Journal of Big Data

RESEARCH

Open Access

A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: empirical evidence from India

R. L. Manogna[1*], Vijay Dharmaji[1] and S. Sarang[1]

*Correspondence:
manognar@goa.bits-pilani.ac.in

[1] Department of Economics and Finance, Birla Institute of Technology and Science, Pilani, K K Birla Goa Campus, Zuarinagar, Sancoale, Goa 403726, India

Abstract

This study presents a comprehensive analysis of agricultural price volatility forecasting using a hybrid long short-term memory (LSTM)-Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Agricultural price volatility poses critical challenges for food security, economic stability, and the livelihoods of millions, particularly in developing countries like India. Accurately forecasting these price fluctuations is vital for effective policymaking and strategic decision-making in agricultural markets. This study investigates the potential of deep learning models, specifically LSTM, and their integration with GARCH for forecasting agricultural commodity price volatility. Using extensive historical price data for 23 commodities across 165 markets in India from February 2010 to June 2024, the proposed hybrid model demonstrates significantly enhanced accuracy and robustness compared to standalone econometric or deep learning models. The results suggest that this hybrid approach effectively addresses price instability, offering improved predictive capabilities. These findings provide valuable implications for policymakers and stakeholders, emphasizing the adoption of advanced machine learning techniques for better market risk management and policy interventions tailored to agricultural price dynamics.

Keywords: Price volatility, Neural networks, Forecasting, Agricultural commodities, Hybrid model, Deep learning

JEL Classification: C45, E17, C53, G17

Introduction

Agricultural price volatility is a persistent and critical issue that has significant implications for global economies, particularly those that rely heavily on agriculture for economic stability and food security. Unlike industrial commodities, where price movements are largely influenced by controlled production cycles and predictable demand patterns, agricultural commodities are subject to a confluence of unpredictable factors. These include natural disasters, climate change, pest infestations, shifts in trade policies, and geopolitical conflicts. The inherently volatile nature of agricultural prices stems from the sector's dependence on biological processes, which are sensitive to environmental variations, and its integration into complex global supply chains. The unpredictability of agricultural prices has far-reaching consequences, influencing the lives of billions, shaping government policies, and determining the trajectory of economic growth in many developing and developed nations [1, 20, 29].

The problem of agricultural price volatility has become increasingly urgent in the twenty-first century due to several interrelated factors. Climate change is arguably one of the most significant drivers of volatility, introducing erratic weather patterns and increasing the frequency of extreme events such as floods, droughts, and heatwaves. These phenomena disrupt planting and harvesting schedules, leading to reduced crop yields and, in some cases, complete crop failures. Rising global temperatures are also causing long-term shifts in agricultural productivity, with some regions becoming less suitable for traditional crops, thereby exacerbating global supply imbalances [23, 25]. For instance, a single drought in a major grain-producing region can create ripple effects across international markets, driving up prices and reducing affordability for consumers worldwide [14].

Another critical driver of volatility is the interconnected nature of modern agricultural markets. The globalization of trade has linked the fates of producers and consumers in disparate regions, exposing markets to external shocks. A trade policy shift in one country, such as the imposition of export bans or the lifting of tariffs, can have cascading effects on global prices. Similarly, the use of agricultural commodities for non-food purposes, such as biofuels, has introduced a new layer of complexity to price dynamics. Policies that promote biofuel production often divert significant portions of crops like maize and sugarcane from food markets, driving up prices and exacerbating volatility during periods of low supply [7, 24]. Speculative trading in agricultural futures markets adds another dimension, as it amplifies price swings by decoupling prices from fundamental supply–demand dynamics. While futures markets are designed to mitigate risks and provide price stability, excessive speculation has, in many cases, increased the magnitude and frequency of price fluctuations [5].

However, traditional econometric models, such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, struggle to capture the non-linear dependencies and intricate patterns in agricultural market data, leading to inaccurate forecasting. This lack of precision makes it difficult for farmers, policymakers, and stakeholders to anticipate price fluctuations, exacerbating financial instability, food insecurity, and economic disruptions. Price volatility directly impacts food security, as fluctuating prices can lead to food shortages, increased hunger, and malnutrition, particularly in developing countries where a significant portion of the population relies on agriculture for sustenance [22]. It also disrupts economic stability, discouraging investments, destabilizing supply chains, and reducing market confidence. Small-scale farmers are among the most vulnerable, as unstable income prevents them from making informed investment decisions, often pushing them into cycles of debt and financial distress. Moreover, policymakers struggle to implement timely and effective price stabilization measures due to the limitations of existing forecasting models, leading to delayed interventions that exacerbate inflation and supply imbalances. Climate change-induced disruptions, such as droughts, floods, and extreme weather patterns, further amplify the unpredictability of agricultural prices, making traditional forecasting methods increasingly ineffective. Additionally, speculative trading in agricultural markets can amplify price swings, detaching prices from actual supply–demand dynamics and creating instability. Given these factors, delayed action in improving volatility forecasting can result in severe economic and humanitarian crises. Without accurate price forecasts, governments, policymakers, and market participants cannot effectively manage risks, protect farmers, stabilize markets, or ensure food security [21]. Therefore, addressing this pressing issue through a hybrid deep learning-based approach, such as integrating Long Short-Term Memory (LSTM) networks with GARCH, is essential for improving prediction accuracy and enabling proactive risk management in agricultural markets.

The socioeconomic implications of agricultural price volatility are profound and multifaceted. For producers, particularly smallholder farmers in developing countries, volatile prices represent a significant source of financial instability. These farmers often lack access to formal credit systems, crop insurance, and advanced market information, leaving them highly vulnerable to adverse price movements [26]. A sudden drop in prices can result in income losses, forcing farmers to reduce investments in essential inputs such as seeds, fertilizers, and machinery. Over time, this discourages productivity-enhancing practices, perpetuating cycles of low yields and poverty. Conversely, sudden price increases can benefit producers in the short term but often lead to overproduction, which drives down prices in subsequent cycles and destabilizes markets further [9, 15].

Consumers, especially in low-income regions, bear the brunt of price surges. In many developing countries, food expenditures constitute a significant share of household budgets, often exceeding 50%. When staple food prices rise sharply, households are forced to make difficult trade-offs, reducing food consumption or sacrificing expenditure on education, healthcare, and other necessities [13, 27]. The consequences of such adjustments are particularly severe for vulnerable populations, including children, pregnant women, and the elderly, who face long-term health risks due to malnutrition and inadequate diets. Rising prices also contribute to political instability, as food riots and protests often emerge in response to unaffordable staple goods. Historical examples include the 2007–2008 global food price crisis, which triggered widespread unrest in several countries, particularly in Africa and the Middle East [28].

