, India, Russia**,** and Ukraine. The crop is cultivated on every continent except Antarctica and is increasingly being used in food processing industries globally. Asia contributes to over 50% of the global potato output**,** with China being the top producer, followed by India (FAO, 2022) [7]**.** The Asian potato market is scattered rapidly due to increasing demand in both fresh and processed forms. The crop's short growing season and adaptability make it ideal for food security in high-density populated regions (Gulati *et al.* 2022) [9]. India is the second-largest producer of potatoes**,** with about 53 million metric tons produced in 2022-23 (National Horticulture Board). Major potato-growing states include Uttar Pradesh**,** West Bengal**,** and Bihar**.** With high production, the sector faces challenges such as price volatility, inadequate storage, and changing weather conditions (Gulati *et al.* 2022) [9].

However, this growth has not brought relief to vegetable farmers. Faced with excess production, farmers resort to distress sales, burn surplus crops, or abandon them on roadsides. This leads to the question of why farmers are not benefiting even though production is at a record high. The main reasons include broken supply chains, unstable prices, loss of quality and quantity during handling, and a lack of proper processing facilities common issues in India's horticulture sector.

These challenges have diminished India's potential in the global horticulture trade and led to poor returns for farmers. The high variability in potato prices and yields, due to climate, storage, and market factors, forecasts techniques crucial for planning and decision-making. Autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH models, developed by Engle (1982) and Bollerslev (1986) [5] are used in time series data with changing variances, making these models suitable for analyzing agricultural price volatility. These models help forecast fluctuations in potato prices, allowing better management for farmers and traders. ANNs (artificial neural networks) are machine learning models that have a structure similar to that of the human brain. They are highly effective in capturing nonlinear relationships in data, which is often the case in agriculture, due to the complex interplay between weather, soil, inputs, and market conditions. ANNs have been used to predict crop yields, market trends, and weather-based production for crops such as potatoes from historical data (Kaul *et al.* 2005) [11]. By integrating the ARCH/GARCH and ANN approaches, policymakers and farmers can develop more reliable forecasts, reduce market risk, and improve crop planning and income stability.

The growing importance of potato cultivation globally, in Asia and India, highlights its economic and nutritional relevance. To examine the volatility and forecasting of potato production and prices, which directly impact farmers, policymakers, and supply chain in the development of potato price models, researchers have employed various techniques, such as symmetric and asymmetric generalized autoregressive conditional heteroskedasticity (GARCH) and artificial neural network (ANN) models, to estimate market prices; for example, Wang (2009) [21] used a hybrid asymmetric volatility approach in an artificial neural network option-pricing model to increase the ability to estimate derivative securities prices. Thus, the ANNS option-pricing model showed that GJR-GARCH volatility is more predictable than other volatility techniques. Lama *et al.* (2016) [1] evaluated the forecasting accuracy of time-delay neural networks and GARCH models in predicting volatility via monthly price data of edible oils across domestic and international markets. Dhiraj *et al.* (2017) [22] examined fluctuations in the arrivals and prices of potatoes in Agra, the leading potato-producing district in Uttar Pradesh. The results confirmed both negative and positive relationships across months between market arrivals and prices, as indicated by the correlation coefficients. Mitra & Paul (2017) [16] studied hybrid time series models for forecasting agricultural commodity prices, and for the data under consideration, the ARIMA-ANN hybrid model outperforms other combinations as well as their individual counterparts. Ahmed *et al.* (2018) [2] investigated the multivariate Granger causality relationships between oil prices, gold prices, and the KSE100 index using Johansen cointegration and GARCH models. Results from the ARCH and GARCH (1, 1) models indicated that crude oil prices (COP) had a significant impact on the volatility of returns in the KSE100 stock index. Dinku (2021) [6] studied price volatility in selected agricultural markets in Ethiopia using various GARCH models. The findings showed that the TGARCH model was the most suitable for capturing and forecasting the return volatility of Teff and Red Pepper prices in the country. Amirshahi and Lahmiri (2023) [4] applied deep learning techniques to enhance the forecasting performance of GARCH-type models, regardless of the assumed distribution. They found that incorporating informative features from GARCH-type model forecasts significantly improved the prediction accuracy of deep learning models, specifically the Deep Feedforward Neural Network (DFFNN) and Long Short-Term Memory (LSTM) models. Shankar *et al.* (2024) [19] compared time series models for potato price volatility and found that the EEMD-ARIMA model gave the most accurate forecasts for the Dehradun market, with a MAPE of 12.97%. Kumar *et al.* (2024) [14] studied the potato price forecasting using a hybrid model combining Singular Spectrum Analysis (SSA) with a Time Delay Neural Network (TDNN). They compared this approach with ARIMA, SSA-R, and SSA-ARIMA models across Agra, Delhi, and Bengaluru markets. Using the Diebold–Mariano test, they found the hybrid SSA-TDNN model delivered the most accurate forecasts.

