https://doi.org/10.7251/EMC2603811M
UDK: 338.242.2:[336.77:519.2

Datum prijema rada: 6. februar 2026.
Submission Date: February 6, 2026
Datum prihvatanja rada: 4. septembar 2026.
Acceptance Date: September 4, 2026

Časopis za ekonomiju i tržišne komunikacije
Economy and Market Communication Review

Godina/Vol. **XVI** • Br./No. **III**
str./pp. 811-825

ORIGINALNI NAUČNI RAD / ORIGINAL SCIENTIFIC PAPER

FORECASTING COMMODITY MARKET VOLATILITY: A COMPARATIVE ANALYSIS OF TRADITIONAL AND MACHINE LEARNING MODELS

Ivana Miklošević | Assistant Professor, Department of Economics, University of Slavonski Brod, Slavonski Brod, Croatia, imiklosevic@unisb.hr; ORCID ID: 0000-0002-5882-6409

Abstract: *This study examines how conventional and machine learning models perform in predicting and forecasting volatility in the Indian commodity market and study involved seven key commodities i.e., gold, silver, crude oil and metals. The study applied GARCH, GARCH plus LSTM, and machine learning models like Random Forest, BiLSTM and XGBoost. The study also uses Principal Component Analysis. This study employs daily data for the period spanning from January 2003 to December 2025. Data was sourced from the Multi Commodity Exchange (MCX).*
While traditional GARCH models are helpful in capturing volatility persistence, they are insufficient for accurately forecasting volatile commodity markets.
It was concluded that Machine learning models performed better, especially hybrid ones that mix GARCH and LSTM. Findings matter for investors, policymakers and analysts who need better ways to understand risk in fast-changing markets. It supports the use of data driven tools to manage market risk in a volatile environment, specifically focusing on Indian commodity market.

Keywords: *Volatility Forecasting, Commodity Markets, Machine Learning, GARCH, Financial Risk*

JEL Classification: *C53, G13*

INTRODUCTION

In recent years, commodity markets have gained significant attention from academicians, economic agents and policymakers, specifically in the context of economic volatility (Pindyck, 2004); (Ezeaku, 2020). Unlike equity or bond markets, commodity market is highly sensitive and reactive to global shocks, war, trade disruptions and geopolitical tensions and macroeconomic volatility around the world (Kilian, 2022). Over the past decade, events like the COVID-19 pandemic, Russia Ukraine war and Israel-Hamas have exposed the commodity markets to shocks triggering fluctuations and volatility in prices of gold, silver, crude oil and other metals. All this has led to a challenge in predicting commodity prices with accuracy. All this has led to uninformed decisions by financial institutions, risky investments by investors and business houses.

Further, in emerging markets like India commodity market plays a crucial role in influencing and shaping many macroeconomic variables, decision making by policy makers, fiscal planning and trade. India being an emerging economy is heavily reliant on import of metals , energy, oil and remains highly vulnerable to external shocks and volatility in global commodity prices (Dutta, 2021); (Agnihotri, 2022). India's commodity futures market driven by the Multi Commodity Exchange (MCX) plays a key role in discovering price and risk management for important raw materials that affect regional production systems and supply chains. Commodity market volatility in India goes far beyond financial markets as it directly influences input costs, working capital needs and revenue stability for regionally concentrated producer networks, particularly energy oriented industries and metal-intensive manufacturing clusters. MCX functions within a distinct regulatory and microstructure environment driven by SEBI oversight, liquidity concentration in near-month contracts and presence of both hedgers and speculators.

These market-focused features make India a key regional setting for examining whether traditional and machine learning-based volatility approaches remained effective under emerging-market conditions.

Volatility in commodity markets is generally modelled by employing traditional methods like ARCH and GARCH models. These models are used to examine volatility clustering and heteroskedasticity in commodity market prices. But these models come with the limitation of linearity and stationarity due to which these models struggle to give accurate results under complex conditions (Yilmaz, 2012); (Liu S. Z., 2024). In contrast to this, previous literature has witnessed adoption of machine learning models like Random Forest model, Long Short-Term Memory (LSTM) networks and XGBoost algorithms to increase the predictive accuracy and precision as these models are free from rigid assumptions.

This study will try to address these critical gap by performing a comparative analysis of traditional and machine learning-based forecasting models for major commodities traded in India. The current study also seeks to convey that there is absence of long-term integrated model that provides a framework for predicting commodity market volatility. Lastly, the paper deals with an emerging economy like India with increasing exposure to external commodity volatility thus providing implications at the regional level which is missing from existing studies.

