FORECASTING FINANCIAL MARKET VOLATILITY: A COMPARATIVE STUDY OF TRADITIONAL MODELS AND MACHINE LEARNING TECHNIQUES

Naman Aggarwal (202456971)

STRATHCLYDE BUSINESS SCHOOL M.Sc. Business and Management

Abstract

Volatility forecasting in financial markets is an important part of effective risk management, portfolio optimization, and decision-making. This dissertation examines the comparative effectiveness of traditional statistical models - Exponentially Weighted Moving Average (EWMA) & Generalized Autoregressive Conditional Heteroskedasticity (GARCH) as well as advanced machine learning techniques such as Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) in predicting stock market volatility. Utilizing a decade-long dataset spanning from 2014 to 2024, including indices such as the S&P 500, NASDAQ-100, and NIFTY-50, alongside prominent companies like Apple, Microsoft, and JPMorgan Chase, the study highlights the nuanced strengths and limitations of each methodology. While GARCH and EWMA excel in capturing volatility clustering and steady trends, they fall short in dynamic market shifts. Conversely, XGBoost emerges as the most robust performer, capturing non-linear relationships and adapting to extreme volatility, although its computational intensity poses challenges. LSTM demonstrates promise in modeling sequential data but struggles during abrupt market transitions. By integrating rigorous data preprocessing, feature engineering, and comprehensive evaluation metrics, the research underscores the importance of aligning model selection with specific use cases. The findings advocate for a hybrid approach, leveraging the strengths of both traditional and advanced methods to enhance predictive accuracy and facilitate actionable insights for diverse stakeholders, from institutional investors to individual traders.

Keywords

Volatility; GARCH; EWMA; LSTM; XGBoost; Forecasting; S&P500

Table of Contents

List of Tables

List of Figures

Introduction

The task of financial market forecasting has become a sufficiently popular activity among both researchers and practitioners due to its potential to have far-reaching consequences for decision-making, risk management, and portfolio optimization. Among the key parameters focused on within a broad framework of financial analysis, stock market volatility is the extent to which asset prices vary over time. It is important to understand market volatility, as it evidences investor uncertainty, creates opportunities for traders, and increases the risk of financial instability. History is replete with events that have shown how unpredictable market fluctuations can cause massive destruction: the 2008 global financial crisis and the COVID-19 market shock in 2020 are two such cases. These instances underpin the need for strong forecasting techniques which would help investors and institutions navigate such disruptions with aplomb.

Volatility plays a critical role in modern finance. It is a key factor in valuing derivative instruments like options, where it is one of the primary components for pricing. Beyond that, understanding and forecasting volatility is essential for creating effective hedging strategies and optimizing investment portfolios, helping to achieve the right balance between risk and return. Accurate volatility predictions provide a valuable advantage, allowing investors and businesses to anticipate and respond to market movements with confidence. However, forecasting volatility remains a challenge due to the ever-changing and unpredictable nature of financial markets, compounded by the noise and complexity of high-frequency trading data.

Traditional models, like GARCH, are widely applied in volatility forecasting because they are effective at modeling time series with changing variances. These models, however, are limited to linear assumptions about the relationships they model, assumptions that often do not hold in these markets. A new direction has emerged over the past two decades: the application of machine learning to financial forecasting. Techniques like Random Forest and Gradient Boosting have shown great performance in handling nonlinear relationships and processing large datasets efficiently. This dichotomy between traditional and ML-based approaches represents a fertile ground for exploration, particularly as technology continues to transform financial decision-making.

My personal motivation for the dissertation comes through the interest in forecasting and the long-held fascination with prediction of future trends. This has been driven by interests in understanding from this research, whether unforeseen events-can a market crash-be anticipated from advanced methodologies. My academic founding in finance and data analytics has therefore fueled this quest to inspire in me the ability to examine these traditional statistical model intersections with their ML techniques to foresee instances of volatility. Therefore, this dissertation attempts to meaningfully contribute to the fast-growing domain of financial analytics through in-depth analysis and a critical review of such themes.

