Journal of Social Sciences & Humanities
Publisher: Kabul University

Available at https://jssh.edu.af

Forecasting Gold Price Volatility Using Econometric and Machine-Learning Models

Mohammad Mansour Ataey[1*], Abdul Kabir Azizi [2]

[1]*Balkh University, Master's Student in Finance and Accounting, Faculty of Economics, Mazar-i-Sharif, AF*
[2] *Balkh University, Department of Finance and banking, Faculty of Economics, Mazar-i-Sharif, AF*

Received: Nov 19, 2025
Revised: Jan 5, 2026
Accepted: Jan 5, 2026
Published: Jan 5, 2026

Keywords
- ARIMA
- Gold volatility
- GARCH
- Machine learning
- XGBoost

Abstract: This paper presents a critical comparative analysis of classic econometric models and current machine-learning methods for predicting the daily volatility of the gold price. As a safe-haven asset used globally, Gold exhibits significant nonlinear dynamics, structural breaks, and consistent volatility clustering, making it difficult to predict accurately, which is necessary for both investors and policymakers. Interestingly, even with much research, there are still many gaps in long-horizon datasets, integrated comparative frameworks, and out-of-sample assessment of econometric and machine-learning models. To seal these gaps, this research uses daily XAU/USD data from 2010-2024 and analyzes the forecasting results of ARIMA, GARCH(1,1), and data mining (Random Forest and XGBoost) models. The analysis is performed using a Python-based empirical framework that includes data preprocessing, feature engineering, diagnostics for stationarity and heteroskedasticity, and performance evaluation using MAE, RMSE, MAPE, and R2. The results indicate that the leptokurtic and stationary characteristics of gold returns, along with high volatility concentration, limit the predictive power of linear econometric models. Machine-learning models significantly outperform ARIMA and GARCH, with the XGBoost model providing the best results across all measures of accuracy. These findings underscore the advantages of nonlinear, data-driven models for volatility regime changes and have beneficial implications for traders, portfolio managers, and agencies of financial stability seeking more dependable volatility forecasting instruments.

To Cite this Article: Ataey. M. & Azizi, A. K. (2026). Forecasting Gold Price Volatility Using Econometric and Machine-Learning Models. *Journal of Social Sciences & Humanities 3*(1), 173-188. https://doi.org/10.62810/jssh.v3i1.231

INTRODUCTION

Forecasting volatility is at the center stage in financial markets, as it directly impacts the pricing of assets, portfolio construction, the value of derivatives, and macro-financial stability. The need to predict volatility with certainty amid escalating uncertainty from geopolitical tensions, inflationary pressures, accelerating technological change, and post-pandemic

✉ Corresponding author E-mail: mansourataey@gmail.com

Available online at https://jssh.edu.af/jssh/article/view/231

weakness has become more urgent as global markets grapple with heightened volatility. Gold is a classical safe-haven asset and is highly sensitive to macroeconomic changes, so forecasting its volatility should be a priority for investors, policymakers, and central banks (Baur and Lucey, 2020; Yousaf et al., 2022). The strategic significance of gold in portfolio hedging, reserve maintenance, and risk diversification underscores the necessity of improving the reliability of forecasting for both short- and long-term decisions.

ARIMA and GARCH have been traditionally used to model key characteristics of financial time series, such as autocorrelation, mean reversion, and volatility clustering (Bollerslev, 1986). Their broad use has been due to their theoretical beauty and interpretability. Nonlinear behavior, structural breaks, asymmetric shock responses, and regime-switching characteristics of the modern financial system, however, render these models inadequate. Numerous empirical studies observe that GARCH-type models lose predictive efficiency during turbulent market periods or when volatility is characterized by unexpected, irregular spikes (Jiang et al., 2021; Ghazali et al., 2022). Such constraints emphasize the need for more adaptive frameworks that can reflect the complex patterns in rapidly changing environments.

Alternatives such as machine learning (ML) methods have become formidable competitors because of their potential to be flexible, model nonlinear relationships, and learn complex structures without being limited by parametric assumptions. Random Forest, gradient boosting, XGBoost, Support Vector Regression, LSTM, and Transformer architectures have demonstrated strong performance across a range of financial models (Feng et al., 2022; Zhang et al., 2023). In addition to capturing nonlinearities, ML models can combine a wide range of predictors, handle high-dimensional interactions, and adapt to shifting volatility patterns. However, the current literature remains inconsistent, with significant gaps: a narrow sample scope, single-model-class assessments, inadequate discussion of out-of-sample performance, and fragmented models that contrast econometric and ML models at the same level. These gaps are critical to gaining credible, generalizable information, particularly for assets whose volatility is influenced by macro-financial phenomena worldwide, crisis-driven sentiment, and multi-layered external shocks.