LITERATURE REVIEW

In previous studies, dealing with volatility in commodity markets has been the key focus, specifically due to global crises, uncertainty and the barriers attached with traditional forecasting methods. Academicians have shifted from traditional approaches of GARCH to machine learning models and hybrid frameworks as they tend to improve forecast accuracy and policy insights.

The origin of volatility modelling is seen in work of (Bollerslev, 1986), who introduced the GARCH model that was capable of capturing volatility clustering in financial returns. Since then, various EGARCH, TGARCH and GJR-GARCH models have evolved. However, these models are not capable of measuring non-linear patterns and structural breaks. Therefore the transition to machine learning models has been observed.

(Lili, Xinya, Yanjiao, & Houjian, 2023) made use of LSTM networks on crude oil futures and stated that they outperformed ARIMA and GARCH approaches while assessing phases of high market fluctuation. Their model was capable of capturing long-memory behaviour and performed better in out-of-sample forecasting.

Similarly, (Kristjanpoller & Minutolo, 2016) developed a hybrid artificial neural network–GARCH model to forecast oil price volatility. Their results demonstrated that the hybrid framework significantly improved forecasting performance compared with traditional GARCH models.

Similarly, (Junting, Haifei, Wei, & Xiaojing, 2021) applied XGBoost with wavelet decomposition to make forecasting in gold price. Their hybrid approach performed better than SVM and ANN models in terms of RMSE and MAE. They stated that hybrid models are more suitable for noisy data conditions than single-method approaches.

(Guan & Gong, 2023) proposed a hybrid deep learning model that integrated variational mode decomposition (VMD) with bidirectional long short-term memory (BiLSTM) networks to predict monthly crude oil prices.

(Liu, Zhang, Wang, & Feng, 2024) developed a hybrid carbon price forecasting model that integrates GARCH and LSTM components to capture both conditional heteroskedasticity and long-term nonlinear dynamics in China's carbon trading markets. The GARCH–LSTM framework significantly improved predictive accuracy compared with standalone approaches.

The growing utilization of ML techniques for bolstering the precision of volatility projections substantiates the ascendancy of ML methodologies over conventional econometric models (Goodell, Kumar, Lim, & Pattnaik, 2021).

(Reboredo & Ugolini, 2016) investigated the dependence structure between oil prices and stock market returns and found that the relationship varies across different market conditions, highlighting the complex dynamics present in commodity and financial markets. (Jakub Michańków, 2023). The study proposed a hybrid volatility forecasting model combining GARCH specifications with Gated Recurrent Unit (GRU) neural networks. The model was tested on financial assets including the S&P 500 index, gold and Bitcoin. Results showed that hybrid GARCH–GRU models improved volatility forecasting accuracy compared with traditional approaches. (Palic, 2025) presented a multivariate GARCH (MGARCH/DCC) model of volatility transmission for European Union countries. The transmission mechanism was observed across bond, equity and FX markets. These studies served as motivation for including machine learning models in this study.

(Costa, et al., 2021) investigated the use of machine learning techniques such as Random Forest, XGBoost and regularization models for oil price forecasting. Using a large macroeconomic dataset, the study demonstrated that machine learning methods can significantly improve forecasting accuracy compared with traditional econometric models. (Kumar, Rao, & Dhochak, 2025) examined the use of hybrid machine learning models for volatility prediction in financial risk management. The study integrated traditional econometric models with machine learning techniques to improve forecasting performance. Their findings indicated that hybrid ML frameworks capture nonlinear market dynamics more effectively and provide better volatility prediction than standalone models.

(Araya, Aduda, & Berhane, 2024) examined volatility prediction in financial markets using a hybrid modelling framework. The study integrated ARIMA, GARCH family models, CNN and BiLSTM networks to improve stock price volatility forecasting. The results showed that the proposed hybrid model significantly reduced forecasting errors (MAE and RMSE) compared to conventional models, demonstrating higher accuracy in capturing nonlinear market dynamics.

(Zeda Xu, 2024) The study proposed a hybrid **GARCH-Informed Neural Network (GINN)** model for improving stock market volatility forecasting. The model combines the strengths of traditional GARCH models with the nonlinear learning capability of LSTM deep neural networks. The results indicated that the hybrid framework captures complex volatility dynamics more effectively than standalone models. Empirical findings showed superior out-of-sample forecasting performance in terms of R^2, MSE, and MAE. (Kocaman & Hazar, 2025) recommended use of machine learning approaches for better forecasting and improving predictions. (Barisic, 2025) also employed the use of Random Forest model as a machine learning approach for better results.