The urgency of addressing agricultural price volatility has spurred considerable interest in developing effective forecasting tools. Accurate volatility forecasts can enable governments, producers, and private sector stakeholders to make informed decisions that mitigate risks and enhance market stability. For policymakers, reliable forecasts can guide interventions such as releasing strategic food reserves, adjusting subsidies, or implementing trade policies to stabilize prices [13]. For producers, particularly small-scale farmers, accurate predictions of price trends can inform planting, harvesting, and marketing strategies, reducing financial uncertainty and enhancing resilience. For private-sector stakeholders, including commodity traders, food processors, and retailers, improved forecasting capabilities enable better inventory management, risk hedging, and cost optimization.

Traditional econometric models have long been the foundation of volatility forecasting. Autoregressive Conditional Heteroskedasticity (ARCH) models, introduced by Engle [3], and their extension, Generalized ARCH (GARCH), developed by Bollerslev [2], have been widely used to analyze time-varying volatility. These models are particularly effective in capturing volatility clustering, a phenomenon where periods of high volatility are followed by more high volatility, and periods of low volatility are followed by more low volatility. Extensions such as the GJR-GARCH model, which accounts for asymmetric responses to positive and negative shocks, and the Exponentially Weighted Moving Average (EWMA) model, which dynamically adjusts weights for recent observations, have further refined volatility modeling techniques [6, 13]. While these methods have provided valuable insights into price dynamics, they are limited by their reliance on linear assumptions and predefined structures, which constrain their ability to capture the complex, nonlinear, and stochastic behavior of agricultural price movements [30].

The limitations of traditional models have driven the adoption of machine learning (ML) and deep learning (DL) approaches, which offer greater flexibility and predictive power. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, have gained particular attention for their ability to model sequential data and capture long-term dependencies. LSTMs have demonstrated strong performance in financial markets, where they excel at identifying patterns in time-series data and predicting price trends [4, 8]. In agricultural markets, LSTMs have shown promise in capturing the seasonal and cyclical patterns inherent in commodity prices, as well as in incorporating external factors such as weather conditions, policy changes, and global economic trends.

Despite their advantages, standalone LSTM models face challenges in forecasting volatility. Their "black-box" nature raises interpretability concerns, making it difficult for stakeholders to understand and trust their predictions. Additionally, LSTMs may struggle to account for volatility clustering—a critical feature of agricultural price data. To address these shortcomings, researchers have increasingly turned to hybrid models that combine the strengths of econometric and machine learning techniques.

The hybrid LSTM-GARCH model represents a significant advancement in volatility forecasting. By integrating the interpretability and volatility modeling strengths of GARCH with the pattern recognition and sequence learning capabilities of LSTM, this approach provides a comprehensive framework for analyzing agricultural price dynamics. The LSTM component captures long-term dependencies and complex nonlinear relationships, while the GARCH component focuses on modeling short-term volatility clustering in the residuals. This dual approach enhances predictive accuracy and addresses the limitations of standalone models, making the hybrid LSTM-GARCH model particularly well-suited for agricultural markets characterized by irregular and extreme price movements.

Empirical studies have demonstrated the superiority of the LSTM-GARCH model across diverse commodities, including cereals, pulses, and oilseeds. For instance, the model has shown exceptional performance in predicting the prices of highly volatile commodities like onion and mustard, which are prone to sharp fluctuations due to climatic variations, supply chain disruptions, and policy interventions. By achieving lower prediction errors and better capturing the complexities of agricultural price movements, the LSTM-GARCH model has emerged as a powerful tool for stakeholders seeking to manage risks and optimize decision-making.

This paper aims to contribute to the growing body of literature on agricultural price volatility by evaluating the potential of hybrid models like LSTM-GARCH. Through a detailed analysis of traditional econometric models, advanced machine learning techniques, and their integration, this study seeks to advance the understanding of agricultural price dynamics and provide actionable insights for policymakers, producers, and private-sector stakeholders. The findings underscore the transformative potential of hybrid approaches in creating more resilient and sustainable agricultural markets, ultimately contributing to global food security and economic stability.

Literature review

Agricultural price volatility has long been a critical issue globally, particularly in developing economies where agricultural income constitutes a significant portion of household earnings. Volatility leads to unpredictable market conditions, impacting both producers and consumers. For producers, fluctuating prices create uncertainty in income, limiting their ability to plan investments and production cycles [15]. Consumers, particularly in low-income regions, experience heightened food insecurity as price spikes force adjustments in consumption patterns [28]. Traditional econometric approaches such as Autoregressive Conditional Heteroskedasticity (ARCH) and its extension Generalized ARCH (GARCH) have been extensively used in volatility modeling [2, 3]. These models are designed to capture time-varying volatility and have been widely applied in financial and commodity markets [17]. Extensions like the GJR-GARCH [6] model account for asymmetric effects, while Exponentially Weighted Moving Average (EWMA) models provide a simplistic framework for updating volatility estimates dynamically [18]. Although these models have been effective in capturing volatility clustering, they often fail to address non-linear dynamics and external shocks, which are common in agricultural markets. Studies indicate that traditional GARCH models perform well under stable conditions but struggle with abrupt changes caused by external factors such as weather anomalies and policy interventions [30]. Additionally, the reliance on fixed lag structures limits their adaptability to complex agricultural data patterns [19].

Recent advancements in machine learning, particularly deep learning, have paved the way for more robust volatility forecasting techniques. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, have been recognized for their ability to model sequential dependencies in time-series data [8]. Studies have demonstrated the success of LSTM in predicting financial market trends and stock price volatility [4]. In agricultural markets, LSTM models have shown potential in capturing seasonal trends and the effects of external shocks, such as weather anomalies [10]. While LSTMs offer significant improvements over traditional models, they are not without limitations. The "black-box" nature of deep learning makes the results less interpretable compared to traditional econometric models. Furthermore, LSTMs often require large datasets and extensive computational resources, making them less accessible for certain applications [11]. Hybrid models that integrate traditional econometric models with deep learning techniques have emerged as a promising solution to address the limitations of standalone methods [12]. The LSTM-GARCH model combines the interpretability and volatility modeling strength of GARCH with LSTM's ability to capture non-linear patterns and long-term dependencies [30]. By leveraging the strengths of both approaches, hybrid models aim to achieve higher accuracy and robustness in volatility forecasting. Empirical evidence supports the effectiveness of hybrid models in financial markets. For instance, studies by Paul and Yeasin [16] and Kim and Won [10] demonstrate that hybrid models outperform standalone econometric and deep learning models in predicting price trends. However, their application in agricultural markets remains limited, and the generalizability of these findings across diverse commodities and markets has not been fully explored.