Alhussan *et al.* (2024) [3] compared ARIMA and ETS models for forecasting global potato production, focusing on China, India, and the USA. The ETS model outperformed ARIMA in long-term predictions, highlighting China’s and India’s continued dominance by 2027. The study also stressed the challenges in agricultural forecasting and its relevance to global food security. Kumari *et al.* (2024) [15] emphasized the importance of accurate potato price forecasting for improving production planning, marketing, and inventory control. Their findings demonstrated that the stacked ensemble model significantly enhanced prediction accuracy, supporting more efficient decision-making and resource allocation in the potato industry.

Data and Methodology

Monthly time-series data of average wholesale prices (₹/quintal) of potato were obtained from the AGMARKNET portal from January 2010 to December 2023. This data is compiled and maintained by the Directorate of Marketing and Inspection, functioning under the Ministry of Agriculture, Government of India. The markets selected for the study, Haryana, Delhi, and Kanpur, were chosen based on the maximum arrivals of potato. (Source: https://agmarknet.gov.in/).

Nonlinear Time-Series Models

In time series, the widely used Box-Jenkins ARIMA model cannot effectively capture price volatility or deal with situations where the variance changes over time. This is because it assumes the data series is linear, stable (stationary), and has constant variance (homoscedasticity), which may not be true in agricultural commodity price markets, stock markets, and financial data. Therefore, to handle the volatile behavior of the data, it is important to consider using nonlinear time series models. The Autoregressive Conditional Heteroscedasticity (ARCH) model, introduced by Engle in 1982, is the most primitive parametric nonlinear time series models developed to analyse volatility. However, the ARCH model has certain limitations, including a rapid decrease in the autocorrelation of squared residuals, which affects its ability to model persistent volatility. To overcome these limitations, Bollerslev (1986) [5] extended the model by introducing the Generalized ARCH (GARCH) model. Moreover, to improve the performance of individual models, researchers have created several hybrid models that integrate multiple time series methods (Khashei *et al.*, 2012; Pagan and Schwert, 1990) [12, 17]. It is important to note that standard parametric models are limited because they depend on specific distribution assumptions. In contrast, nonparametric models are more flexible and have been shown to perform more efficiently in a wide range of situations. Against this background, nonparametric nonlinear time series modelling has increasingly captured the interest of researchers. Among these methods, machine learning techniques, particularly Artificial Neural Networks (ANNs), are an effective approach in modern time series analysis. While the widely used GARCH model is helpful in modelling volatility, it has limitations in capturing the asymmetric impact of positive and negative shocks on volatility. Since GARCH models focus on the squared values of returns, they do not distinguish between upward and downward movements. However, asymmetric models are designed to account for the leverage effect where an unexpected drop in prices leads to a larger increase in volatility compared to a similar unexpected price rise. In simpler terms, the leverage effect refers to the negative relationship between past returns and current volatility. In 1993, Glosten, Jagannathan, and Runkle introduced an improved asymmetric model known as the GJR-GARCH model. This model extends the standard GARCH by allowing for seasonal variations in volatility and by incorporating nominal interest rates to better predict conditional variance. The GJR-GARCH model is particularly useful for capturing the asymmetric behavior of volatility in response to market shocks.

Generalized Autoregressive Conditional Heteroscedasticity Model

Autoregressive conditionally heteroscedastic (ARCH) models were introduced by Engle (1982) and their generalized autoregressive conditionally heteroscedastic (GARCH) extension is due to Bollerslev (1986) [5]. In these models, the main concept is the conditional variance, that is, the variance conditional on the past. In the classical GARCH models, the conditional variance is expressed as a linear function of the squared past values of the series. At the same time, it is simple enough to allow for a complete study of the solutions. We first present definitions and representations of GARCH models.

The ARCH model is used when the error variance in a time series follows an autoregressive (AR) model, if an autoregressive moving average (ARMA) model is assumed for the error variance, the model is a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The general functional forms of GARCH (m, n) models to be considered are:

Mean equation of the model

$$y_t = \theta_0 + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t$$

Where,

y_t: *the value of the time series at time t*

θ_0: *intercept term*

$\theta_1,\ \theta_2, \dots \dots \theta_p$: *coefficients that measure the influence of the past values*

$y_{t-1}, y_{t-2}, \dots \dots .. y_{t-p}$: past p observations (lags).