The major goal of the current dissertation is to study and compare the effectiveness of traditional and machine learning approaches in forecasting the volatility of stock markets. It carries out the study of various financial instruments: three major indices, such as S&P 500, NASDAQ-100, and NIFTY-50, and three major corporations such as Apple Inc., Microsoft, and JPMorgan Chase. This paper, therefore, considers a total tenure of ten years, starting from January 2014 to December 2024, considering the best performance evaluation of a comprehensive dataset under fluctuating market conditions, apart from studying the feasibility of diverse techniques under time series forecasting. The current research is dedicated to the performance of the prediction and applicability of classic models such as GARCH, EWMA, and Rolling Volatility. Simultaneously, these are being compared with the machine learning techniques of Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks to determine their relative strengths. The study further seeks to provide actionable insights into the advantages and limitations of these models, offering a nuanced understanding of their implications for individual and institutional investors alike. The current analysis thus answers some essential questions that concern financial analytics, such as how traditional models weigh against their ML counterparts, if market crashes could be predicted with the use of the reviewed methodologies, and what concrete benefits the volatility forecast offers by considering risk assessment and portfolio management.

The academic value of this dissertation is an interdisciplinary study that brings together finance, data analytics, and machine learning. With the extensiveness of the studies in volatility forecasting of financial econometrics, the introduction of ML to it has drawn a whole new, transformational dimension that still largely remains in a comparative study of the underexplored. A study like this was conducted, within strictness across both camps and added to a corpus of existing, increasing literatures related to financial forecasting-hence, turning it into substance. The study's alignment with my academic background further enhances its relevance, drawing upon my expertise in financial management, risk analysis, and programming to tackle a pressing and complex issue.

Another angle that has an equally strong drive for this research involves its practical implications: the accurate volatility forecast would bring a sea of change in decision-making, enhance risk management strategies, and optimize portfolio allocation for investors, portfolio managers, and risk analysts.

Automation and algorithmic trading further underscore the need to understand what these predictive models can and cannot do. This dissertation intends to develop valuable insights related to real-world applicability and performance evaluation of such models, which shall be useful to financial practitioners in an increasingly data-driven environment.

This study is divided into several structured sections. The introduction stipulates the focus of the study, its relevance, and its objectives developed in subsequent chapters. The literature review examines the extant volatility forecasting studies, both traditional and ML-based, while identifying gaps and opportunities. The methodology outlines the research design, data sources, and analytical techniques employed, followed by the results and discussion, which present a comparative analysis of the models and interpret their implications. The conclusion summarizes some of the main findings, considers the limitations of this study, and outlines further directions of research-as indeed it does in the completeness of volatility forecasting.

Literature Review

The earliest study on volatility forecasting was conducted by (Kuen and Hoong, 1992), who compared the naive method based on historical sample variance, the EWMA method, and the GARCH model to determine the most effective method for volatility forecasting in the Singapore stock market.

(Cumby et al., 1994) address the issue of volatility and correlation forecasting using EGARCH models with time variation in volatilities and correlations across US broad asset class returns. The emphasis of the study was out-of-sample forecasting performance, with correct forecasts of future volatility levels being particularly important.

In the light of this review, (Day et al. 1993) presented a comparison between different volatility forecasting techniques on crude oil futures prices, some GARCH family models, and implied volatilities derived from call options on crude oil futures. This paper examines various volatility forecasting methods to determine those which best forecast volatility in futures markets.

(Walsh et al. 1998) compared the naive approach, an improved extreme-value method (because of its ease of application), the ARCH/GARCH class of models, and the EWMA of volatility in volatility forecasting. (Poon and Granger, 2003) review 93 papers published over the last two decades with the aim of assessing volatility forecasting in financial markets. The focus is placed on forecasting rather than modeling; hence, light has been shed on issues such as critical evaluation, data frequency, extreme values, and measuring "actual" volatility. In a sequel study, (Poon and Granger, 2005) address practical forecasting issues and propose four models. The most important issue is the role of option-implied volatility, as opposed to time-series models, for effective forecasting. The literature underlines the crucial role that accurate forecasts have for good risk management and investment decisions in financial markets.