The application of hybrid models, particularly those integrating econometric and deep learning techniques, in agricultural markets remains underexplored. While financial markets have benefited from the advancements of hybrid approaches like LSTM-GARCH, agricultural markets present unique challenges, including seasonality, supply chain disruptions, and policy interventions, which require specialized adaptations. Existing studies predominantly focus on standalone econometric or deep learning models, leaving a significant gap in understanding how hybrid models perform under the complex and irregular conditions characteristic of agricultural price data. Another critical gap lies in the lack of comparative analyses between traditional econometric models, deep learning methods, and hybrid approaches across a diverse range of commodities and market conditions. Most research limits its scope to specific commodities or geographic regions, which undermines the generalizability of the findings. This narrow focus fails to capture the heterogeneity of agricultural markets, where different commodities exhibit varying volatility patterns influenced by factors like seasonality, perishability, and global trade dynamics. Furthermore, the integration of econometric and machine learning techniques presents practical challenges, such as parameter optimization and ensuring model interpretability without compromising accuracy. Hybrid models like LSTM-GARCH are theoretically promising but have yet to be rigorously evaluated for their scalability and real-world applicability in agricultural markets. Addressing these integration challenges is crucial for broader adoption and effective implementation. A summary of the literature review has been provided in Table 1.

Table 1 Summary of literature review

Study/approach	Findings	Gaps in literature
ARCH and GARCH models [2, 3]	Effectively captures time-varying volatility but assumes linear relationships	Fails to capture non-linear dependencies and complex patterns in agricultural markets
GJR-GARCH model [6]	Accounts for asymmetric effects but still relies on linear assumptions	Still struggles with capturing non-linearity and external market shocks
EWMA model [18]	Provides a simplistic framework for volatility estimation but lacks adaptability to external shocks	Limited in handling non-linear agricultural price dynamics and supply chain disruptions
Traditional GARCH models in agricultural markets [30]	Performs well under stable conditions but struggles with abrupt external shocks	Relies on fixed lag structures, limiting adaptability to agricultural data
LSTM for financial markets [4]	LSTM models successfully predict stock price volatility by capturing long-term dependencies	Black-box nature makes results less interpretable for decision-makers
LSTM for agricultural price forecasting [10]	LSTM models capture seasonal trends and external shocks but face interpretability challenges	Requires large datasets and high computational resources, making accessibility difficult
Hybrid LSTM-GARCH models [16], Kim and Won [10]	Hybrid models outperform standalone methods in financial markets, demonstrating improved accuracy	Limited research on applying hybrid models to agricultural markets and ensuring scalability

Methodology

Data description

The daily wholesale data from 2010–02–26 to 2024–06–11 for 23 different commodities across 165 markets have been collected from the AGMARKNET Portal (http://agmarknet.gov.in/). These 23 commodities include 3 vegetables (Potato, Onion and Tomato), 4 oilseeds (Mustard, Sesamum, Groundnut and Soybean), 5 pulses (Greengram, Kabuli Channa, Lentil, Bengal Gram and Arhar), 4 spices (Turmeric, Dry Chillies, Cumin and Coriander) and 6 cereals (Paddy, Jowar, Maize, Wheat, Ragi and Bahra) and Cotton which is a fiber crop. Different classes of commodity have been selected for their different levels of volatility. The selection of the markets and commodities in each class was based on their maximum share and representation. The analysis of the data is carried out in Python software with 70% for training, 10% for validation and 20% for testing the data.

The commodity with the highest variation, as measured by the coefficient of variation (CV), is Onion, with a CV of approximately 0.765. The descriptive statistics can be seen in Table 2.

Forecasting models

Generalized Autoregressive Conditional Heteroscedastic (GARCH) Model

The Generalized Autoregressive Conditional Heteroscedastic (GARCH) model, introduced by Bollerslev in 1986, represents a significant evolution in econometric modeling for analyzing and forecasting time-varying volatility in financial and economic time series. Building upon the foundation laid by Engle's Autoregressive Conditional Heteroscedasticity (ARCH) model in 1982, the GARCH model addresses several limitations of the ARCH framework by introducing a more parsimonious structure that accommodates long-term dependencies in volatility.

The GARCH model is designed to model the conditional variance of a time series, which refers to the variance of the current observation given past observations. This approach is particularly effective in capturing volatility clustering, a phenomenon frequently observed in financial and commodity markets, where large price changes are followed by more large changes (of either sign) and small changes are followed by small changes. The model's ability to describe time-varying volatility has made it a cornerstone of volatility forecasting across diverse fields, including finance, economics, and agriculture.

In its most basic form, the GARCH(p,q) model is defined as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \cdot \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \cdot \sigma_{t-j}^2 \tag{1}$$

σ_t^2 is the conditional variance at time t; ω is the constant term; α_i are the ARCH parameters; β_j are the GARCH parameters; ϵ_t is the error term at time t.

The GARCH model assumes that the conditional variance is a linear function of past squared errors and past conditional variances. This formulation offers a more compact representation compared to the ARCH(q) model, which requires a large number of lagged squared errors to achieve comparable accuracy. The GARCH model remains a

Table 2 Descriptive statistics of wholesale prices of agricultural commodities

Commodity	Mean	Median	Maximum	Minimum	Std Dev.	CV
Spices						
Turmeric	11,421	11,000	1,10,000	4250	3850	0.34
Coriander	5718	5540	12,005	1975	1990	0.35
Dry chillies	6151	5750	15,000	1300	2288	0.37
Cumin	14,429	13,000	61,500	5240	5607	0.39
Oilseeds						
Groundnut	4536	4255	8955	900	1283	0.28
Soybean	4160	4300	9200	1900	1121	0.27
Mustard	3852	3500	13,500	1875	1012	0.26
Sesamum	9082	9000	15,500	4000	2744	0.3
Cereals						
Wheat	1735	1830	4000	935	417	0.24
Ragi	1588	1600	3350	680	569	0.36
Bajra	1569	1460	3500	675	442	0.28
Maize	1748	1853	2100	850	311	0.18
Jowar	2160	2000	5200	300	694	0.32
Paddy	1338	1360	2395	692	280	0.21
Pulses						
Arhar	4917	4600	11,000	2105	1643	0.33
Bengal gram	4119	4123	8915	1750	1088	0.26
Lentil	3699	3420	7015	1000	1117	0.3
Greengram	6936	7800	9400	2000	1929	0.28
Kabuli chana	6064	5350	14,000	1200	2140	0.35
Vegetables						
Onion	1635	1300	10,000	320	1250	0.77
Tomato	1509	1300	8100	200	986	0.65
Potato	857	810	3800	240	497	0.58
Others						
Cotton	4909	4650	10,750	1705	1501	0.31

Author's own representation

foundational tool in volatility forecasting, combining theoretical elegance with practical utility. While it has limitations, its adaptability and extensibility make it a valuable framework for understanding and predicting agricultural price dynamics. By capturing volatility clustering and providing a basis for hybrid models, GARCH continues to play a pivotal role in addressing the challenges of time-varying volatility in complex markets.

The GARCH model remains a foundational tool in volatility forecasting, combining theoretical elegance with practical utility. While it has limitations, its adaptability and extensibility make it a valuable framework for understanding and predicting agricultural price dynamics. By capturing volatility clustering and providing a basis for hybrid models, GARCH continues to play a pivotal role in addressing the challenges of time-varying volatility in complex markets.