Despite using both econometric and machine learning models significant progress has been made in forecasting commodity price volatility there still remains several gaps. Most previous studies focused either on global markets or a single commodity with limited exploration of cross-commodity behaviour specifically in the Indian context. Moreover, while hybrid models like GARCH-LSTM have shown promising results there is no consensus on the most efficient admixture or combination. Many models have ignored the role of sudden external shocks such as geopolitical events, war or policy reforms which are highly relevant in volatile emerging markets. Furthermore, very few past literature has made an attempt to compare traditional and AI-based models on the same dataset across different commodity categories (like metals, energy and agriculture). These gaps in past studies highlight the need for a comprehensive comparative analysis. This study has selected commodity market data of India and tried to compare the forecasting performance of traditional and ML-driven models in a real-world volatile environment.

Research Objectives and Questions

This study aims to explore the comparative effectiveness of traditional econometric models and modern machine learning techniques in forecasting commodity price volatility for selected commodities. This study tries to address the constraints of fragmented forecasting approaches by using a uniform evaluation framework across different commodities and modelling techniques.

The focus of this study is to examine which model performs better under situations of volatility particularly during periods which are influenced by global shocks and structural changes. This study also investigates if hybrid approaches such as GARCH–LSTM could offer improvements over individual models in terms of forecast accuracy and in capturing volatility dynamics. In addition, this research tries to identify the underlying factors that drive variations in forecasting performance across commodities by employing a dimensionality-reduction approach.

Based on these objectives and the identified research gaps, the study addresses the following key questions:

1. Which forecasting models-traditional econometric approaches (such as GARCH) or machine learning techniques (such as LSTM and Random Forest) provide better estimates in predicting commodity price volatility?
2. Do hybrid models such as GARCH–LSTM significantly outperform their standalone counterparts across different commodity including metals, energy and agriculture?
3. What are the prominent drivers of variability influencing forecasting performance across commodities and how do such factors like price volatility, external shocks and structural patterns contribute to these differences?

METHODOLOGY

Data Description

This study employed daily price data spanning from January 2003 to December 2025. Daily data frequency was employed to capture short-term price fluctuations and volatility clustering.

This study aims to analyse volatility dynamics in India's commodity markets. The examination considered seven major commodities which are traded on the Multi Commodity Exchange (MCX) i.e. gold, silver, crude oil, zinc, copper, lead and aluminium. These commodities represent different segments of the Indian commodity market, for instance precious metals, energy and base metals. These are also the commodities which are widely used by investors, hedgers and policy institutions.

The rationale behind selecting these commodities was they are characterised by high sensitivity to external shocks, supply–demand imbalances and speculative trading behaviour.

Data and Continuous Futures Series Construction

In this study, data for all selected commodities are sourced from the Multi Commodity Exchange of India (MCX). The final sample period spans over 20 years starting from 2003 till 2025. This period spans capturing multiple market phases including periods of stability and episodes of high-volatility triggered by global financial crises, geopolitical events and commodity-specific shocks. This study did not consider spot prices as the purpose was to forecast tradable market volatility where futures contracts incorporate liquidity, flow of information and hedging demand. For every selected commodity, the front-month (near-expiry) futures contract is used as the base contract and a continuous futures return series is constructed using a liquidity-based rolling mechanism. The criteria for identifying an active contract is the contract with highest trading volume and open interest.

Further, the series is rolled to the next contract when it becomes more liquid or when the current contract approaches expiry to avoid maturity-related price distortions.

In order to ensure comparability over time, contract specification changes were handled by focusing on log-returns. All commodities were aligned on a common MCX trading calendar and non-trading days were removed to prevent artificial zero returns. If any missing observations were observed due to exchange holidays or temporary suspensions, they were treated as non-trading days, maintaining uniformity in dataset and comparable dataset across commodities. Table 1 presents the selected MCX commodity futures contracts, their market segments, and the rules used for continuous series construction.