Different methods, such as GARCH models, implied volatilities, and time series forecasts, were studied and compared to find which one works better in volatility forecasting. Given that market conditions evolve dynamically and that financial markets are gradually becoming more complex, more studies should be carried out to develop better and more robust volatility forecasts.

Volatility forecasting has been a key concern in financial markets, especially after the failure of traditional forecasting models during the 2007–2008 Financial Crisis, which underestimated market risk and volatility. Since then, a great spate of theoretical developments and methodologies has come up to make more precise forecasts of volatility and assessments of market risk. Several works have tried different methods of volatility forecasting using a combination of machine learning techniques, sentiment analysis, and economic policy uncertainty. Plugins of machine learning techniques involve Gradient Descent Boosting, Random Forest, Support Vector Machine, and Artificial Neural Networks, among others, in which these algorithms are stacked to predict volatility in assets like the S&P 500.

The proposed hybridized stacked model based on an artificial neural network will improve the performance of volatility forecasting by combining the strengths of multiple machine learning algorithms.

Another line of investigation involves the examination of the influence of social media sentiment on volatility forecasting. For example, (Lehrer et al., 2021) assess how social media sentiment and model uncertainty can affect the performance of a volatility forecast. Sentiment analysis drawn from social media platforms is introduced into the models to capture public opinion and market sentiments on asset-price volatility.

Research has also focused on how economic policy uncertainty (EPU) influences the volatility of different assets. (Liu et al., 2021) investigate the relationship between EPU and European Union carbon futures price volatility by utilizing the GARCH-MIDAS model, evaluating different EPU indexes to see whether economic policy uncertainty can predict future volatility in financial markets. Apart from conventional financial risk factors, researchers have also tried to incorporate environmental, social, and governance (ESG) risks into volatility forecasting models. (Capelli et al., 2021) investigate the impact of integrating ESG risk alongside financial risk factors to enhance volatility forecast accuracy. By considering broader perspectives of risk factors, researchers strive for more complete and robust predictions of volatility.

Furthermore, volatility forecasting has lately adopted deep learning models such as DNN and LSTM. (Jia and Yang, 2021) illustrate, through an empirical approach, how employing a likelihood-based loss function when training a machine learning model—particularly DNN and LSTM—can be effective for forecasting stock index volatility. By applying deep learning methods, researchers aim to improve the predictive performance of volatility models. On the other hand, the role of uncertainty indexes in volatility forecasting has been studied even within the context of cryptocurrencies. Although your reference list has (Xia et al., 2023), here it appears as (Xia et al., 2022) in the text—this might be a mismatch. In either case, the authors discuss the impact of uncertainty indexes on Bitcoin volatility using the GARCH-MIDAS approach. Studies in volatility forecasting, by incorporating uncertainty indexes, aim to provide fresh perspectives on how this uncertainty interplays with asset-price volatility.

By integrating these various approaches—machine learning techniques, sentiment analysis, economic policy uncertainty, ESG risk integration, and deep learning models—researchers hope to arrive at better and more accurate volatility forecasts for various financial markets and asset classes.

Types of Volatility

Historically, there have been three measures of volatility that are widely used in research and practice for the analysis and forecast of volatility, namely, historical volatility, implied volatility, and realized volatility. Historical volatility, otherwise known as statistical volatility, refers to the dispersion of past movements in an asset's price over a predetermined period. It is computed with the standard deviation of its returns and thus has its roots essentially in a backward-looking measure of market volatility. For instance, one could estimate historical volatility through daily percentage fluctuations in an asset's price and further applying statistical formulas to establish volatility, usually annually for easier comparisons across assets and other periods. High historical volatility reflects large swings in price, while low values reflect more stable performance. However, because it is a backward-looking measure, historical volatility cannot predict what the future might hold regarding market conditions and hence is not effective when market disruption or fundamental change is prevalent.

Implied volatility, however, is forward-looking and is obtained from options contract prices. Implied volatility conveys the expectations of market participants about future volatility and is thus an important input to options pricing models, such as the formula by Black-Scholes. Unlike historical volatility, the implied one does not indicate the direction of movement but shows the size of the possible movements. As traders expect wider swings in prices, for example, when economic turmoil has taken over the market stage, options premiums are widening due to their higher implied volatility. This measure serves valuable insights into market sentiments but is definitely not without any limitations. It relies on market expectations, thus being susceptible to subjective interpretations and not always reflecting real market behavior.