Glosten, Jagannatan, Runkle-GARCH (GJR-GARCH) model

The Glosten, Jagannathan, Runkle-GARCH (GJR-GARCH) model, introduced by Glosten, Jagannathan, and Runkle in 1993, is a significant extension of the GARCH framework. The model specifically addresses one of the key limitations of standard GARCH models: their inability to account for the asymmetric impacts of positive and negative shocks on volatility. This phenomenon, often referred to as the leverage effect, is particularly evident in financial and commodity markets, where negative shocks (e.g., unexpected losses or price drops) tend to have a more pronounced effect on volatility than positive shocks of the same magnitude.

Definition and formulation The GJR-GARCH(p,q) model modifies the conditional variance equation of the standard GARCH model by introducing an indicator function that differentiates the effects of positive and negative shocks. The model is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i + \gamma_i \cdot I_{t-i}) \cdot \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \cdot \sigma_{t-j}^2 \quad (2)$$

I_t is an indicator function that takes the value of 1 if $\epsilon_t < 0$ and 0 otherwise; γ_i is the asymmetric parameter that captures the leverage effect.

The GJR-GARCH model allows for different impacts of positive and negative shocks on volatility, with negative shocks having a larger impact when $\gamma_i > 0$.

The GJR-GARCH model represents a significant advancement in volatility modeling by introducing asymmetry into the standard GARCH framework. Its ability to capture the leverage effect and differential impacts of shocks makes it particularly suited for markets with pronounced negative shocks, such as agricultural markets. While it has limitations, its integration into hybrid approaches and its adaptability to diverse datasets make it a valuable tool for understanding and forecasting volatility in complex market environments.

Exponentially weighted moving average (EWMA) model

The Exponentially Weighted Moving Average (EWMA) model, often referred to as the RiskMetrics model, is a widely used volatility forecasting technique known for its simplicity and computational efficiency. Developed by J.P. Morgan for its RiskMetrics framework in the mid-1990s, the EWMA model has been extensively applied in financial risk management, portfolio optimization, and market analysis. Its ability to capture time-varying volatility while requiring minimal parameterization makes it a practical tool for forecasting in dynamic and uncertain environments, including agricultural markets.

The EWMA model is defined as:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \quad (3)$$

σ_t^2 is the conditional variance at time t; λ is the decay factor, typically set to 0.94 for daily data and 0.97 for monthly data; r_t is the return at time t.

The EWMA model's central feature is its use of exponentially decaying weights to calculate volatility. This approach ensures that more recent observations are given greater importance, reflecting the intuition that recent market conditions are more relevant for predicting near-term volatility. This weighting mechanism allows the model to adapt quickly to changing market conditions, making it particularly effective during periods of high volatility. The EWMA model remains a widely used tool for volatility forecasting due to its simplicity, adaptability, and intuitive weighting mechanism. Although it has limitations in addressing the complexities of modern markets, its computational efficiency and effectiveness in capturing short-term volatility dynamics make it a valuable component in both standalone and hybrid forecasting frameworks. In agricultural markets, where timely and actionable volatility forecasts are critical for decision-making, the EWMA model offers a practical solution, particularly when combined with more advanced techniques to address its shortcomings.

Multiplicative error model (MEM)

The Multiplicative Error Model (MEM), introduced by Engle and Gallo in 2006, represents a novel approach to modeling non-negative financial variables. Unlike traditional volatility models such as GARCH, which are primarily designed for squared returns or variances, the MEM is tailored to non-negative time series data, including trading volumes, realized volatility, duration between trades, and other financial measures where values are inherently non-negative. By leveraging a multiplicative error process, the MEM provides a flexible framework for capturing the dynamics of these variables, making it particularly valuable for financial and commodity market analysis.

Definition and formulation The MEM is based on the premise that the observed variable x_t can be expressed as the product of a conditional mean component μ_t and a stochastic error component ϵ_t, as follows:

$$x_t = \mu_t \epsilon_t \tag{4}$$

$$\mu_t = \omega + \sum_{i=1}^{q} \alpha_i \cdot x_{t-i} + \sum_{j=1}^{p} \beta_j \cdot \mu_{t-j} \tag{5}$$

x_t is the non-negative variable of interest; μ_t is the conditional mean; ϵ_t is the error term; ω, α_i, and β_j are the model parameters.

While this discussion emphasizes agricultural markets, the MEM has broader applications in financial markets, energy markets, and macroeconomic forecasting. Its versatility in modeling non-negative variables makes it a valuable tool across disciplines. The Multiplicative Error Model (MEM) provides a powerful framework for modeling and forecasting non-negative variables in time series data. Its unique formulation and flexibility make it particularly valuable for applications involving trading volumes, realized volatility, and other non-negative measures. Although it has limitations, its adaptability and ability to handle dynamic data make it an important tool in agricultural markets and beyond. By integrating MEM with other econometric or machine learning models, researchers and practitioners can further enhance its utility and predictive power in complex market environments.

Hybrid LSTM-GARCH model

The hybrid LSTM-GARCH model architecture is designed to leverage the strengths of both LSTM (Long Short-Term Memory) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) to predict volatility in financial time series data, such as stock returns. Each of these components addresses a specific characteristic of financial data—LSTM captures temporal patterns, while GARCH models volatility clustering. This combination allows for a more comprehensive understanding of time series trends and their associated risks.

The LSTM component is the first part of the model. LSTM networks are a type of recurrent neural network (RNN) that are particularly well-suited for handling sequences of data, like stock returns, due to their ability to retain information over long time periods. The architecture starts with an input layer that receives historical data in the form of sequences. These sequences represent windows of past observations, such as stock returns over a certain number of days. The input is passed through one or more LSTM layers, where each LSTM unit contains memory cells that maintain a temporal understanding of the data. In a typical stacked LSTM architecture, the first LSTM layer outputs a sequence of hidden states, which are further processed by subsequent LSTM layers. The final hidden state, which encodes information from the entire sequence, is passed to a fully connected (dense) output layer that predicts the next step's return.

Once the LSTM predicts the next time step's return, the residuals (the differences between the actual returns and the LSTM predictions) are calculated. These residuals often exhibit volatility clustering—a phenomenon where large swings in returns are followed by more large swings, and small swings follow small swings. The LSTM, while excellent at capturing temporal trends, may not fully account for these clusters of volatility. This is where the GARCH model comes in. The GARCH component is applied to the residuals from the LSTM model. GARCH models are designed to handle heteroscedasticity (time-varying variance) by modeling the volatility of the residuals as a function of both past residuals and past volatility. A GARCH(1,1) model, which includes one lag for both residuals and variance, is commonly used in financial applications due to its simplicity and effectiveness.