ε_t: *white noise error term at time t*

Variance equation of the GARCH model

$$\sigma_t^2 = \alpha_0 + \sum_{i=i}^{m} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{n} \beta_j \sigma_{t-j}^2$$

Where,

σ_t^2: *the conditional variance (volatility) at time t*

α_0: *a constant,* α_i: *coefficients for the lagged squared residuals*

ε_{t-i}^2: *representing ARCH effect*

β_j: *coefficients for the lagged conditional variance*

σ_{t-j}^2: *representing GARCH effect*

Asymmetric Volatility Model

In the case of a volatile series, the effects of positive and negative shocks on the volatility are not always equal. There is a negative link between changes in return volatility and stock returns, meaning that when there's bad news and returns are lower than expected, volatility usually increases. However, when there's good news and returns are higher than expected, volatility doesn't rise as much. To capture this kind of behavior, the GJR-GARCH model is a good choice.

A commonly used model to incorporate Asymmetric volatility was developed by Glosten, Jagannathan, and Runkle (1993). The model is called GJR-GARCH and an advantage of the model is that the variance is directly modelled and does not use the natural logarithm like the E-GARCH model. This means that the GJR-GARCH is simpler to implement in practice. Studies that have applied several GARCH models have deemed the GJR-GARCH the most sufficient in forecasting volatility. The GJR-GARCH model is stated in the equation below

$$\sigma_t^2 = \omega + \alpha_1 e_{t-1}^2 + \beta_1 s_{t-1}^2 + \gamma_1 e_{t-1}^2 I_{t-1}$$

Where,

ε_{t-1} is the lagged residual,

$I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$ (indicating a negative shock), and 0 otherwise,

α_0 sets a baseline level of volatility,

$\alpha_1, \gamma_1 and\ \beta_1$ are model parameters,

γ_1 captures the asymmetric effect and leverage effect,

σ_t^2 is the actual estimated volatility

Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are effective in capturing non-linear patterns in time series data. Their main strength lies in being universal function approximators, meaning they can model a wide variety of relationships accurately. This ability is largely due to their parallel processing structure. Another key advantage is that ANNs do not require any prior assumptions about the data or underlying model, making them flexible and adaptive in the modelling process.

Let, y_t is time series consist of N $(y_1, y_2, y_3, \dots, y_N)$ observations, it is divided into three sets N_1 $(y_1, y_2, y_3, \dots, y_{N_1})$, N_2 $(y_{N_{1+1}}, y_{N_{1+2}}, y_{N_{1+3}}, \dots, y_{N_2})$ and N_3 $(y_{N_{2+1}}, y_{N_{2+2}}, y_{N_{2+2}}, \dots, y_N)$ for training, testing and validation respectively. The input nodes X_1, X_2,..., Xn, which represent lagged values from the training dataset (N_1), are essential for capturing the autocorrelation patterns in a time series. The number of output nodes is generally straightforward to determine. In this study, a single output node is used, and multi-step ahead forecasting is carried out iteratively. For example, the first input node X_1 contains the values y_1, y_2, y_3,..., y_N₁ and is used to predict y_(N₁+1). Similarly, the second input node X_2 contains y_2, y_3,..., y_(N₁+1) to forecast y_(N₁+2), and this process continues in the same way.