Realized volatility, or alternatively, actual or observed volatility, refers to a measure of the actual changes in price an asset experiences over a chosen period. Unlike implied volatility, which is calculated from market perspectives, realized volatility relies on historical price data and looks at the actual outcomes. It is computed as a summation of daily returns across a certain evaluation period and after that annualized based on their standard deviation. For example, if a stock has huge intra-day price variabilities, such a stock will show great realized volatility reflective of high market uncertainty and riskiness. As clearly seen from these definitions, an advantage of the notion of realized volatility is its better clarity regarding past perspectives of market attitude and therefore realized volatility can be used as benchmark to check all predictions. However, the biggest drawback of using actual volatility is having the same shortfall as historic volatility: it also looks back into the past period.

Apart from that, short-run anomalies and noise may hide broader trends in data and hence are difficult to interpret.

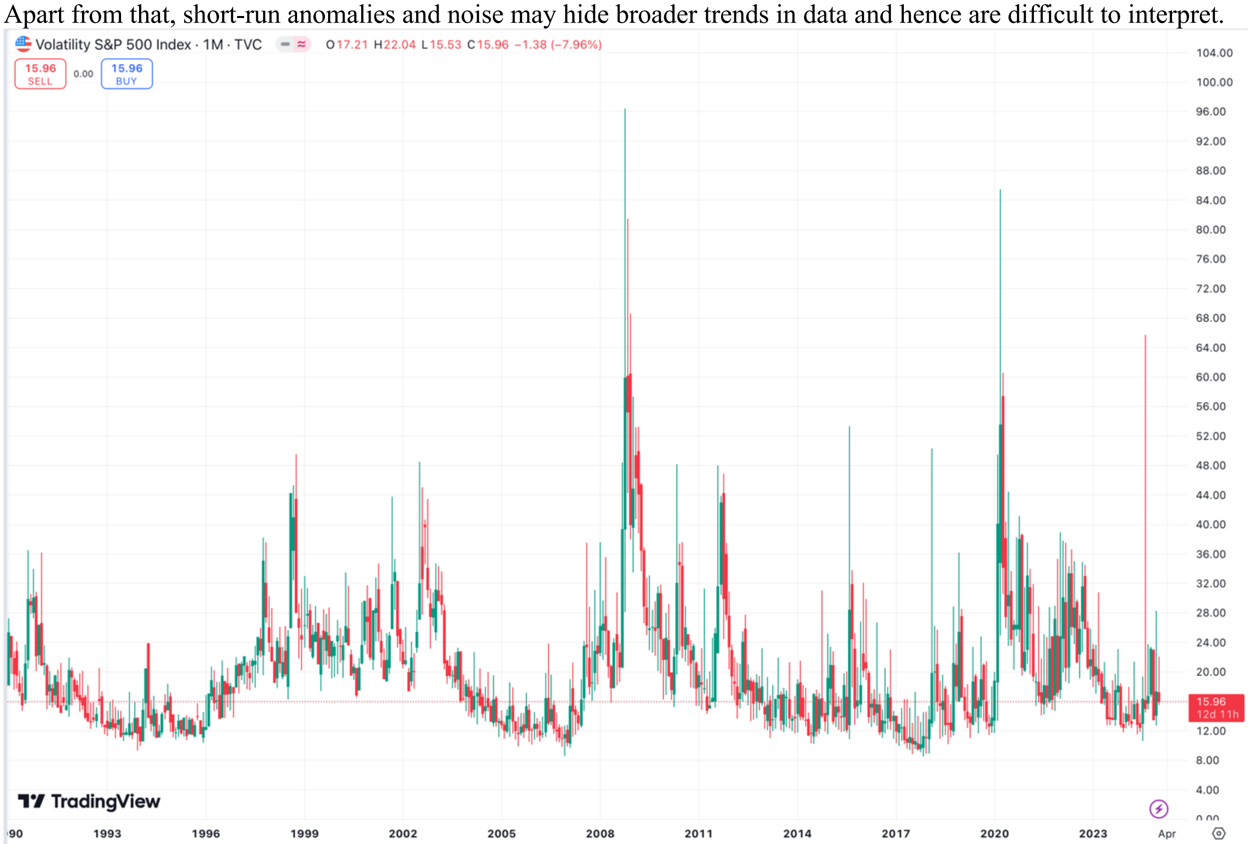

Figure 1: VIX

Complementing the above measures, in volatility analysis, the role of the Chicago Board Options Exchange Volatility Index is paramount, popularly known as a "fear index." The VIX calculates implied volatility from 30-day S&P 500 index options and offers a snapshot of expected price fluctuations over the following month. As the VIX is widely regarded as a barometer of market sentiment, it reflects the level of uncertainty and fear among investors. High readings indicate high market stress, while low readings reflect confidence and stability. The VIX has shown its

predictive capabilities during major financial events, such as the 2008 global financial crisis and the 2020 COVID-19 pandemic, when its levels surged, signaling extreme market volatility and uncertainty.
Despite its usefulness, there are some shortfalls to the VIX. First, it to a great extent reflects the conditions for the S&P 500 and therefore might not fully express volatility in other segments or regions of the market. Since the VIX depends on option pricing, it is going to be influenced by market liquidity and trading volumes; thus, its indications change when the market is not very active. All things considered, the VIX remains an extremely valuable tool for both traders and risk managers, yielding significant insight into the market that inspires the creation of financial products such as VIX futures and exchange-traded funds, which let investors trade volatility directly. By integrating these measures, volatility analysis offers a broad framework for the understanding of market behavior, assessing risks, and making financial decisions; hence, its critical role in both theoretical and practical applications in finance.