The GARCH model architecture is relatively straightforward. It uses the residuals from the LSTM model to estimate future volatility based on past errors and volatility patterns. This model fits the residuals and predicts the next period's variance (volatility), using parameters learned during the model fitting process. The key parameters in a GARCH(1,1) model are (ω) (the constant), (α) (the weight of past squared residuals), and (β) (the weight of past variance). These parameters allow the GARCH model to capture how volatility evolves over time.

Finally, the LSTM and GARCH outputs are combined to form the final volatility prediction. The LSTM provides a trend-based prediction for the next time step's return, while the GARCH model adjusts this prediction based on the predicted volatility of the residuals. This combination allows the model to account for both long-term patterns in the data (via LSTM) and short-term volatility dynamics (via GARCH). Depending on the application, the final prediction might be calculated by adding the LSTM's return prediction to the GARCH's volatility adjustment, providing a comprehensive forecast that incorporates both trend and risk components.

In summary, the hybrid LSTM-GARCH model is a powerful tool for predicting volatility in financial time series. By combining the LSTM's ability to model sequential dependencies with GARCH's expertise in handling volatility clustering, the model can provide more accurate predictions of future market behavior. The architecture's flexibility allows it to be adapted to a wide range of financial assets and timeframes, making it a valuable approach for tasks such as risk management, portfolio optimization, and volatility forecasting.

Prediction accuracy

The prediction accuracy of different models compared based on four error measures. Namely, Root Mean Squared Error (RMSE) Eq. (6), Relative Normalised Mean Squared Error (RNMSE) Eq. (7), Mean Absolute Error (MAE) Eq. (8) and Mean Absolute Percentage Error (MAPE) Eq. (9). The formulas for each are given below.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

$$\mathrm{RNMSE}_{\mathrm{range}} = \frac{\mathrm{RMSE}}{\max(y) - \min(y)} \tag{7}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{9}$$

Results and discussions

The data provided below in Tables 3, 4, 5 and 6 represents RMSE, RNMSE, MAE and Mean Absolute Percentage Error (MAPE) values for five different models (GARCH, GJR-GARCH, EWMA, MEM, and LSTM-GARCH) across various commodities. The plots for the same analysis can be seen in Figs. 1, 2, 3 and 4 respectively. These measures are widely used accuracy metrics, measures the accuracy of models in predicting volatility by showing the percentage deviation of the predicted values from the actual values. A lower value indicates better predictive performance, and the results clearly establish the LSTM-GARCH model as the best-performing approach across all commodities.

The GARCH model (Generalized Autoregressive Conditional Heteroskedasticity) and its extension, GJR-GARCH (which accounts for asymmetric effects), demonstrate limited accuracy in capturing the nonlinear and complex volatility patterns prevalent in agricultural markets. While GJR-GARCH offers slight improvements over GARCH in certain cases, the overall performance of these models is constrained by their reliance on linear assumptions and limited flexibility. The Exponentially Weighted Moving Average (EWMA) model generally performs poorly across most commodities, reflecting its simplicity and inability to handle the complex and nonlinear dynamics characteristic of agricultural price volatility. While it effectively gives more weight to recent data, this approach struggles to adapt to irregular or extreme price movements. The MEM (Multiplicative Error Model) shows moderate performance and occasionally outperforms GARCH and EWMA. However, it still falls short of the predictive accuracy achieved by LSTM-GARCH. While MEM demonstrates some

Table 3 Prediction performance based on RMSE

Commodity	GARCH	GJR-GARCH	EWMA	MEM	LSTM-GARCH
Turmeric	1.51	1.12	1.08	0.96	0.9
Mustard	4.23	1.18	1.11	0.96	0.91
Paddy	1.73	0.87	1.87	0.83	0.81
Greengram	1.64	1.2	1.1	0.96	0.95
Cotton	6.23	0.88	1.89	0.81	0.79
Dry chillies	3.11	0.36	1.36	0.36	0.38
Kabuli chana	2.35	1.06	1.07	0.87	0.84
Sesamum	2.65	0.95	2.95	0.93	0.91
Cumin	2.28	0.98	1.96	0.88	0.84
Soybean	2.96	0.93	2.93	0.9	0.87
Potato	9.43	0.97	2.98	0.98	0.95
Groundnut	3.02	1.01	1.02	0.9	0.88
Jowar	3.79	0.89	1.88	0.82	0.78
Lentil	5.46	0.94	1.93	0.86	0.83
Maize	6.41	1.14	1.04	0.91	0.93
Arhar	9.63	1.16	1.15	0.89	0.85
Onion	5.44	1	1.94	0.83	0.7
Tomato	7.67	1.16	1.11	0.91	0.79
Bengalgram	4.52	0.94	1.94	0.92	0.89
Wheat	6.48	1.01	1.02	0.98	0.97
Ragi	3.93	0.89	1.89	0.89	0.87
Coriander	5.71	0.91	2.91	0.9	0.88
Bajra	2.86	0.92	1.92	0.82	0.77

adaptability, its limitations in modeling nonlinear patterns prevent it from addressing the full complexity of agricultural price movements.

The LSTM-GARCH model consistently outperforms all other models across the dataset, solidifying its position as the most effective approach for forecasting agricultural price volatility. By integrating the strengths of both LSTM networks and GARCH models, the hybrid approach captures both long-term temporal dependencies and short-term volatility clustering with exceptional accuracy. This makes it uniquely suited to handle the complex and dynamic nature of agricultural price data. The superior performance of LSTM-GARCH underscores its ability to adapt to diverse commodities and varying volatility patterns, where traditional models struggle. Its hybrid structure enables it to address the limitations of standalone econometric or deep learning models, offering a robust solution for agricultural price forecasting.

A closer examination of commodity-level results further highlights the robustness of LSTM-GARCH. Commodities with high levels of price volatility, which are challenging for traditional models to predict accurately, are effectively handled by LSTM-GARCH. Its adaptability to irregular or extreme patterns sets it apart from other models, which often fail to account for such complexities. The consistent superiority of LSTM-GARCH across diverse commodities emphasizes its reliability and versatility, making it an essential tool for stakeholders seeking to mitigate risks and improve decision-making in agricultural markets.