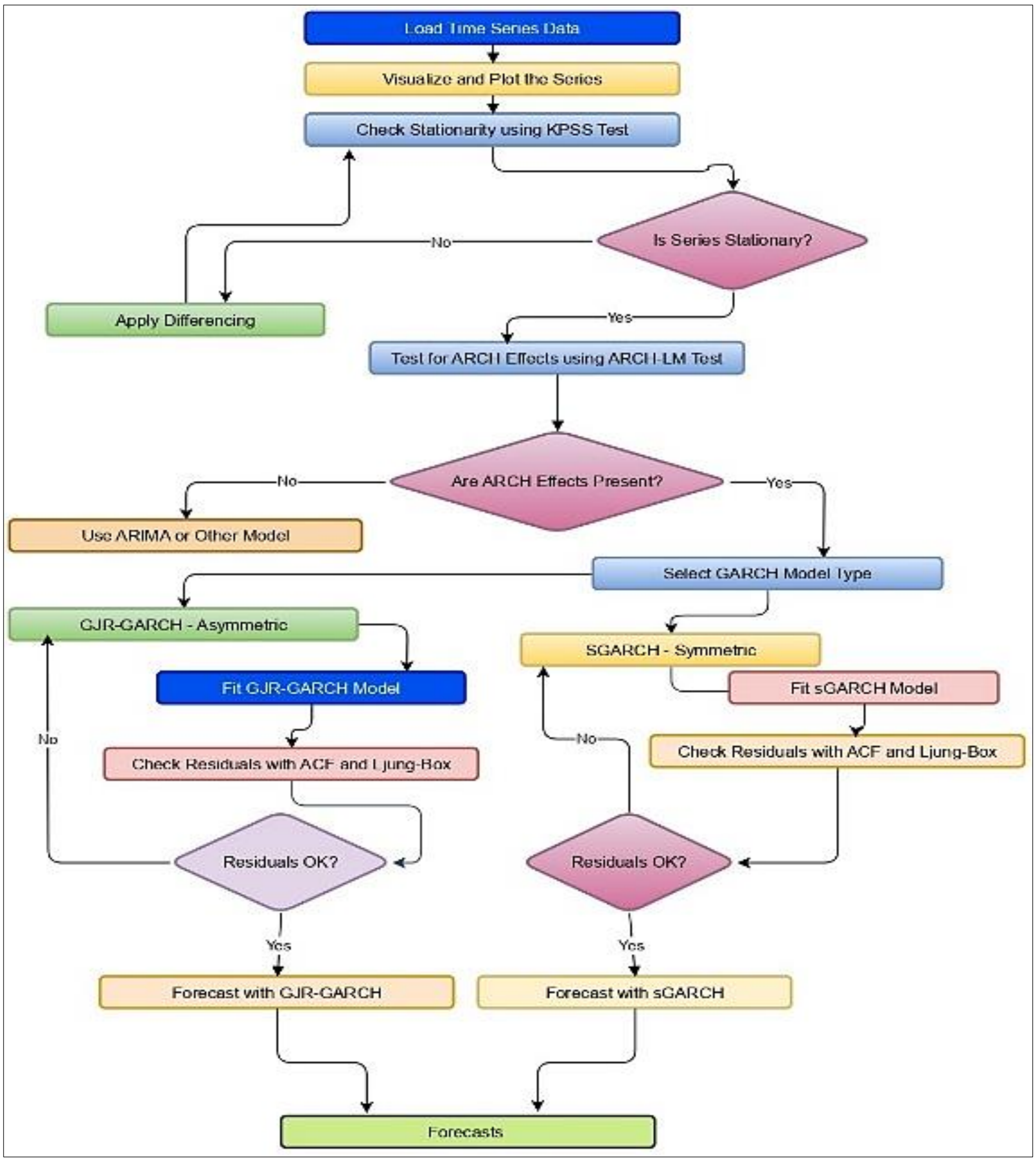

Fig 1: Flowchart of SGARCH and GJR-GARCH model

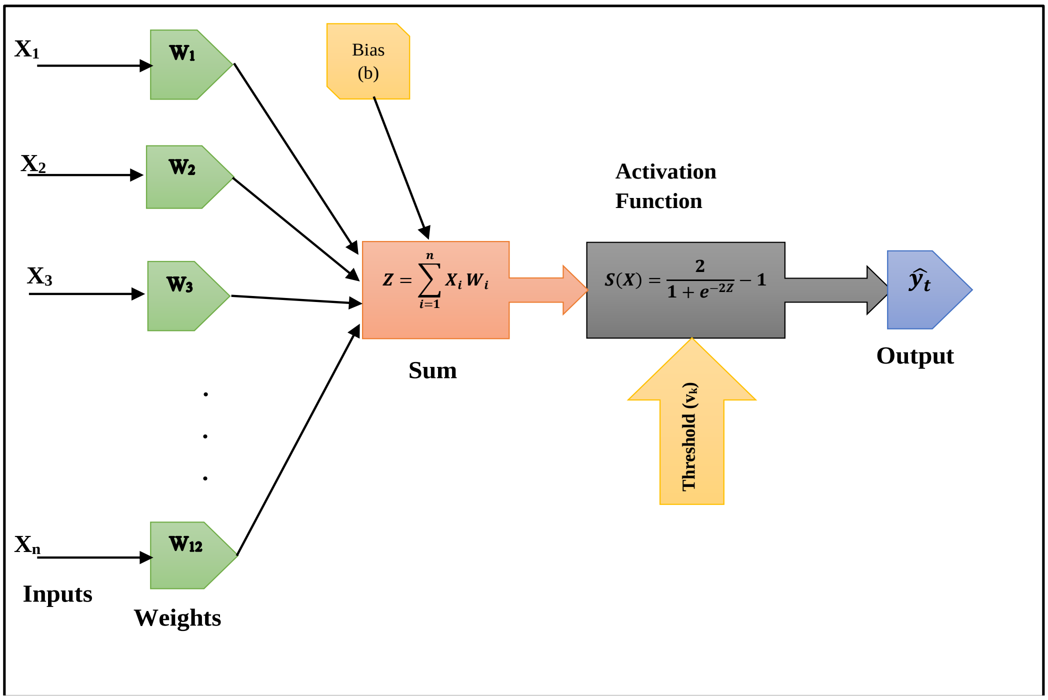

Fig 2: Artificial Neural Network structure based on feed-forward back-propagation

The output is determined as

$$g(X_i) = b + g\left(\sum_{j=1}^{h} W_j f\,(Z) + v_k\right)$$

Where, b is the bias term (constant), $g(.)$ is a linear output function, W_j is the connection weights between hidden layers and output units, and v_k is the threshold term. The threshold term is the baseline input to a node in the absence of any other inputs. $f(.)$ is the activation function of the hidden layer