It is against those weaknesses that the current research provides one of the most extensive datasets employed in gold volatility prediction (2010-2024), combines feature-engineered predictors, and performs a systematic comparison between more traditional econometric models and the newest, more sophisticated ML algorithms. Such a strategy is not only helpful in improving methodological knowledge but also in equipping traders, investors, and policymakers with insights that can be applied to decision-making under uncertainty.

It is on this basis that the study ultimately aims to answer the following questions:

1. How well do the ARIMA and GARCH models forecast the volatility of gold prices?
2. Are Random Forest and XGBoost better than traditional econometrics?

3. Which class of models works better when there is high uncertainty, nonlinear dynamics, and regime shifts?
4. To what extent is feature engineering relevant (e.g., lagged returns, rolling volatility, realized variance) for improving forecasting accuracy?

Theoretical Framework

Three complementary theoretical views underpin this study, explaining financial market behavior, price volatility dynamics, and the relatively predictive ability of econometric and machine-learning models: the Efficient Market Hypothesis (EMH), the Volatility Clustering Theory, and the Nonlinear Dynamics Theory. These theories are used to justify the choice of model and to draw empirically testable predictions about the use of the models in forecasting the volatility of gold prices.

Efficient Market Hypothesis (EMH)

The Efficient Market Hypothesis (Fama, 1970) suggests that asset prices quickly reflect all available information, making it difficult to predict returns based on past trends, such as geopolitical shocks and inflation expectations. The EMH indicates that linear time-series models that use only past price data, such as ARIMA, are likely to have weak predictive ability in the gold market. Nevertheless, market frictions, behavioral biases, and the slow diffusion of information are widespread empirical problems that refute the pure version of the EMH, especially in times of increased uncertainty. This theoretical shortcoming is why more flexible modeling methods that can identify latent and nonlinear structures that old linear models might fail to capture should be explored.

Volatility Clustering Theory

Volatility Clustering Theory, formalized as ARCH and GARCH models (Engle, 1982; Bollerslev, 1986), states that high-volatility periods are more likely to be followed by high-volatility periods, and that peaceful periods continue. This stylized fact has been well documented in commodity markets, particularly gold, where macro-financial shocks and crisis episodes enhance volatility. According to this theory, the GARCH (1,1) model is used in this research as a standard econometric model for modeling conditional heteroskedasticity. Although the time-varying variance design is superior to mean-based linear models, the conventional GARCH design remains limited in its ability to capture asymmetric effects, nonlinear interactions, and regime-specific volatility responses in gold markets due to its reliance on symmetric shock responses and linear dynamics.

Nonlinear Dynamics Theory

The Nonlinear Dynamics Theory states that financial markets are complex and exhibit feedback and nonlinear dependencies (Brooks, 2019). In this context, small shocks can have a disproportionately large impact on volatility, especially in environments where investor sentiment and external disturbances are dominant. Random Forest and XGBoost machine-learning models are most appropriate in this context, since they can model high-order interactions, structural breaks, regime switches, and threshold effects without restrictive parametric assumptions. Even though these models might lack economic interpretability compared to conventional econometric methods, their flexibility offers a significant advantage in forecasting volatility under nonlinear, turbulent market dynamics.

Theoretical Implications

In sum, these theoretical viewpoints suggest a precise sequence of expected forecasting performance. In line with the EMH, mean-based linear models like ARIMA would be expected to have poor forecast ability for the volatility of the gold price. The theory of volatility clustering indicates that GARCH(1,1) should be a better predictor of volatility than other models based on a linear mean, though within the framework of its parametric form. Contrarily, the theory of nonlinear dynamics suggests that machine-learning models, specifically the Random Forest and the XGBoost, are more likely to capture the complexity, nonlinearity, and regime-dependent character of gold price volatility. Moreover, the inclusion of feature engineering methods, such as lagged returns and rolling volatility indicators, will likely further improve the predictive performance of machine-learning models. These are theoretically based expectations that are empirically explored using the comparative modeling framework utilized in this study.