Table 4 Prediction performance based on RNMSE

Commodity	GARCH	GJR-GARCH	EWMA	MEM	LSTM-GARCH
Turmeric	144.34	140.51	20.13	18.95	2.78
Mustard	15.17	14.98	1.42	1.27	0.18
Paddy	1.68	1.66	0.06	0.06	0.02
Greengram	50.62	50.21	4.57	4.20	0.02
Cotton	23.50	23.23	1.77	1.77	0.39
Dry chillies	19.05	17.07	0.62	0.62	10.73
Kabuli chana	37.04	36.32	5.16	5.10	8.97
Sesamum	85.53	84.29	6.86	6.85	5.07
Cumin	218.76	215.77	27.67	25.25	8.85
Soybean	17.27	17.13	1.17	1.15	0.07
Potato	0.95	0.92	0.23	0.24	0.01
Groundnut	20.57	20.44	1.69	1.71	0.10
Jowar	4.43	4.32	0.38	0.37	0.16
Lentil	13.35	13.18	1.11	1.08	0.31
Maize	3.10	3.08	0.13	0.11	0.00
Arhar	24.53	24.11	3.56	3.55	0.99
Onion	3.29	2.77	1.55	1.39	0.34
Tomato	2.85	2.56	1.17	1.08	0.60
Bengalgram	17.39	17.17	1.06	1.05	0.18
Wheat	3.14	3.12	0.18	0.18	0.00
Ragi	2.55	2.50	0.26	0.26	0.12
Coriander	33.47	33.08	3.20	3.18	0.29
Bajra	2.32	2.28	0.17	0.17	0.05

Conclusion

In conclusion, the comparative analysis of volatility prediction models on agricultural commodity prices underscores the significant advantages of hybrid deep-learning models over traditional time-series approaches. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and its asymmetric variant, the GJR-GARCH model, demonstrated moderate performance, effectively capturing volatility clustering and asymmetric shocks. However, their reliance on linear assumptions limits their ability to model the complex and nonlinear dynamics of agricultural price movements. Similarly, the Exponentially Weighted Moving Average (EWMA) model, while straightforward to implement, struggled to accurately predict commodities with irregular volatility patterns. The Multiplicative Error Model (MEM) showed slightly better performance in some cases but fell short in addressing the intricacies of nonlinear volatility.

In contrast, the hybrid LSTM-GARCH model consistently outperformed all other models, providing the most accurate predictions across diverse agricultural commodities. By integrating Long Short-Term Memory (LSTM) networks with GARCH, this hybrid approach combines LSTM's ability to model long-term dependencies and nonlinear patterns with GARCH's strength in addressing volatility clustering. The superior accuracy of the LSTM-GARCH model highlights its robustness in handling the unique characteristics of agricultural price data, including seasonality, supply chain disruptions, and extreme weather conditions. Its adaptability across both highly volatile and more stable commodities demonstrates its versatility and reliability for agricultural market

Table 5 Prediction performance based on MAE

Commodity	GARCH	GJR-GARCH	EWMA	MEM	LSTM-GARCH
Turmeric	15.51	0.88	0.85	0.81	0.74
Mustard	23.81	0.78	0.75	0.77	0.71
Paddy	12.2	0.66	0.66	0.65	0.61
Greengram	12.43	0.78	0.73	0.85	0.83
Cotton	47.8	0.59	0.59	0.59	0.56
Dry chillies	27.07	0.25	0.25	0.27	0.31
Kabuli chana	27.5	0.76	0.77	0.76	0.71
Sesamum	21.16	0.78	0.78	0.81	0.75
Cumin	16.74	0.51	0.50	0.53	0.5
Soybean	20.47	0.76	0.77	0.78	0.73
Potato	46.69	0.74	0.73	0.76	0.72
Groundnut	24.57	0.76	0.77	0.75	0.73
Jowar	37.04	0.65	0.64	0.65	0.59
Lentil	39.49	0.7	0.7	0.73	0.67
Maize	58.97	0.81	0.77	0.77	0.77
Arhar	86.69	0.8	0.79	0.73	0.68
Onion	50.18	0.63	0.6	0.7	0.57
Tomato	53.64	0.75	0.72	0.7	0.59
Bengal gram	39.26	0.67	0.67	0.68	0.63
Wheat	33.6	0.77	0.77	0.79	0.77
Ragi	38.39	0.74	0.74	0.74	0.71
Coriander	47.24	0.69	0.69	0.7	0.66
Bajra	16.06	0.67	0.66	0.67	0.61

analysis. The hybrid model's performance makes it particularly suitable for applications where precise volatility forecasting is critical. Tasks such as risk management, portfolio optimization, and market stabilization can benefit significantly from the improved accuracy of LSTM-GARCH predictions. Traditional models, constrained by their linear frameworks, fail to match the dynamic adaptability of this hybrid approach, especially in the context of complex agricultural markets.

The findings of this study carry critical implications for policymakers and other stakeholders in agricultural markets. Accurate volatility forecasts provided by the LSTM-GARCH model can play a pivotal role in designing and implementing effective market stabilization strategies. Governments and international organizations can utilize these forecasts to anticipate price spikes or drops and implement timely interventions such as releasing strategic food reserves, adjusting subsidies, or modifying import and export policies to ensure market stability. Small-scale farmers, who are often the most vulnerable to price fluctuations, can benefit significantly from the insights generated by LSTM-GARCH forecasts. With better predictions, farmers can make informed decisions about planting, harvesting, and storage, thereby reducing the financial risks associated with volatile markets. This could help break the cyclical trap of poverty and debt that often results from extreme price volatility, fostering more sustainable agricultural practices and financial stability among producers. Food security policies, particularly in low-income regions where price fluctuations directly impact affordability, can also be enhanced using these forecasts. Policymakers can use

Table 6 Prediction performance based on MAPE

Commodity	GARCH	GJR-GARCH	EWMA	MEM	LSTM-GARCH
Turmeric	1878.36	2030.87	1361.26	1338.13	12.87
Mustard	1597.34	1657.17	1913.26	1782.61	5.79
Paddy	1267.09	1267.57	1252.56	1175.32	7.86
Greengram	6050.44	5888.69	13,192.35	12,881.59	1.38
Cotton	2186.36	2182.29	2109.34	1930.07	7.02
Dry chillies	46.17	45.77	46.06	43.66	49.52
Kabuli chana	3147.69	3139.53	4659.42	4384.64	37.99
Sesamum	3946.95	3947.55	3746.46	3614.45	16.76
Cumin	2301.92	2315.49	3281.49	3125.87	7.56
Soybean	1491.24	1479.74	1421.14	1355.28	4.76
Potato	4073.02	4111.21	4163.99	3948.67	5.08
Groundnut	2538.33	2527.78	2731.01	2654.73	3.85
Jowar	1339.31	1320.26	1612.20	1522.92	12.99
Lentil	2767.37	2768.07	3116.56	2955.92	7.98
Maize	4572.77	5026.65	3440.31	3317.89	2.82
Arhar	1487.12	1493.03	2332.12	2218.80	13.58
Onion	774.02	791.66	1571.25	1318.11	45.51
Tomato	1075.95	1080.79	1502.50	1353.89	39.90
Bengalgram	2482.33	2479.70	2563.87	2421.57	8.77
Wheat	7058.99	7015.67	6750.07	6638.81	1.42
Ragi	1787.70	1779.01	1839.13	1774.46	16.42
Coriander	2489.79	2488.26	2572.36	2434.46	6.23
Bajra	1122.34	1124.10	1193.11	1091.30	8.01