Forecasting Models

The modeling of financial market volatility has evolved significantly over the years, with traditional statistical models forming the backbone of early advancements. These models provide a structured approach to understanding and forecasting volatility, making them essential tools in financial econometrics. Among the most prominent models are the Autoregressive Conditional Heteroskedasticity (ARCH) model, its generalized form (GARCH), and alternative methods such as Historical Rolling Volatility and Exponentially Weighted Moving Average (EWMA). Each of these methods has contributed uniquely to the study of financial time series, offering distinct advantages and limitations in practical applications.

Autoregressive Conditional Heteroskedasticity (ARCH)

Introduced by Robert Engle in 1982, the ARCH model revolutionized financial econometrics by introducing a systematic way to model time-varying volatility. Before its introduction, most econometric models assumed constant variance, which failed to account for the phenomenon of volatility clustering—a common characteristic in financial markets where large price changes are followed by large changes and small price changes by similarly small changes. The ARCH model addresses this limitation by allowing variance to change dynamically over time based on past information, making it particularly suited for analyzing financial time series.

At its core, the ARCH model defines the conditional variance of a time series as a function of past squared residuals. The model comprises two primary components: the mean equation, representing the observed returns, and the variance equation, capturing the time-varying conditional variance. The variance equation incorporates a constant term and lagged squared residuals, where the number of lags determines the model's order. This structure enables the ARCH model to capture the persistence of volatility effectively.

1. **Mean Equation**:

$y_t = \mu + \epsilon_t,$

$$\epsilon_t \sim N(0, \sigma_t^2)$$

Here:

- y_t: Observed return at time t.
- μ: Mean of the series.
- ϵ_t: Error term or innovation, assumed to have a zero mean and variance σ_t^2.

2. **Variance Equation**:

$$\sigma_t^2 = \omega + \sum_{\{i=1\}}^{q} \alpha_i \epsilon_{\{t-i\}}^2$$

Where:

- σ_t^2 : Conditional variance at time t.
- $\omega > 0$: Constant term ensuring a positive variance.
- $\alpha_i \geq 0$: Coefficients representing the impact of past squared errors.
- q: Lag order, determining how many past periods influence the current variance.