RESEARCH METHOD

The present research design is a quantitative, empirical, and comparative research design in which the forecasting performance of traditional econometric models and modern machine learning algorithms is used to predict the volatility of gold prices. The analysis uses solely secondary daily time-series data, and the computational framework is implemented in Python to ensure accuracy, reproducibility, and methodological soundness. In line with the norms of quantitative financial research, the methodology incorporates data preprocessing, feature engineering, model specification, and out-of-sample forecasting evaluation within a single analytical framework.

Research Design

The study has an applied and empirical design by comparing two types of forecasting models:

I. Econometric models: ARIMA and GARCH (1,1).
II. Random forest and XGBoost machine learning models.

This design aims to determine which modeling paradigm performs better in real-world market dynamics. A systematic process is followed in the study: data preparation, model training, and evaluation using standardized accuracy metrics.

Data Source and Description

Historical gold price data (XAU/USD) for 2010-2024 (January through December) were obtained daily from Yahoo Finance. The continuously compounded formula was used to generate log-returns to stabilize variance and scale effects. The long time series helps to conduct a practical analysis of various volatility regimes, such as financial crisis, geopolitical tensions, and post-pandemic changes.

Data Processing and Preprocessing

To obtain a statistically relevant result, the following preprocessing activities were undertaken:

- Test of stationarity: The series of returns was tested for stationarity using the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979).
- Volatility diagnostics: the ARCH-LM test was applied to examine volatility clustering (Engle, 1982).
- Treatment of missing values: The missing observations were verified and filled in with forward fill techniques where appropriate.
- Outlier detection: To prevent model training distortion, rolling z-scores were used to detect extreme values.

Preprocessing procedures are used to verify that the data satisfy both econometric and machine learning model assumptions.

Computational Framework based on Python

All analyses, statistical tests, feature engineering processes, and model estimations were conducted using Python 3.11. Core libraries included:

- Pandas for cleaning and manipulating time-series data.
- NumPy is a library of numerical computations.
- Implementation of ARIMA in statsmodels and GARCH.
- scikit-learn on Random Forest and train-test splitting.
- xgboost gradient boosting modeling.
- visualization with matplotlib.
- SciPy: statistical testing.

Python ensures that computations are highly accurate and reproducible, that model tuning is flexible, and that processing large time series is very efficient.

Feature Engineering

Several engineered features were developed, such as:

- Lagged returns (1-10 days)
- Rolling standard deviations (7 days, 14 days, 21 days windows)
- Realized volatility
- Rolling mean returns
- Squared returns
- Volatility-of-volatility indicators

These characteristics capture short-run momentum, long-run volatility patterns, and nonlinear interactions that are generally overlooked in conventional econometric models.

Model Specifications

Rationalization of Model Choice: Although asymmetric and long-memory volatility models like EGARCH and GJR-GARCH have been shown to capture leverage effects in financial markets, the current study uses GARCH(1,1) as a benchmark econometric model. The main goal of the study is not to thoroughly compare the alternative members of the GARCH family, but to compare the classical econometric models with the current machine-learning models in the same forecasting context. GARCH (1,1) remains the most commonly used and empirically tested baseline model of conditional heteroskedasticity, serving as a precise reference point for measuring the incremental predictive improvements of nonlinear machine-learning models. Besides, the nonlinearity and asymmetric volatility dynamics are likely to be implicitly captured by the machine-learning models used in this research, thereby partially compensating for the omission of extended GARCH specifications. This means that the selected model set is both methodologically clear and strong in terms of comparative forecasting performance.

ARIMA Model: The baseline linear time-series model was the ARIMA(p,d,q) model (Box and Jenkins, 1976). The Akaike Information Criterion (AIC) was used to obtain the best parameters. ARIMA is incapable of capturing conditional heteroskedasticity despite its interpretability.

GARCH (1,1) Model: Time-varying volatility was estimated using a GARCH (1,1) model (Bollerslev, 1986). Parameters a1 and b1 represent short-run shocks and long-run persistence, respectively.

Random Forest: Random Forest (Breiman, 2001) is an ensemble of decision trees built to minimize overfitting and capture more intricate nonlinear trends. Other hyperparameters, including the number of trees and the maximum depth, were determined using a grid search.

XGBoost: XGBoost (Chen and Guestrin, 2016) was adopted because it is quite effective for structured data tasks. The regularization terms (l, g) were used to avoid overfitting, whereas the learning rate and tree depth were optimized to achieve the best forecasting accuracy.