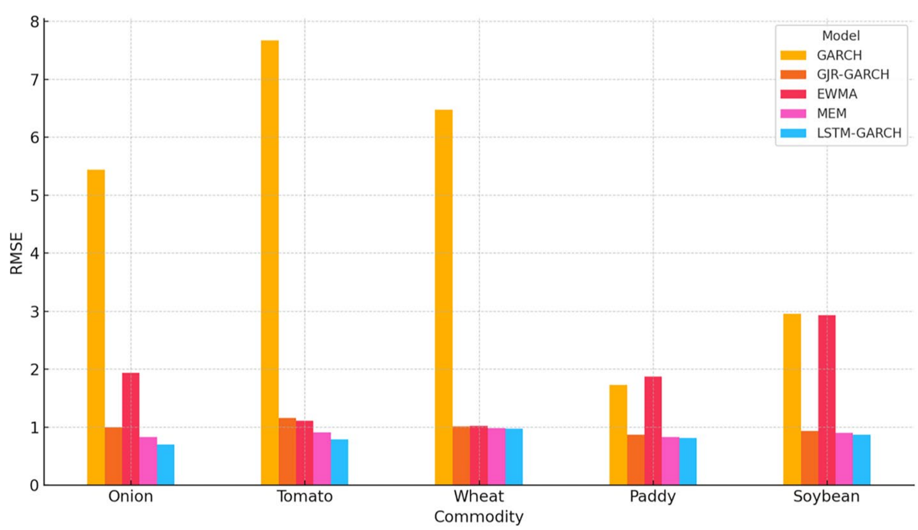

Fig. 1 Comparison of RMSE across models

the improved predictive accuracy of the LSTM-GARCH model to ensure a stable supply of essential commodities and mitigate the adverse effects of sudden price changes on vulnerable populations. By doing so, they can help safeguard access to nutritious food and prevent the socioeconomic disruptions caused by extreme price swings. In

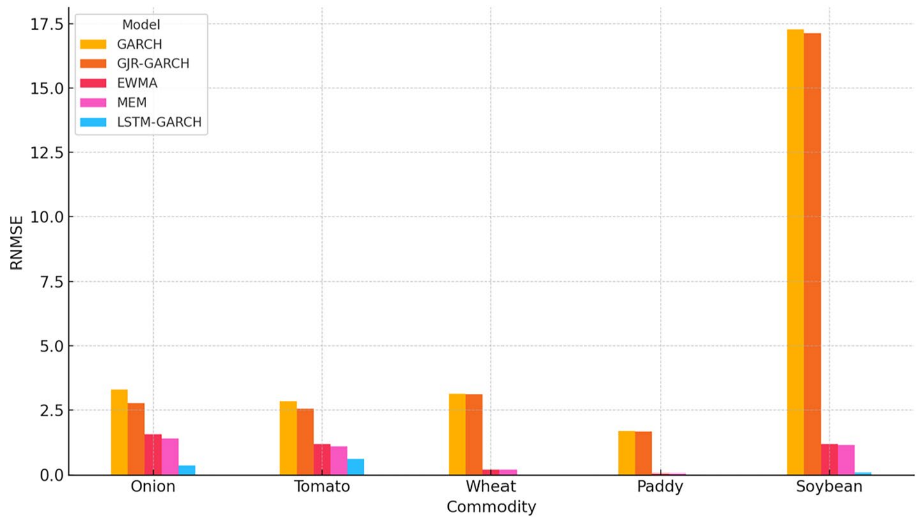

Fig. 2 Comparison of RNMSE across models

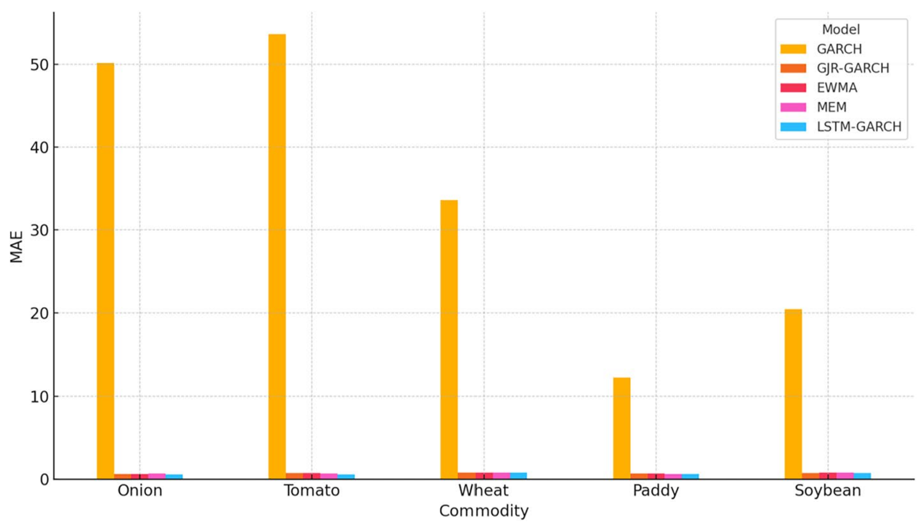

Fig. 3 Comparison of MAE across models

the private sector, businesses involved in agricultural supply chains, such as commodity traders, food processors, and retailers, can leverage the improved forecasts to optimize procurement strategies and manage risks more effectively. Accurate predictions of price trends allow these stakeholders to hedge against volatility, plan inventories better, and reduce costs, ultimately contributing to greater efficiency and stability across the supply chain.

The LSTM-GARCH hybrid model represents a transformative advancement in agricultural price volatility forecasting. By addressing the limitations of traditional econometric and deep-learning models, it provides a robust framework for improving predictive accuracy in markets characterized by high volatility and nonlinear dynamics. Beyond its academic contributions, this model offers practical applications in risk

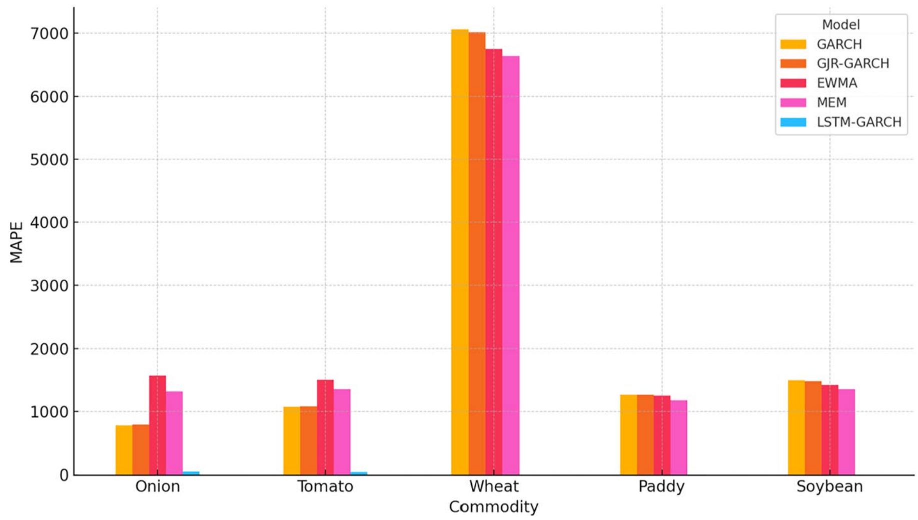

Fig. 4 Comparison of MAPE across models

management, policymaking, and market interventions. Policymakers, producers, and private sector stakeholders can all benefit from its insights, fostering a more resilient and stable agricultural market system. The findings of this study underscore the importance of adopting advanced hybrid models like LSTM-GARCH to mitigate the challenges of price volatility and improve decision-making in the agricultural sector.