The following steps used in the backpropagation algorithm

1. Determine the rate at which the error changes when the activity of an output unit changes. This error rate of change (EA) is the difference between the actual and predicted activity.

$$\mathcal{E}A_i = \frac{\partial \mathcal{E}}{\partial y_i} = (y_i - \hat{y}_i)$$

2. Determine how quickly the error changes when the total input received by an output unit varies?

$$\mathcal{E}I_i = \frac{\partial \mathcal{E}}{\partial Z_i} = \frac{\partial \mathcal{E}}{\partial y_i} \times \frac{\partial y_i}{\partial Z_i} = \mathcal{E}A_i y_i (1 - y_i)$$

3. Determine how quickly the error changes when the weight of the connection to an output unit is adjusted?

$$\mathcal{E}W_{ij} = \frac{\partial \mathcal{E}}{\partial W_{ij}} = \frac{\partial \mathcal{E}}{\partial y_i} \times \frac{\partial Z_i}{\partial W_{ij}} = \mathcal{E}I_i y_i$$

4. Determine how fast the error changes in response to a change in the activity of a unit from the preceding layer?

$$\mathcal{E}A_i = \frac{\partial \mathcal{E}}{\partial y_i} = \sum_i \frac{\partial \mathcal{E}}{\partial y_i} \times \frac{\partial Z_i}{\partial y_i} = \sum_i \mathcal{E}I_i\, W_{ij}$$

This step allows backpropagation to be applied in multilayer networks. By using the second and fourth steps, we can pass the error derivatives from one layer to the layer before it. This process is repeated as many times as needed to compute the error derivatives for all previous layers. Once a unit's EA is known, we can calculate the error derivatives for the weight of its connections by applying the second and third steps.

The structure of an artificial neural network (ANN) for time series forecasting involves the number of layers and the number of nodes in each layer. Since there is no fixed theoretical approach to define these parameters, they are selected through experimentation. In this study, a neural network with a single hidden layer is used. Using fewer nodes in the hidden layer, is generally preferred, as it improves forecasting accuracy and helps prevent overfitting.

Results and Discussion

Advanced time-series models such as GARCH and ANN have been applied to analyse wholesale price volatility of potato in selected. These models have also been employed to capture both linear and nonlinear patterns in monthly wholesale prices across selected markets. In Fig.1, the step-by-step process has been given for modelling the time series data using GARCH techniques. It begins with checking for stationarity & ARCH effects, then guides model selection between symmetric (SGARCH) and asymmetric (GJR-GARCH) forms of GARCH. Finally, model adequacy has been assessed through residual diagnostics. Fig 2, illustrates the structure of an artificial neurons. In this, inputs were multiplied by corresponding weights and summed with a bias term and the resulted sum has been passed through an activation function to generate the final output $\hat{y}_t$. The results of the modelling and their comparative performances are presented below**:**

Table 1: KPSS, LM, and Jarque-Bera tests for prices of potato

Markets	KPSS test		LM test		Jarque-Bera	
	Statistic	**p -Value**	**Statistic**	**p -Value**	**Statistic**	**p -Value**
Haryana	0.04	0.10	105.47	0.02	163.94	<0.01
Delhi	0.04	0.10	96.76	<0.01	443.26	<0.01
Kanpur	0.04	0.10	116.14	0.01	111.80	<0.01

Note: H0: Returns series is stationary H0: no ARCH effect in Returns series H0: Returns series is normally distributed

H_1: Returns series is not stationary H_1: ARCH effect in Returns series H_1: Returns series is not normally distributed

In Table 1, the values of KPSS test have been found to be not significant at the 5% level of significance which indicate that return series are stationary. The values of LM test shows a significant presence of heteroscedasticity (volatility clustering) in selected markets as p-values less than 0.05. The results of Jarque-Bera test show that price distributions deviate from normality (p<0.01). Thus, the volatility behavior of potato prices in Haryana, Delhi, and Kanpur markets is non-normal.

One of the key techniques for analysing the properties of time series data is the correlogram., which usually represents the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). The ACF measures the correlation between X_t and X_{t+k}, where k is the number of lead periods in future. And, the PACF is the correlation between X_t and X_{t+k} after removing the effect of intermediate observations. The phenomenon of time-varying variance, known as volatility clustering, is observed when periods of high returns tend to follow other high returns, and periods of low returns follow low returns. To evaluate the statistical characteristics of the price return series, both ACF and PACF were computed. Figures 3 to 5 illustrate the ACF and PACF plots, respectively, for each price return series. The PACF helps to identify the autoregressive order (p), while the ACF is used to determine the moving average order (q).