One of the key strengths of the ARCH model is its ability to reflect volatility clustering, a hallmark of financial data. It also introduces time-varying variance into econometric models, moving beyond the static assumptions of earlier approaches. However, its reliance on a finite number of lags (q) poses challenges, as selecting the appropriate lag order is critical to balancing model complexity and predictive accuracy. Additionally, ARCH models are limited in their capacity to capture long-term volatility patterns, as they depend solely on past squared residuals.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

Building on the foundation of the ARCH model, Bollerslev introduced the GARCH model in 1986 to address some of its limitations. While the ARCH model effectively captures short-term volatility dynamics, it often requires a large number of lags to model persistent volatility, leading to computational inefficiencies and potential overfitting. The GARCH model overcomes these issues by incorporating lagged conditional variances into the variance equation, reducing the number of parameters needed while improving its ability to capture long-term dependencies.

The GARCH model generalizes the ARCH framework by combining past squared residuals and lagged variances in the conditional variance equation. This parsimonious design allows the GARCH model to account for the persistence of volatility over time, a feature often observed in financial markets. The inclusion of lagged variances ensures that the model can efficiently represent the memory of volatility without requiring excessive lags, as would be necessary in high-order ARCH models.

The GARCH (p, q) model generalizes ARCH by expressing the conditional variance σ_t^2 as a function of both past squared residuals and past variances:

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

Here:

- $\omega > 0$: A constant term ensuring a positive variance.
- α_i : Coefficients for the squared residuals(ϵ_t^2), capturing the impact of past shocks.
- β_j : Coefficients for lagged variances ($\sigma_{\{t-j\}}^2$), modeling the persistence of volatility over time.
- p: Number of lagged variances.
- q: Number of lagged squared residuals.

While the GARCH model retains many of the strengths of ARCH, such as capturing volatility clustering and time-varying variance, it introduces new capabilities in modeling the long memory of volatility. Despite these advancements, GARCH models have limitations. They assume symmetry in the impact of positive and negative shocks, which empirical evidence shows is often unrealistic. Negative shocks tend to have a greater effect on volatility, a phenomenon known as the leverage effect. Extensions such as Exponential GARCH (EGARCH) and Threshold GARCH (TGARCH) have been developed to address this asymmetry, providing further flexibility to model volatility dynamics.

Historical Rolling Volatility

Historical rolling volatility, or rolling standard deviation, is a straightforward yet widely used approach for tracking changes in volatility over time. It calculates the standard deviation of asset returns within a fixed rolling window, updating the calculation as the window advances. This dynamic method enables real-time assessment of an asset's risk profile, making it a valuable tool for practitioners seeking to monitor volatility trends.

The rolling standard deviation for a time series y_t is calculated as:

$$\sigma_{\{t-w\}} = \sqrt{\{}\frac{1}{w}\sum_{i=1}^{w}\left(y_{\{t-i\}} - \bar{y}_{\{t-w\}}\right)^2\}$$

Where:

- w: Window size (e.g., 10 days, 30 days).
- y_t : Observed value at time t.
- $\bar{y}_{\{t-w\}}$: Mean of the observations in the rolling window.

The mathematical framework for rolling volatility involves computing the standard deviation of returns over a specified window, which can vary in size depending on the desired focus on short-term or long-term volatility. This measure is particularly useful for analyzing market behavior during periods of financial stress, as sudden spikes in rolling volatility can signal increased uncertainty and risk. However, rolling volatility has its limitations. It assigns equal weight to all observations within the window, potentially overlooking the greater relevance of recent data. Additionally, the choice of window size can significantly impact results, with shorter windows being more sensitive to noise and longer windows potentially missing short-term changes.

Exponentially Weighted Moving Average (EWMA)

The Exponentially Weighted Moving Average (EWMA) method improves on rolling volatility by assigning exponentially decreasing weights to older observations. This approach ensures that recent data carries more significance, allowing the model to adapt quickly to sudden changes in market conditions. The conditional variance in EWMA is calculated using a smoothing parameter, typically set close to 1 (e.g., 0.94 as recommended by RiskMetrics), which controls the weight assigned to historical observations.