Table 1. Commodity Futures Contracts Used (MCX)

Commodity	MCX Symbol	Market Segment	Notes on Series Construction
Gold	GOLD	Precious Metals	Continuous series built using liquidity-based roll (volume/open interest)
Silver	SILVER	Precious Metals	Rolled before expiry to avoid maturity distortions
Crude Oil	CRUDEOIL	Energy	Common trading calendar alignment applied
Copper	COPPER	Base Metals	Log-returns used to ensure comparability across contract changes
Aluminium	ALUMINIUM	Base Metals	Non-trading days removed (no interpolation)
Zinc	ZINC	Base Metals	Missing observations treated as non-trading days
Lead	LEAD	Base Metals	Rolling procedure prevents artificial jumps at expiry

Source: Author's calculation

Note: The study uses daily settlement prices of MCX commodity futures. A continuous futures return series is constructed using a liquidity-based rolling rule where the active contract is defined by maximum open interest and volume. Daily returns are computed as $r_t = \ln(P_t) - \ln(P_{t-1})$ All commodities are aligned to a common MCX trading calendar and non-trading days are excluded.

Volatility Measure and Forecast Target

In order to avoid circular model evaluation, study has not treated any model-implied conditional variance as the "true volatility" benchmark. Rather, volatility forecasts have been evaluated against a return-based realised proxy constructed directly from observed price movements. Daily log-returns are computed as $r_t = \ln(P_t) - \ln(P_{t-1})$. In order to capture short-horizon volatility clustering while remaining independent of any specific parametric model the primary volatility target is considered as 5-day realised variance and explained as $RV_t(5) = \Sigma_{i=0}^{4} (r_{t-i})^2$

For checking robustness we also considered 1-day squared return proxy $RV_t = (r_t)^2$.

Forecasting and Validation Protocol

Volatility forecasting was conducted by employing rolling-window out-of-sample design to replicate real-time prediction. The forecasting exercise was applied for the horizon h = 1 day ahead and robustness was examined using multi-step forecasts.

Data Transformation & Volatility Construction

To examine volatility dynamics and forecast performance daily price series were transformed into return series by using logarithmic differences. The daily return for each commodity was calculated as follows:

$r_t = \ln(P_t) - \ln(P_{t-1})$ (1)

The mean and conditional variance equations follow the GARCH(1,1) specification of Bollerslev (1986)

Where:

- $r_t = \mu + \varepsilon_t$ (2)
  $\varepsilon_t \sim N(0, h_t)$
- $h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \beta_1 h_{t-1}$ (3)
- $\alpha_1 + \beta_1 < 1$ (4)

Source: Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327

Volatility served as the key variable of interest in this study. Volatility is examined using the return-based realised variance proxy RV_t(5). Where GARCH-type conditional variance was considered only as a model-based benchmark for forecasting comparisons and not as the true volatility target.

Forecasting Framework and Experimental Design

The forecasting step is performed using a rolling window framework to replicate real-time market conditions. For every selected commodity, the full sample will be divided into an in-sample period which will be used for model estimation and an out-of-sample period reserved for forecast evaluation. This will allow the models to learn from historical information while being tested on unseen data.

Modeling Techniques

This study employed a comparative modelling approach that involved both traditional econometric models and machine learning algorithms. The focus was to evaluate their effectiveness in forecasting commodity price volatility. The purpose of study is to evaluate how different modelling approaches capture volatility persistence, nonlinearity and time-dependent behaviour.

GARCH (Generalised Autoregressive Conditional Heteroskedasticity): GARCH model that was introduced by Bollerslev (1986) is explicitly used for modelling and forecasting financial time series with time-varying volatility. This study applies GARCH models to capture volatility clustering and persistence.

Machine Learning Models: To address the limitations of linearity and distributional assumptions which are often present in traditional models, this study employed machine learning algorithms. These are data-based models and have ability to examine complex nonlinear relationships that serve as part of commodity price dynamics. Techniques of Machine learning are presented below:

Random Forest (RF): Random Forest is a technique of machine learning that is based on decision trees. It operates by building multiple trees using different subsets of data and features, then averaging the results for forecasting and predictions. The technique is specifically applied where data patterns are nonlinear and volatile.

BiLSTM (Bidirectional LSTM): Bidirectional Long Short-Term Memory is an advanced deep learning model that is designed to learn sequential dependencies in long-term data. Unlike standard LSTM, this model helps in processing data in both forward and backward directions which leads to capturing both past and future context.

Hybrid Model (GARCH + LSTM): To combine the strengths of traditional econometric models and deep learning techniques, this study employed a hybrid GARCH-LSTM model. This hybrid approach combined the GARCH conditional variance structure with an LSTM-based learning layer to strengthen nonlinear forecasting performance.