Fig 3: Plots of Original series, return series, ACF of Log returns, PACF of Log returns, ACF of squared Log returns, and PACF of squared Log returns, for prices of Potato in Haryana

Fig 4: Plots of Original series, return series, ACF of Log returns, PACF of Log returns, ACF of squared Log returns, and PACF of squared Log returns, for prices of Potato in Delhi

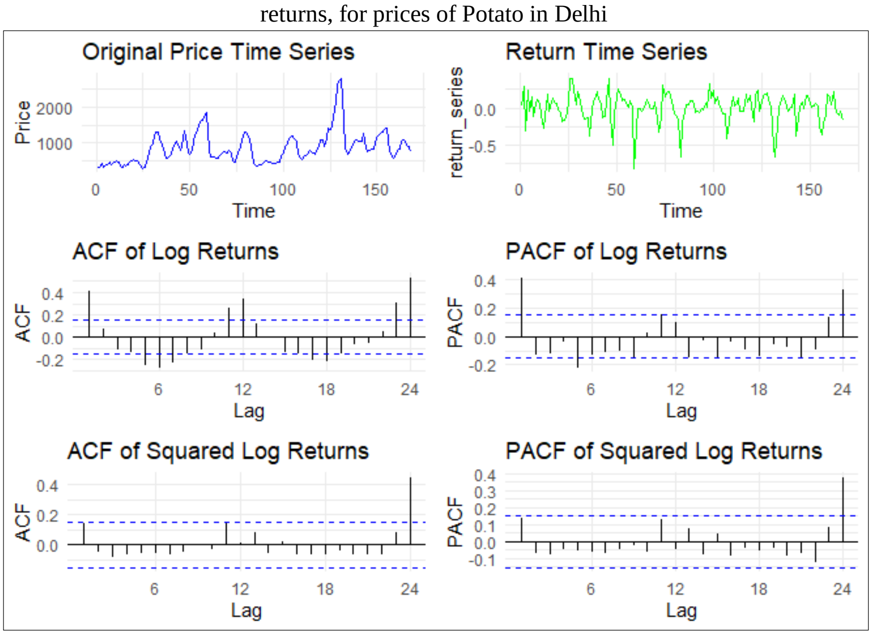

Fig 5: Plots of Original series, return series, ACF of Log returns, PACF of Log returns, ACF of squared Log returns, and PACF of squared Log returns, for prices of Potato in Kanpur

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) have been used to determine the appropriate model. The results of AIC and BIC suggest that ARMA (1,1)-GJR-GARCH (1,1), ARMA (1,1)-GJR-GARCH (1,1), and ARMA (1,1)-SGARCH (1,1) based on Student's -t-distribution provide better fits for evaluating price volatility in Haryana, Delhi, and Kanpur, respectively.

Table 2: Estimated parameters of ARMA (1,1)-GJR-GARCH (1,1), ARMA (1,1)-GJR-GARCH (1,1), and ARMA (1,1)-SGARCH (1,1) models for prices (returns series) of potato:

Model	Parameter	Estimate	S. E.	p-value
Haryana ARIMA (1,1)-GJRGARCH (1,1)	μ	0.47	0.19	<0.01
	AR_1 ($\emptyset_1$)	0.92	0.04	<0.01
	MA_1(θ_1)	-0.99	0.03	<0.01
	Omega	159.45	334.30	0.60
	α_0	0.85	1.17	0.40
	β_1	0.11	1.22	0.90
	Skew	1.05	0.10	<0.01
	Shape	3.22	3.63	<0.01
Delhi ARIMA (1,1)-GJRGARCH (1,1)	μ	4.70	1.56	<0.01
	AR_1 ($\emptyset_1$)	0.15	0.26	0.50
	MA_1(θ_1)	0.01	0.27	0.90
	Omega	72.14	184.62	0.60
	α_0	2.68	0.18	<0.01
	β_1	9.54	0.13	<0.01
	γ_1	0.01	0.17	0.90
	Skew	0.69	0.11	<0.01
	Shape	2.25	0.24	<0.01
Kanpur ARIMA (1,1)-SGARCH (1,1)	μ	3.03	1.42	0.03
	AR_1 ($\emptyset_1$)	0.23	0.11	0.03
	MA_1(θ_1)	0.15	0.11	0.16
	Omega	0.56	5.10	0.01
	α_0	0.01	0.01	0.01
	β_1	0.99	0.04	<0.01
	Skew	1.10	0.26	<0.01
	Shape	3.35	0.30	<0.01

In Table 2, the values of estimated parameters (AR_1, MA_1, α_1, β_1, γ_1 skew, shape) of the selected models have been found highly significant at 5% level of significance for all selected markets. The sufficient condition of conditional variance for the SGARCH and GJR-GARCH model has also been satisfied as the value of β_1 is greater than zero and less than one. The conditional variance equations for the selected models are given as follows:

ARMA (1,1)-GJR-GARCH (1,1) model for Prices of potato in Haryana

$y_t = 0.660 + 0.394y_{t-1} - 0.320\sigma_{t-1}^2 + \varepsilon_t$
(conditional mean equation)