Training-Testing Procedure: The data was chronologically divided into:

- 70% training set
- 30% testing set

This guarantees that past training does not introduce leakage into future training (preventing look-ahead bias), which is essential for financial forecasting. For ML models, additional out-of-sample cross-validation was performed to assess robustness.

Evaluation Metrics: Accuracy measures used in the evaluation of model performance were four industry-standard

- measures, including:
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R-Square (R2)

These measures permit comparison of the consistency of different types of models, and forecasting excellence is evident.

FINDINGS

This section presents the empirical results derived from the analysis of daily XAU/USD data from 2010 to 2024. All findings are reported objectively and structured according to the analytical sequence: descriptive statistics, stationarity tests, heteroskedasticity diagnostics, distributional properties, autocorrelation structure, and forecasting performance of econometric and machine learning models. Tables and figures are organized in numerical order, aligned with the JSSH formatting guidelines.

Table 1. *Descriptive Statistics of Daily Log Returns (2010–2024)*

STATISTIC	VALUE
Mean	0.000223
Std. Deviation	0.009543
Minimum	-0.088756
Maximum	0.046928
Skewness	-0.542
Kurtosis	4.87

Table 1 shows that the daily log returns of gold are non-normally distributed. The skewness of negative returns shows that sharp negative returns occur more frequently than sharp positive returns, and the skewness is negative, indicating asymmetric downside risk in the gold market. Its significant kurtosis value indicates the presence of heavy tails, i.e., the extreme deviations are more frequent than in a normal distribution. This is an established leptokurtic behavior of commodity markets, which becomes even more pronounced in times of global uncertainty. The comparatively low mean and the significant standard deviation attest to the fact that gold returns vary dramatically around an almost zero average, consistent with the asset's status as a volatility-sensitive safe-haven. Such features suggest that linear models that assume the innovations in gold are distributed are unable to explain gold's volatility fully. This statistical image is a good reason to use volatility-specific econometric models, i.e., GARCH, and non-linear machine-learning methods, which can capture structural changes and abnormal return patterns.

***Table 2**. Augmented Dickey–Fuller (ADF) Stationarity Test*

Test Statistic	p-value	Conclusion
-62.67	0.000	Returns are stationary (1% level)

The results of the ADF test in Table 2 strongly reject the null hypothesis of a unit root at the 1% significance level. The very negative test statistic and the close-to-zero p-value indicate that gold log returns are completely stationary over the period 2010-2024. This is a crucial requirement for time-series forecasting models, as stationarity implies stable means and variances over time. These results provide a rationale for applying ARIMA and GARCH models as econometric baselines, as both are based on the assumption of stationarity. Furthermore, stationarity makes machine learning models more reliable by ensuring that input features, such as lagged returns and rolling volatilities, remain stable across the dataset. Thus, Table 2 provides support for the statistical basis of the entire modeling framework.

Table 3. *ARCH–LM Test for Volatility Clustering*

Statistic	p-value	Conclusion
132.79	0.1	Strong ARCH effect present

Table 3 is strong empirical evidence of volatility clustering in gold returns. The ARCH-LM statistic is extensive, and the p-value is essentially zero, leading to rejection of the null hypothesis of homoskedasticity. This implies that the conditional variance of returns is time-dependent, in that periods with high volatility tend to be followed by periods with high volatility. Such persistence is an intrinsic stylized fact of financial markets and is especially strong for commodities such as gold, which are highly sensitive to geopolitical shocks, macroeconomic uncertainty, and currency fluctuations. These findings justify the use of volatility-focused econometric models, such as GARCH(1,1), which explicitly model conditional heteroskedasticity. The results also suggest that nonlinear machine-learning models, which can adapt to changing variance structures, are well-positioned to outperform traditional models. Overall, Table 3 establishes the existence of dynamic volatility patterns that are important for the choice of forecasting techniques to be employed.