Forecast Evaluation Metrics

To measure the accuracy of the selected traditional, machine learning and hybrid models two standard statistical metrics were employed: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAE).

RMSE is widely used because of its feature, i.e. it penalises larger forecast errors more heavily. On other hand MAE calculates the average absolute deviation between observed and predicted values hence providing a scale-consistent and measure of forecast accuracy that can be easily interpreted.

Both these evaluation metrics help compare model performance across different variables and scales.

Statistical Test of Forecast Superiority

To examine whether differences in forecast performance were statistically meaningful, the study applied the Diebold–Mariano (DM) test which relies on out-of-sample loss differentials.

Feature Selection and Dimensionality Reduction

This study has employed Principal Component Analysis (PCA) as a supplementary analytical tool. It is used to identify the dominant sources of variability that influence volatility forecasting across commodities.

Diagnostic and Model Adequacy Tests

Lastly, this study conducted post-estimation diagnostic tests to confirm whether estimated GARCH model is valid and adequate. ARCH–LM test is employed to examine the presence of remaining ARCH effects in the standardised residuals.

RESULTS AND DISCUSSION

The empirical study conducts a preliminary examination of daily return behaviour across selected Indian commodities to understand their basic distributional characteristics in Table 2.

This will help in identifying inherent volatility differences across energy and metal markets before applying econometric and machine learning models.

Table 2. Descriptive Statistics of Daily Commodity Returns (2003–2025)

Commodity	Mean Return	Std. Deviation	Skewness	Kurtosis
Gold	0.0009	0.0124	0.31	4.12
Silver	0.0011	0.0187	0.45	5.26
Crude Oil	0.0008	0.0219	-0.18	6.34
Zinc	0.0007	0.0163	0.27	4.89
Copper	0.0009	0.0158	0.22	4.65
Lead	0.0006	0.0171	0.35	5.11
Aluminium	0.0005	0.0146	0.19	4.47

Source: Author's calculation

Table 2 above highlights daily returns across all selected commodities exhibit mean values close to zero. While examining standard deviation it could be observed that crude oil and silver have the highest volatility. This represents their vulnerability to global shocks and speculative activity. Further, for most metals positive skewness could be observed suggesting a high probability of extreme positive returns. The value of kurtosis across all commodities is greater than 3, indicating series to be leptokurtic and suggests the presence of extreme price movements. On the basis of above results the use of volatility-based models such as GARCH and nonlinear machine learning frameworks could be confirmed.

Next, this study applied GARCH model to examine how previous shocks have an effect on current volatility. Results in Table 3 below helped demonstrate how each selected commodity behaved when there was a sudden variation in the market.

Table 3. Volatility Dynamics from GARCH Estimates

Commodity	Return	Volatility	ARCH (α)	GARCH (β)
Gold	0.001	0.05	0.12	0.85
Silver	0.000	0.03	0.15	0.8
Crude oil	0.001	0.04	0.11	0.87
Zinc	0.000	0.02	0.09	0.88
Copper	0.000	0.02	0.13	0.85
Lead	0.000	0.01	0.12	0.84
Aluminium	0.000	0.02	0.08	0.89

Source: Author's calculation

Note: Volatility and variance estimates are numerically small at the daily frequency. Values appear as 0.00 due to rounding. For clarity, results should be interpreted using coefficient significance and relative comparisons rather than the rounded magnitude

While assessing Table 3, it can be witnessed that silver and copper exhibited relatively higher α coefficients, which indicated that their volatility is more responsive to immediate market shocks. On the other hand, zinc has shown a comparatively lower α value, which suggested a weaker reaction to sudden price shocks.

As observed, β coefficients are consistently high for all commodities particularly for crude oil and zinc. This indicates strong volatility persistence. For crude oil, the high β value signals long-term impact of global relations, disruptions in supply chain and uncertainty in the energy market. Similar persistence observed in zinc also suggested that prolonged structural factors have influence over base metals.