$\sigma_t^2 = 19.47 + 0.283\varepsilon_{t-1}^2 - 0.077\gamma_1\varepsilon_{t-1}^2 I_{t-1} + 0.99\sigma_{t-1}^2$
(conditional variance equation)

ARMA (1,1)-GJR-GARCH (1,1) model for Prices of potato in Delhi

$y_t = 4.70 + 0.155y_{t-1} + 0.012\sigma_{t-1}^2 + \varepsilon_t$
(conditional mean equation)

$\sigma_t^2 = 72.14 + 2.684\varepsilon_{t-1}^2 + 0.012\gamma_1\varepsilon_{t-1}^2 I_{t-1} + 9.540\sigma_{t-1}^2$
(conditional variance equation

ARMA (1,1)-SGARCH (1,1) model for Prices of potato in Kanpur

$y_t = 3.036 + 0.234y_{t-1} + 0.157\sigma_{t-1}^2 + \varepsilon_t$
(conditional mean equation)

$\sigma_t^2 = 0.565 + 0.001\varepsilon_{t-1}^2 + 0.995\sigma_{t-1}^2$
(conditional variance equation)

Table 3: ARCH-LM Summary Statistics

Markets	Ljung- Box		ARCH-LM	
	Statistic	p-value	Statistic	p-value
Haryana	0.20	0.65	0.12	0.72
	0.73	0.91	0.74	0.81
	28.8	0.77	2.73	0.56
Delhi	3.08	0.07	0.52	0.47
	4.35	0.06	1.13	0.69
	7.45	0.09	1.92	0.73
Kanpur	0.21	0.60	0.38	0.53
	1.21	0.80	1.12	0.695
	2.48	0.83	2.11	0.69

Note H_0: no autocorrelation in residuals H_0: no ARCH effect in residuals H_1: autocorrelation in residuals H_1: ARCH effect in residuals

From Table 3, the values of Ljung-Box and ARCH-LM tests revealed that there is no significant autocorrelation or heteroskedasticity in the residuals of prices of tomato in all selected markets.

Artificial Neural network: The Artificial Neural Networks (ANNs) are effective in capturing the non-linear patterns of time series data. For this, the Brock- Dechert-Scheinkman (BDS) has been used to check the linearity of the data. The results of BDS test have been found statistically significant ($p < 0.05$), confirming the non-linearity in the prices of potato in all the selected markets.

Table 4: Selection criteria of ANN models

Neural network structure					**Training set**		**Testing set**	
Activation function	Input node	Hidden node	Rep.	Output node	RMSE	MAPE	RMSE	MAPE
Haryana								
Logistic	12	3	1	1	50.53	5.64	31.15	2.75
	12	7	1	1	41.33	3.86	25.47	1.91
Tanh	12	5	1	1	20.28	1.08	11.38	0.81
	12	6	1	1	11.16	0.34	10.36	0.61
Delhi								
Logistic	12	7	1	1	40.74	4.01	44.92	5.03
	12	9	3	1	19.80	0.93	14.38	1.55
Tanh	12	6	1	1	11.71	0.46	9.62	0.55
	12	8	1	1	13.16	0.40	6.44	0.41
Kanpur								
Logistic	12	3	1	1	22.00	1.35	11.04	0.65
	12	1	1	1	27.86	1.48	21.04	1.74
Tanh	12	1	1	1	22.86	1.01	17.79	1.30
	12	6	1	1	11.65	0.43	4.29	0.37

In Table 4, the neural network models with tanh activation function outperformed as compared to logistic activation function based on values of RMSE and MAPE in both training and testing sets. Also, the ANN with hidden nodes ranging from 5 to 8 and 12 input nodes gave better results based on RMSE and MAPE values.

Comparative performance of the selected models

Table 5: Comparison of selected GARCH and ANN models for Potato markets based on RD (%)

Month	**RD%**					
	GJR-GARCH (1, 1) (1, 1) Student's t distribution	ANN (12-6-1) (tanh)	GJR-GARCH (1, 1) (1, 1) Student's t-distribution	ANN (12-6-1) (tanh)	SGARCH (1, 1) (1, 1) Student's t-distribution	ANN (12-6-1) (tanh)
	Haryana		Delhi		Kanpur	
Jan	3.67	11.08	1.59	9.19	1.17	1.32
Feb	1.58	3.33	8.10	13.41	5.58	18.62
Mar	1.53	8.54	10.81	12.86	7.42	31.88
April	2.03	13.54	1.26	2.93	2.70	17.25
May	1.22	3.53	1.81	13.76	1.22	18.34
June	2.20	8.93	7.32	8.09	2.38	28.12
July	2.83	4.57	7.39	7.68	1.97	2.99
Aug	1.92	2.83	7.81	8.29	2.98	3.14
Sept	4.56	12.42	8.43	8.89	1.98	13.82
Oct	3.14	9.00	8.41	8.81	4.16	20.14
Nov	1.18	3.33	8.76	9.22	7.25	25.89
Dec	5.64	15.27	2.99	10.56	3.20	29.59

Table 5, revealed that GJR-GARCH model gave better results as compared to ANN models in Haryana and Delhi, with lower RD% values throughout the year. But in Kanpur, the SGARCH model gave the better results, with lower RD% values than ANN model in all months. Thus, overall, the GARCH models were more accurate than the ANN models in estimating volatility of potato prices in all selected markets.