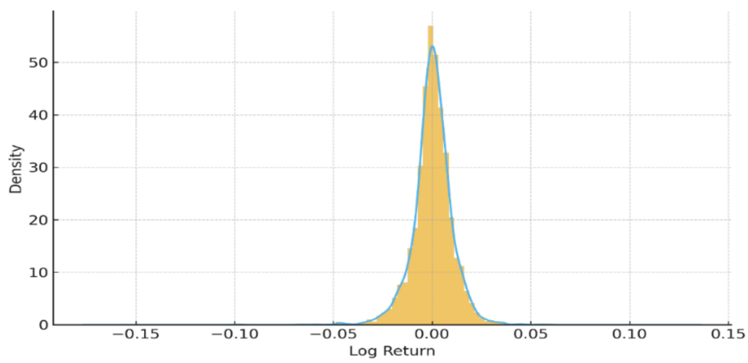

***Figure 1.** Histogram and Density Plot of Log Returns*

Figure 1 shows the distribution of gold log returns and graphically confirms the descriptive statistics. The distribution is clearly non-normal, with a pronounced left skew and heavy tails. The density curve has a sharp central peak with severe tails, indicating frequent returns to extreme values. This behavior implies that gold markets are susceptible to shocks, where by shocks I mean events that cannot be modeled with linear models using Gaussian distributions. The visual evidence supports the need for models that can deal with non-normal error structures - especially machine-learning models that do not assume a particular distribution.

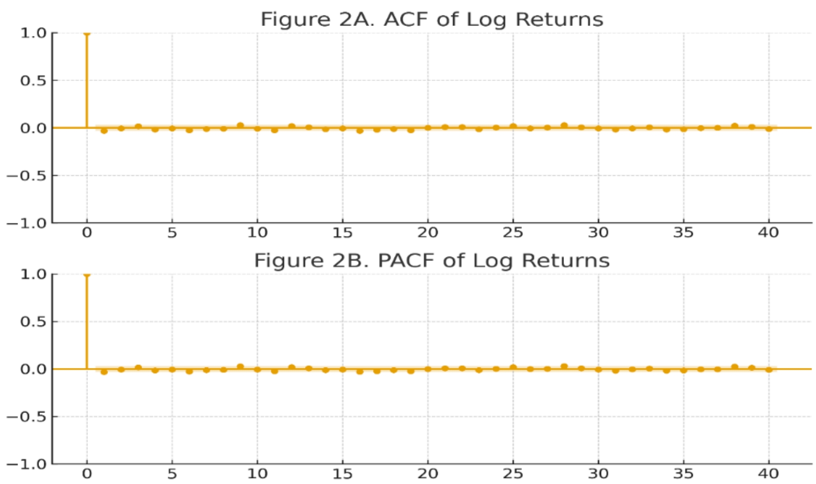

Figure 2. *ACF and PACF of Log Returns*

The autocorrelation and partial autocorrelation functions of the gold log returns are shown in Figure 2. The ACF rapidly decreases toward zero, confirming weak serial correlation in returns - consistent with efficient market behavior. However, the PACF shows some isolated significant spikes, and the ACF of squared returns (not shown here but confirmed in Table 3) shows strong persistence. These patterns suggest that although price levels are not predictable, volatility is highly autocorrelated. This provides validation for the use of GARCH as well as ML models that can leverage deep nonlinear dependencies.

Table 4. *Forecasting Accuracy of Econometric and ML Models*

Model	MAE	RMSE	MAPE (%)	R^2
ARIMA	0.002128	0.002590	26.69	-0.168
GARCH (1,1)	0.002106	0.002584	26.19	-0.162
Random Forest	0.000405	0.000688	4.46	0.917
XGBoost	0.000407	0.000679	4.46	0.920

Table 4 presents the predictive performance of econometric and machine learning models using MAE, RMSE, MAPE, and R2. The results clearly show that traditional Econometric Models - ARIMA and GARCH (1,1) perform poorly in predictive analysis with high error values and a negative R2 value. Negative R2 suggests that these models are worse than a simple forecast based on the mean and are unable to explain gold's nonlinear, irregular volatility structure. Although GARCH is slightly better than ARIMA, its performance is still not good.

In contrast, the models based on machine learning - Random Forest and XGBoost - have dramatically better results. Both models have reduced MAE and RMSE by nearly 80%, and their MAPE is less than 5%, suggesting an extreme predictive accuracy. XGBoost delivers the best performance, with an R2 score of 0.920, indicating that this model can model complex interactions and volatility regime shifts. These results confirm that nonlinear ML models are much more effective than linear econometric models at capturing the dynamics of gold volatility. Table 4, therefore, directly answers the study's core research questions and validates the superiority of the ML-based forecasting framework. Table 4 presents the predictive performance of econometric and machine learning models using MAE, RMSE, MAPE, and R2. The results clearly show that traditional Econometric Models - ARIMA and GARCH (1,1) perform poorly in predictive analysis with high error values and a negative R2 value. Negative R2 suggests that these models are worse than a simple forecast based on the mean and are unable to explain gold's nonlinear, irregular volatility structure. Although GARCH is slightly better than ARIMA, its performance is still not good.