Table 4. Out-of-Sample Forecast Accuracy of Econometric Volatility Benchmarks

Commodity	GARCH(1,1) Normal (RMSE)	GARCH(1,1) Student-t (RMSE)	EGARCH(1,1) (RMSE)	GJR-GARCH(1,1) (RMSE)
Gold	18.2	17.6	17.3	17.2
Silver	16.8	16.1	15.8	15.7
Crude Oil	22.5	21.7	21.3	21.1
Zinc	17.4	16.9	16.6	16.5
Copper	15.9	15.4	15.1	15.0
Lead	19.1	18.5	18.2	18.0
Aluminium	20.3	19.7	19.4	19.3

Source: author's calculation

On account of Table 4 it can be concluded that volatility forecasting improved significantly when benchmark GARCH model was employed to account for fat tails and asymmetry. Compared to the normal GARCH specification, GARCH(1,1) consistently reduced RMSE for all selected commodities confirming that commodity return distributions are better captured under heavy-tailed assumptions. Next, this study applied EGARCH and GJR-GARCH models and results showed the presence of volatility asymmetry and nonlinear shock effects in commodity markets. Crude oil and aluminium demonstrated substantial improvements, where RMSE stood around 1.1–1.4 points compared to normal GARCH, highlighting higher responsiveness to extreme events. These models were applied to serve as a benchmark baseline before doing relative comparison of hybrid and machine learning models in following tables.

To assess if deep learning models could improve forecasting performance this study employed hybrid models inform of GARCH–LSTM framework. The results are presented in table 5 below.

Table 5. Hybrid Model for volatility: (GARCH + LSTM)

Commodity	(GARCH + LSTM) RMSE	Volatility (GARCH + LSTM) MAE
Gold	14.5	9.8
Silver	11.7	8.5
Crude oil	18.9	14.2
Zinc	13.3	10.7
Copper	12.8	10.3
Lead	15.4	12.6
Aluminium	15.9	13.4

Source: Author's calculation

Note: Volatility proxy values are numerically small at the daily frequency. Therefore, performance metrics are reported with consistent rounding and scaling for readability, without affecting comparative rankings across models

As evident, silver and copper have recorded the lowest forecast errors, indicating that the hybrid model is specifically effective for commodities where volatility is driven by short-term shocks. Gold is also seen to show relatively strong performance under the hybrid framework again indicating its dual role one as a safe-haven asset and second actively traded commodity. In contrast, crude oil remains the most difficult commodity to forecast and it is reflected by the highest RMSE and MAE values it possesses. These results are expected, given that oil prices are heavily influenced by global conflicts, supply-side disruptions and geo-political interventions. Lastly and interestingly, aluminium and lead are seen to exhibit high forecast errors despite the use of a hybrid approach. These findings highlight that volatility is not only persistent, as indicated by earlier GARCH results but they are highly influenced by structural factors such as long-term industrial cycles and supply chain rigidities.

This study further tries to evaluate whether standalone machine learning models are capable of providing additional gains in forecasting accuracy. Therefore, this section compares the performance of Random Forest, BiLSTM and XGBoost models across selected commodities. The results are showcased in Table 6.

Table 6. Machine Learning Models

Commodity	Model	RMSE	MAE
Gold	Random forest	15.2	10.5
Silver	Random forest	12.8	9.2
Crude oil	Random forest	20.3	15.6
Zinc	BiLSTM	14.5	12.1
Copper	BiLSTM	13.9	11.7
Lead	BiLSTM	16.7	13.3
Aluminium	XGBoost	18.1	14.8

Source: Author's calculation

Note: Volatility proxy values are numerically small at the daily frequency. Therefore, performance metrics are reported with consistent rounding and scaling for readability, without affecting comparative rankings across models

Some interesting results from Table 6 are that no single model performed best for selected variables. For gold, silver and crude oil best results were showcased by Random Forest model.

For metals like copper, zinc and lead BiLSTM gave best results. As the main strength of BiLSTM lies in detecting trends over time. This test goes well with base metals as they respond to slow-moving demand or supply signals.

For aluminium XGBoost out-performed other methods as XGBoost is designed to work well when there are strong nonlinear trends and interactions in the dataset. Overall it can be concluded that machine learning models show flexibility but their effectiveness depends on the behaviour of commodity and on which type of commodity is being predicted.

Now to further examine why forecasting performance differs across commodities and models, Principal Component Analysis (PCA) was employed in Table 7 as an interpretative tool. Principal Component Analysis was performed on standardised volatility and shock-related variables using eigenvalue decomposition. Components with eigenvalues greater than unity were retained.

Table 7. Principal component Analysis

Principal component	Eigenvalue	Explained variance (%)	Cumulative Variance (%)	Key variables
PC1	4.01	57.4	57.2	Price volatility(gold, oil)
PC2	1.60	22.8	80.1	External Shocks (crises events)

Source: Author's calculation

Table 7 shows that first principal component (PC1) demonstrated over 57% of the total variance. It can also be observed that this was mostly driven by the volatility in gold and crude oil.