Author contributions

The corresponding author, Prof. Manogna RL has contributed towards conceptualization, literature, data, results discussion and drafting the paper. The co-authors Mr. Vijay and Mr. Sarang has contributed towards data extraction, methodology, analysis and conclusion along with editing the draft. All authors have read and approved the manuscript, and ensure that this is the case.

Funding

Open access funding provided by Birla Institute of Technology and Science. No funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 November 2024 Accepted: 11 March 2025
Published online: 11 April 2025

References

1. Abbott PC, Borot de Battisti A. Recent global food price shocks: causes, consequences, and lessons for African governments and donors. J Afr Econ. 2011;20(1):12–62.
2. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. J Econometr. 1986;31(3):307–27.
3. Engle RF. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 1982;50(4):987–1007.
4. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. Eur J Oper Res. 2018;270(2):654–69.
5. Gilbert CL, Morgan CW. Food price volatility. Phil Transact Roy Soc B Biol Sci. 2010;365(1554):3023–34.

6. Glosten LR, Jagannathan R, Runkle DE. On the relation between the expected value and the volatility of the nominal excess return on stocks. J Fin. 1993;48(5):1779–801.
7. Headey D, Fan S. Anatomy of a crisis: the causes and consequences of surging food prices. Agric Econ. 2008;39(suppl):375–91.
8. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
9. Ivanic M, Martin W. Implications of higher global food prices for poverty in low-income countries. Agric Econ. 2008;39(suppl):405–16.
10. Kim J, Song Y, Choi B. LSTM-based agricultural price prediction with seasonality considerations. Agric Syst. 2020;177:102722.
11. Ma Q, Li M, Zhang L. A hybrid forecasting model based on extreme learning machine and GARCH-type models for financial time series. Comput Econ. 2018;52(2):409–31.
12. Manogna RL, Kulkarni N. Does the financialization of agricultural commodities impact food security? An empirical investigation. Borsa Istanbul Rev. 2024. https://doi.org/10.1016/j.bir.2024.01.001.
13. Manogna RL, Kulkarni N, Akshay Krishna D. Nexus between financialization of agricultural products and food security amid financial crisis: empirical insights from BRICS. J Agribus Dev Emerg Econ. 2024. https://doi.org/10.1108/JADEE-06-2023-0147.
14. Manogna RL, Desai D. Nexus of monetary policy and productivity in an emerging economy: supply-side transmission evidence from India. J Quant Econ. 2024. https://doi.org/10.1007/s40953-023-00380-9.
15. Minot N. Food price volatility in sub-Saharan Africa: has it really increased? Food Policy. 2014;45:45–56.
16. Paul S, Yeasin M. GARCH-X models incorporating external shocks for agricultural price prediction. J Time Ser Anal. 2022;43(3):567–85.
17. Pindyck RS. Volatility and commodity price dynamics. J Futur Mark. 2004;24(11):1029–47.
18. RiskMetrics. Technical document. New York: J.P. Morgan; 1996.
19. Roberts MJ, Schlenker W. Identifying supply and demand elasticities of agricultural commodities: implications for the US ethanol mandate. Am Econ Rev. 2013;103(6):2265–95.
20. Manogna MJ, Mishra AK. Price discovery and volatility spillover: an empirical evidence from spot and futures agricultural commodity markets in India. J Agribus Dev Emerg Econ. 2020;10(4):447–73. https://doi.org/10.1108/JADEE-10-2019-0175.
21. Manogna RL, Anand A. A bibliometric analysis on the application of deep learning in finance: status, development, and future directions. Kybernetes. 2023. https://doi.org/10.1108/K-04-2023-0637.
22. Manogna RL, Mishra AK. Forecasting spot prices of agricultural commodities in India: application of deep-learning models. Intell Syst Acc Fin Manag. 2021;28(1):72–83.
23. Manogna RL. Innovation and firm growth in agricultural inputs industry: empirical evidence from India. J Agribus Dev Emerg Econ. 2021;11(5):506–19. https://doi.org/10.1108/JADEE-07-2020-0156.
24. Manogna RL, Mishra AK. Does investment in innovation impact firm performance in emerging economies? An empirical investigation of the Indian food and agricultural manufacturing industry. Int J Innov Sci. 2021;13(2):233–48. https://doi.org/10.1108/IJIS-07-2020-0104.
25. Manogna RL, Mishra AK. Market efficiency and price risk management of agricultural commodity prices in India. J Model Manag. 2023;18(1):190–211. https://doi.org/10.1108/JM2-04-2021-0104.
26. Manogna RL, Mishra AK. Financialization of Indian agricultural commodities: the case of index investments. Int J Soc Econ. 2022;49(1):73–96. https://doi.org/10.1108/IJSE-05-2021-0254.
27. Manogna RL, Mishra AK. Agricultural production efficiency of Indian states: evidence from data envelopment analysis. Int J Financ Econ. 2022;27(4):4244–55.
28. von Braun J, Tadesse G. Global food price volatility and spikes: an overview of costs, causes, and solutions. ZEF discussion papers on development policy, No 161. Bonn: University of Bonn, Center for Development Research (ZEF); 2012.
29. Wright BD. The economics of grain price volatility. Appl Econ Perspect Policy. 2011;33(1):32–58.
30. Zhu X, Zhan Z, Ye M, Zhang C. Improving financial volatility forecasting with bidirectional LSTM neural networks. Expert Syst Appl. 2016;42(10):4544–53.

Publisher's Note

R. L. Manogna (Corresponding Author) is currently Assistant Professor in Economics and Finance at BITS Pilani, K.K. Birla Goa Campus. Her interest includes applications of machine learning in commodity markets and finance, financial econometrics and firm performance. She has published more than 30 research papers in reputed international journals. Her core research is in the area of agricultural commodity markets, food security and financialization.

Vijay Dharmaji , is currently graduate student at BITS Pilani, K.K. Birla Goa Campus. His interest includes financial econometrics and machine learning.

S. Sarang , is currently graduate student at BITS Pilani, K.K. Birla Goa Campus. His interest includes financial econometrics and machine learning.