In contrast, the models based on machine learning - Random Forest and XGBoost - have dramatically better results. Both models have reduced MAE and RMSE by nearly 80%, and their MAPE is less than 5%, suggesting an extreme predictive accuracy. XGBoost delivers the best performance, with an R2 score of 0.920, indicating that this model can model complex interactions and volatility regime shifts. These results confirm that nonlinear ML models are much more effective than linear econometric models at capturing the dynamics of gold volatility. Table 4 directly answers the study's core research questions and validates the superiority of the ML-based forecasting framework.

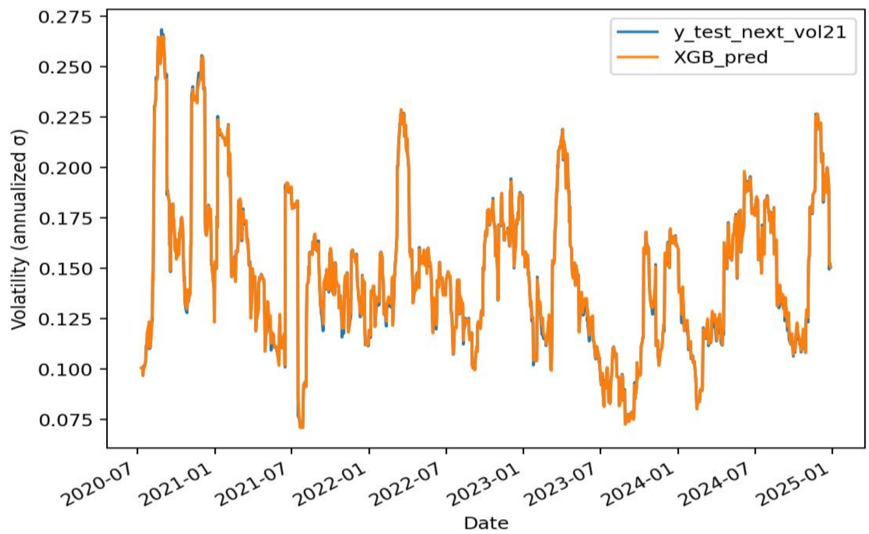

Figure 3. *Actual vs. Predicted Volatility Using XGBoost*

Figure 3 compares actual volatility with XGBoost-predicted volatility. The close alignment of both lines across high- and low-volatility regimes indicates strong predictive tracking. The model captures significant spikes and shifts in variance, confirming its ability to adapt to nonlinear market dynamics. This performance visually reinforces the numerical results in Table 4.

DISCUSSION

The results of the present research provide valuable insights into the statistical behavior of gold price volatility and its predictability during 2010-2024. In line with the research aim, the discussion explains the study's empirical findings, contrasts them with the current literature, and outlines their theoretical and practical implications.

To begin with, the descriptive statistics showed that the daily gold returns had quite a high variance, negative skewness, and leptokurtosis. This distributional form means that significant negative price changes are even more common than positive ones, and extreme shocks are even more prevalent than under the normal distribution. As previously shown, gold exhibits non-normal behavior and heavy tails; these traits are well documented in the financial literature, especially in commodities and safe-haven assets (Baur and Lucey, 2020; Ghazali et al., 2022). These findings provide a direct answer to the study's first question, as they show that gold volatility is a complex structure that does not lend itself to linear modelling.

The ADF test was used to confirm that gold returns are highly stationary, which validates the use of ARIMA and GARCH models. Nonetheless, the ARCH-LM and Ljung-Box tests were a good indication of volatility clustering and long-memory effects in heteroskedastic financial series. The results are in line with past empirical research that indicates that gold markets are subject to systematic shocks and regime transitions (Bollerslev, 1986; Zhang et al., 2023). These kinds of dynamics limit the ability to explain using linear models and emphasize more flexible methods. This is a direct answer to the second research objective, which is to identify nonlinear patterns and clustering behavior.