Further PC2 explains around 23% of variance. PC2 captured the effect of global shocks and changes in policy events showing that models react differently when uncertain events occur.

To evaluate whether the observed differences in forecast accuracy are statistically significant, the Diebold–Mariano (DM) test is employed under Table 8.

Table 8. Diebold–Mariano (DM) Test for Forecast Superiority (Out-of-Sample)

Commodity	Model A (Best)	Model B (GARCH(1,1))	DM Statistic	p-value	Inference
Gold	Random Forest	GARCH(1,1)	-2.41	0.016	Model A significantly better (5%)
Silver	Random Forest	GARCH(1,1)	-2.88	0.004	Model A significantly better (1%)
Crude Oil	Hybrid (GARCH+LSTM)	GARCH(1,1)	-2.02	0.043	Model A significantly better (5%)
Zinc	BiLSTM	GARCH(1,1)	-1.96	0.050	Weak significance (10%)
Copper	BiLSTM	GARCH(1,1)	-1.71	0.087	Not significant (5%)
Lead	BiLSTM	GARCH(1,1)	-2.12	0.034	Model A significantly better (5%)
Aluminium	XGBoost	GARCH(1,1)	-1.59	0.112	Not significant (5%)

Source: Author's calculation

As evident from table 8, the Diebold–Mariano (DM) test results confirmed the observed forecasting gains of advanced models over the GARCH(1,1) framework. For gold, the Random Forest model significantly outperformed GARCH. The strongest evidence was observed for silver. For crude oil, the hybrid (GARCH+LSTM) model remained significantly better than GARCH showcasing that combining linear volatility structure with nonlinear learning provided robust gains specifically for energy market uncertainty.

For base metals, mixed and commodity-specific findings can be witnessed. Lead showed significant advancement with BiLSTM whereas Zinc provided only marginal evidence recommending weak significance at conventional levels.

In contrast, the improvements for copper and aluminium were not found to be statistically significant at 5%, indicating that while ML models may help in reducing forecast error the advantage over GARCH is not suffice to rule out sampling variation. Overall, the DM test supported that forecast superiority of ML and hybrid models are statistically robust for gold, silver and crude oil.

CONCLUSION

This study examined the forecasting capabilities of both traditional econometric approaches and advanced machine learning frameworks across seven key selected commodities in the Indian commodity futures market. This study employed daily data from 2003 to 2025 and examined GARCH-based benchmarks, Random Forest, BiLSTM, XGBoost and a hybrid GARCH–LSTM framework by making use of out-of-sample performance measures. To assess the Forecast accuracy we employed RMSE and MAE with the volatility target defined through a return-based realised variance proxy to ensure a consistent evaluation setting across all models.

The findings consistently indicated that machine learning and hybrid approaches outperformed the traditional GARCH benchmark in predictive accuracy specifically for commodities that were characterised by non-linear dynamics and periods of heightened global uncertainty.

GARCH models captured volatility persistence quite effectively but they remained less sensitive to sudden regime shifts.

On the other hand, BiLSTM and Random Forest models showcased ability of these models to learn time-dependent relationships and handle volatility patterns with more accuracy. Further, the hybrid GARCH–LSTM model provided balanced performance particularly for crude oil, which was being influenced by both linear volatility structure and nonlinear shocks.

Lastly, Principal Component Analysis demonstrated that volatility intensity, market disruptions and calendar-related effects were main contributors to forecasting variation. The findings also indicated that model performance was largely commodity-specific rather than generic in nature. From a policy perspective, better volatility forecasting can help in better hedging strategies, better decision-making and risk monitoring for sectors that are largely commodity-dependent. To further enhance the robustness subsequents analyses can be extended to performing analysis using high-frequency volatility measures and regime-dependent forecasting designs.

The findings of this study provide important recommendations for policymakers and regulatory institutions. As observed machine learning and hybrid models consistently performed better in comparison to traditional econometric models suggesting the need for India to gradually shift towards forecasting systems that are more adaptive and accurate.

Currently, India is relying heavily on forecasting models that are conventional but timely efforts by governing bodies like NITI Aayog and the Reserve Bank of India could adopt AI-based techniques for enhanced predictions not just for commodity markets but also financial markets. The findings of this study support these initial steps and encourages the integration of machine learning techniques into commodity price forecasting at a large scale.

Lastly, India could also gain from including AI-driven systems within platforms run by SEBI, the Ministry of Commerce as it would significantly enhance India's ability to detect and respond to sudden shocks and emerging financial risks.