Conclusion

In the present study, ARMA (1,1)-GJR-GARCH (1,1) model with Student's t-distribution and ANN (12-6-1) model with tanh activation function have been used to estimate the volatility of monthly wholesale prices of potato in Northern India. The GJR-GARCH model was found effective in capturing the volatility of prices & market shocks, while the ANN model handled non-linear patterns well but was found less consistent during highly volatile periods. Also, the GJR-GARCH model have been given more accurate results than the ANN model based on RD (%) values in all selected markets. Therefore, based on the basis of present study, the GJR-GARCH model was found more reliable as compared to ANN model for estimating the volatility of wholesale prices of potato in selected markets in Northern India. These findings can be useful for farmers, traders, and policymakers to reduce the impact of price fluctuations and support better planning and decision-making in the potato supply chain.

References

1. Lama A, Jha GK, Gurung B, Paul RK, Bharadwaj A, Parsad R. A comparative study on time-delay neural network and GARCH models for forecasting agricultural commodity price volatility. 2016.
2. Ahmed RR, Vveinhardt J, Streimikiene D. Multivariate granger causality among oil prices, gold prices, and KSE100: Evidence from Johansen Cointegration and GARCH models. Acta Montan Slovaca. 2018;23(2):216-231.
3. Alhussan AA, Khafaga DS, Abotaleb M, Mishra P, El-Kenawy ESM. Global potato production forecasting based on time series analysis and advanced waterwheel plant optimization algorithm. Potato Res. 2024;1-36.
4. Amirshahi B, Lahmiri S. Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies. Mach Learn Appl. 2023;12:100465.
5. Bollerslev T. A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev Econ Stat. 1987;542-547.
6. Dinku T. Modeling price volatility for selected agricultural commodities in Ethiopia: The application of GARCH models. Int J Environ Sci. 2021;6.
7. Food and Agriculture Organization of the United Nations (FAO). FAO Statistical Databases. 2022.
8. Food and Agriculture Organization of the United Nations (FAO). FAO Statistical Databases. 2023.
9. Gulati A, Wardhan H, Sharma P. Tomato, onion and potato (TOP) value chains. Agric Value Chains India. 2022;3:33-97.
10. Harshith N, Kumari P. Memory based neural network for cumin price forecasting in Gujarat, India. J Agric Food Res. 2024;15:101020.
11. Kaul M, Hill RL, Walthall C. Artificial neural networks for corn and soybean yield prediction. Agric Syst. 2005;85(1):1-18.
12. Khashei M, Bijari M, Ardali GAR. Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs). Comput Ind Eng. 2012;63(1):37-45.
13. Kristjanpoller W, Minutolo MC. Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Syst Appl. 2016;65:233-241.
14. Kumar P, Jha GK, Kumar RR, Lama A, Mazumder C. Leveraging singular spectrum analysis and time delay neural network for improved potato price forecasting. Potato Res. 2024;1-21.
15. Kumari P, Vekariya P, Kujur SN, Macwan J, Mishra P. Predicting potato prices in Agra, UP, India: an H2O AutoML approach. Potato Res. 2024;1-16.
16. Mitra D, Paul RK. Hybrid time-series models for forecasting agricultural commodity prices. Model Assist Stat Appl. 2017;12(3):255-264.
17. Pagan AR, Schwert GW. Alternative models for conditional stock volatility. J Econom. 1990;45(1-2):267-90.
18. Rao A, Sharma GD, Tiwari AK, Hossain MR, Dev D. Crude oil price forecasting: Leveraging machine learning for global economic stability. Technol Forecast Soc Change. 2025;216:124133.
19. Shankar SV, Chandel A, Gupta RK, Sharma S, Chand H, Aravinthkumar A, Ananthakrishnan S. Comparative study on key time series models for exploring the agricultural price volatility in potato prices. Potato Res. 2024;1-19.
20. Singh DK, Pynbianglang K, Pandey NK. Market arrival and price behaviour of potato in Agra district of Uttar Pradesh. Econ Aff. 2017;62(2):341-345.
21. Wang YH. Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach. Expert Syst Appl. 2009;36(1):564-570.
22. Dhiraj K, Sharma A, Singh R. Title of the article. J Agric Econ. 2017;12(3):45-52.