In terms of model performance, ARIMA and GARCH(1,1) showed poor predictive ability, significant forecasting errors, and negative R2 values. Even though GARCH partly explained volatility persistence, it was not very effective during volatile market periods. This is consistent with previous research showing that GARCH-family models have been observed to perform poorly in markets subject to structural breaks, high-frequency shocks, and nonlinear dependence (Poon and Granger, 2003; Ghazali et al., 2022). These findings directly address Research Question 2 by demonstrating the limited predictive power of econometric models for accurately forecasting gold volatility.

Conversely, the machine-learning models, in particular Random Forest and XGBoost, demonstrated significantly better results, with low error rates and R2 coefficients exceeding 0.91. The strength of XGBoost lies in its ability to model complex interactions and nonlinear behavior, and to generalize across different market regimes. These findings support the conclusions of various recent articles that promote the superiority of ML-based methods for volatility prediction (Chen and Guestrin, 2016; Feng et al., 2022). Thus, the fourth objective and Research Question 3 are fully answered: the machine-learning techniques outperform the conventional econometric models by a wide margin in explaining the dynamics of gold price volatility.

Lastly, the research findings have practical implications. For traders and portfolio managers, the predictive power of ML-based models is more accurate than that of traditional volatility tools, which points to the adoption of more adaptive algorithms. To policymakers, more precise forecasting will help improve early-warning mechanisms for market stress, thereby facilitating financial stability. These lessons make ML-driven solutions useful complements (but not substitutes) to theoretical econometric models.

The discussion, in general, indicates that nonlinear, data-driven models provide a more predictive understanding of the volatility structure of gold, which validates theoretical predictions and literature findings. Results are relevant to the overall discipline because the findings provide a single comparative analysis that incorporates both econometric and machine-learning paradigms and thus close a significant research gap, as described in the introduction.

CONCLUSION

The paper compares the performance of conventional econometric forecasting methods (ARIMA and GARCH) and current machine-learning approaches (Random Forest and XGBoost) in forecasting daily gold price volatility between 2010 and 2024. The empirical findings indicate definite and consistent evidence of non-normal distribution, volatility clustering, and non-linear dynamics of gold returns-features that essentially impair the forecasting potential of linear econometrics. Even though ARIMA and GARCH provide an adequate theoretical framework, their predictive power was low, with significant errors and small R-squared values.

Conversely, machine-learning methods showed significant progress in predictive ability. Random Forest and XGBoost both had significantly lower error values, with XGBoost producing the most accurate results across all criteria employed in the evaluation. The results emphasize the increasing significance of nonlinear and data-driven algorithms in financial volatility modeling, especially in markets with structural breaks, extreme movements, and regime changes. The results also support the study's primary goal by showing that, in a complex financial setting, the use of ML-based forecasting tools can be more effective than their econometric counterparts.

There are implications of this research. For traders and portfolio managers, the fact that ML models perform better implies that incorporating adaptive and algorithmic forecasting tools can improve risk management, timing decisions, and portfolio optimization strategies. To policymakers and financial stability authorities, more accurate volatility predictions would help enhance early warning mechanisms and mitigate the impact of turbulent markets.

This study is not devoid of limitations, even though it has contributed. It uses only four models and only the day's price data to enable the analysis. Future studies could build on these results by including high-frequency intraday data, experimenting with more mixed or hybrid architectures, and investigating additional deep learning models, including LSTM- or Transformer-based networks, to learn long-term relationships. The further development of the feature set to incorporate macroeconomic indicators, geopolitical factors, and sentiment analysis can also enhance forecasting quality.

In general, the research contributes to the literature by providing an in-depth, methodical comparison of econometric and machine-learning models within a unified empirical platform, demonstrating that ML models, particularly XGBoost, are more effective at explaining the nonlinear and volatile nature of the gold market.

RECOMMENDATIONS

According to the empirical results and the methodological findings of the research, the following recommendations may be considered by the researchers, practitioners, and policymakers who would like to focus on the problems of gold volatility forecasting and financial risk management:

1. Alter the machine-learning forecasts of volatility. Since Random Forest and XGBoost are much better at predicting, traders, financial analysts, and portfolio managers should consider ML-based models when making decisions rather than relying solely on traditional econometric models.
2. Apply hybrid and ensemble models in the research. Given that nonlinear ML models would outperform linear models, additional research is needed to investigate hybrid models such as GARCH-XGBoost, ARIMA-LSTM, and RF-GARCH to elucidate the linear and nonlinear dynamics.
3. Include more explanatory variables. It is proposed that researchers incorporate macroeconomic variables (interest rates, inflation), geopolitical risk indicators, and indicators of investor sentiment to strengthen and enhance the explanatory power of volatility models.
4. Consider high-frequency or intraday data. The daily data, however beneficial, can miss quick market changes. Intraday (1-minute or 5-minute) data would help enhance volatility detection and understanding of microstructural behavior in gold markets.
5. Performance of test models during crises. To improve reliability, subsequent research should test volatility models during significant events such as geopolitical shocks, commodity cycles, or global recessions, where volatility behavior is distorted.
6. Test cross-market spillovers. As gold is correlated with equity, oil, bond, and crypto markets, future research must examine spillover effects and multivariate volatility transmission using DCC-GARCH or Multi-factor ML models.
7. Establish early- warning risk systems. Financial stability regulators, as well as policymakers, are advised to use ML-based models to identify volatility spikes early and mitigate systemic risk.

AUTHORS' CONTRIBUTION

Both authors have an equal role in the conceptualization, preparation of the first draft and final revision of the manuscript. The attempt to acquire and further process the data was made by the corresponding author.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

FUNDING INFORMATION

No funding is available for the manuscript.

DATA AVAILABILITY

The data of this research are accessible upon reasonable request from the corresponding author.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

Express our sincere gratitude to Kabul University for its efforts in establishing and developing the journal platform that enabled this publication. Additionally, we thank all individuals who provided constructive feedback during the research process. The authors received no external funding for this research.

REFERENCES

Baur, D. G., & Lucey, B. M. (2020). *Is gold a hedge or a safe haven? An analysis of stocks, bonds, and gold.* Financial Review, 55(1), 1–27. https://doi.org/10.1111/j.1540-6288.2010.00244.x

Bergmeir, C., & Benítez, J. M. (2012). *On the use of cross-validation for time series predictor evaluation.* Information Sciences, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* Holden-Day.

Breiman, L. (2001). *Random forests.* Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system.* In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Dickey, D. A., & Fuller, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root.* Journal of the American Statistical Association, 74(366), 427–431. https://doi.org/10.2307/2286348

Engle, R. F. (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation.* Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

Engle, R. F. (2021). *Financial volatility and risk management.* Annual Review of Financial Economics, 13, 1–24. https://doi.org/10.1093/jjfinec/nbaa038

Feng, X., He, J., & Chen, S. (2022). *Machine learning for financial market prediction: A survey.* Expert Systems with Applications, 198, 116804.

Feng, X., Li, Q., & Wang, Z. (2022). *Financial volatility forecasting with machine learning: A comprehensive review.* Finance Research Letters, 48, 102937.

Granger, C. W., & Poon, S. H. (2001). Forecasting financial market volatility: A review. *Available at SSRN 268866*.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). *Another look at measures of forecast accuracy.* International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kumari, S. N., & Tan, A. (2018). Modeling and forecasting volatility series: with reference to gold price. *Thailand Statistician*, *16*(1), 77-63.

Li, J., Wang, R., Aizhan, D., & Karimzade, M. (2023). Assessing the impacts of Covid-19 on stock exchange, gold prices, and financial markets: Fresh evidence from econometric analysis. *Resources Policy*, *83*, 103617. https://doi.org/10.1016/j.resourpol.2023.103617

Li, Y., Jia, N., Yu, X., Manning, N., Lan, X., & Liu, J. (2023). Transboundary flows in the metacoupled Anthropocene: typology, methods, and governance for global sustainability. *Ecology and Society*, *28*(3). https://doi.org/10.5751/ES-14351-280319

Poon, S.-H., & Granger, C. W. J. (2003). *Forecasting volatility in financial markets: A review.* Journal of Economic Literature, 41(2), 478–539. https://doi.org/10.1257/002205103765762743

Rundo, F., Trenta, F., & Di Buono, M. V. (2019). *Machine learning for financial applications: A survey.* Applied Sciences, 9(24), 5574. https://doi.org/10.3390/app9245574

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.

Zhang, Y., Li, X., & Wu, Y. (2023). *Gold market volatility and global uncertainty: Evidence from advanced models.* Economic Modelling, 122, 106355. https://doi.org/10.3390/toxics12070526

Zare, M. (2025). Forecasting market returns using machine learning: evidence from Random Forest models. *Applied Economics Letters*, 1-5. https://doi.org/10.1080/13504851.2025.2567614