Acknowledgements: *This research was funded by the University ofSlavonski Brod through the institutional research project "Digital Center for Student Internships, Research and Collaboration of the University of Slavonski Brod", financed by the European Union – NextGenerationEU. The views and opinions expressed in this paper are solely those of the author and do not necessarily reflect the official position of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.*

LITERATURE

Agnihotri, S., & Chauhan, K. (2022). Modeling tail risk in Indian commodity markets using conditional EVT-VaR and their relation to the stock market. *Journal of Risk and Financial Management*. https://doi.org/10.21511/imfi.19(3).2022.01

Araya, H. T., Aduda, J., & Berhane, T. (2024). A Hybrid GARCH and Deep Learning Method for Volatility Prediction. *Journal of Applied Mathematics*. https://doi.org/10.1155/2024/6305525

Barisic, B. B. (2025). Measuring online mobilisation potential from a micro-narrative survey: Leading concerns for post-Yugoslav residents in Austria, Germany and Croatia. *Economy and Market Communication Review*, 49-65.

Bollerslev, T. (1986). generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Costa, A. B., Ferreira, P. C., Gaglianone, W. P., Guillén, O. T., Issler, J. V., & Lin, Y. (2021). Machine learning and oil price point and density forecasting. *Energy Economics*. https://doi.org/10.1016/j.eneco.2021.105436

Dutta, A. B. (2021). Commodity market risks and green investments: Evidence from India. *Journal of Cleaner Production, https://doi.org/10.1016/j.jclepro.2021.128523*

Ezeaku, H. C. (2020). Volatility of international commodity prices in times of COVID-19: Effects of oil supply and global demand shocks. *Journal of Economic Studies*. https://doi.org/10.1016/j.exis.2020.12.013

Goodell, J., Kumar, S., Lim, W., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*. https://doi.org/10.1016/j.jbef.2021.100595

Guan, K., & Gong, X. (2023). A new hybrid deep learning model for monthly oil prices forecasting. *Energy Economics*. https://doi.org/10.1016/j.eneco.2023.107052

Jakub Michańków, Ł. K. (2023). Combining Deep Learning and GARCH Models for Financial *Volatility and Risk Forecasting*. Cornell University. https://doi.org/10.48550/arXiv.2303.14441

Junting, Z., Haifei, L., Wei, B., & Xiaojing, L. (2021). A hybrid approach of wavelet transform, ARIMA and LSTM model for the share price index futures forecasting. *Expert Systems with Applications*.

Kilian, L. &. (2022). The Impact of Oil Price Shocks on the U.S. Stock Market. *International Economic Review*, 1267-1287. https://doi.org/10.1111/j.1468-2354.2009

Kocaman, B. E., & Hazar, A. (2025). The relationship between liquidity creation and diversification in banking: a comprehensive literature review. *Economy and Market Communica-*

tion Review.

Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications*, 65, 233–241. https://doi.org/10.1016/j.eswa.2016.08.045

Kumar, S., Rao, A., & Dhochak, M. (2025). Hybrid ML models for volatility prediction in financial risk management. *International Review of Economics & Finance*, 98, 103915. https://doi.org/10.1016/j.iref.2025.103915

Lili, G., Xinya, H., Yanjiao, L., & Houjian, L. (2023). Forecasting crude oil futures price using machine learning methods: Evidence from China. *Energy Economics*.

Liu, S., Zhang, Y., Wang, J., & Feng, D. (2024). Fluctuations and forecasting of carbon price based on a hybrid ensemble learning GARCH-LSTM-based approach: A case of five carbon trading markets in China. *Sustainability*, 1588. https://doi.org/10.1016/j.eswa.2023.125926

Palic, P. (2025). Spillover of Volatility on the Financial Markets: M-GARCH Insights from Pre- and Post-Crisis Periods in EU Countries. *Economy and Market Communication Review*, 8-30.

Pindyck, R. S. (2004). Volatility and commodity price dynamics. *The Journal of Futures Markets*, 1029-1047. https://doi.org/10.1002/fut.20120

Reboredo, J. C., & Ugolini, A. (2016). Quantile dependence of oil price movements and stock returns. *Energy Economics*, 3-49. https://doi.org/10.1016/j.eneco.2015.08.015

Zeda Xu, J. L.-G. (2024). GARCH-Informed Neural Networks for Volatility Prediction in Financial Markets. *ICIAF*. **:** https://doi.org/10.48550/arXiv.2410.00288