The conditional variance under EWMA is calculated as:

$$\sigma_t^2 = \lambda\, \sigma_{\{t-1\}}^2 + (1 - \lambda)\epsilon_t^2$$

Where:

- σ_t^2: Conditional variance at time t.
- λ: Smoothing parameter (e.g., 0.94, as recommended by RiskMetrics).
- ϵ_t^2 : Squared residual or return at time t.

EWMA is particularly suited for environments where rapid changes in volatility are frequent, as it emphasizes recent data while maintaining a smooth estimate of overall trends. Its simplicity and sensitivity to recent developments make it a popular choice in risk management applications, such as forecasting Value-at-Risk (VaR). However, EWMA lacks a strong theoretical foundation compared to ARCH and GARCH models and assumes a constant smoothing parameter, which may not reflect evolving market dynamics. Furthermore, excessive sensitivity to short-term volatility spikes can sometimes lead to overreaction in estimates.

Gradient Boosting

Gradient Boosting is a powerful machine learning technique that has gained significant traction in predictive modeling due to its accuracy and adaptability. It belongs to a class of ensemble methods, which combine multiple weak learners—typically decision trees—to create a stronger predictive model. By iteratively correcting the errors of existing models, Gradient Boosting achieves improved performance, making it particularly effective for regression and classification tasks. Its ability to capture complex patterns and relationships in data has made it a popular choice across diverse fields, including finance, healthcare, and marketing.

At its core, Gradient Boosting seeks to minimize a specified loss function by sequentially adding new models to an ensemble. These additional models are designed to predict the residuals, or errors, from the previous iteration, effectively refining the overall predictions. The iterative process begins with a simple model, often the mean value for regression tasks. Each subsequent model contributes incremental improvements, with its predictions weighted by a learning rate that controls the impact of individual models on the final output. This process continues for a predetermined number of iterations or until the model's performance converges. By optimizing the loss function through the gradient descent approach, Gradient Boosting ensures systematic improvements at each step, which is where its name originates.

Mathematically, the model at stage m is expressed as:

$$F_m(x) \ = F_{\{m-1\}}(x) \ + \gamma_m h_m(x)$$

Where:

- $F_m(x)$ is the updated model.
- $F_{\{m-1\}}(x)$ is the model from the previous iteration.
- γ_m is the learning rate, controlling the contribution of the new weak learner.
- $h_m(x)$ is the new weak learner trained to predict the residuals.

XGBoost: Enhancements to Gradient Boosting

XGBoost, or Extreme Gradient Boosting, is an advanced implementation of Gradient Boosting that introduces several enhancements for improved performance and efficiency. Developed by Tianqi Chen, XGBoost was specifically designed to address some of the computational challenges and limitations of traditional Gradient Boosting methods. Its innovative features include regularization techniques to prevent overfitting, parallel processing capabilities for faster computations, and mechanisms for handling missing data. These advancements have made XGBoost one of the most widely used algorithms in data-driven industries.

Regularization is a key strength of XGBoost, allowing it to penalize overly complex models through L1 (Lasso) and L2 (Ridge) penalties. This feature helps reduce the risk of overfitting while maintaining model flexibility. Additionally, XGBoost incorporates sparsity awareness, enabling it to handle missing or sparse data effectively, which is particularly valuable in real-world datasets. Its parallel processing capabilities further enhance its scalability, allowing it to process large datasets efficiently. Tree pruning and depth control are additional features that contribute to its robustness, ensuring that the model remains interpretable while avoiding overfitting.

The practical relevance of Gradient Boosting, and particularly XGBoost, extends across various domains. In financial markets, these methods are frequently used for forecasting stock prices, predicting market volatility, and assessing financial risks. Their predictive power also makes them valuable in credit scoring and fraud detection, where the ability to identify subtle patterns in complex datasets is crucial.

In the healthcare sector, Gradient Boosting is applied to predictive analytics, such as identifying patients at risk of readmission or predicting the likelihood of specific medical conditions. Similarly, in e-commerce and marketing, these models power recommendation systems, predict customer lifetime value, and assist in customer segmentation. Their adaptability to structured and unstructured data makes them equally suitable for tasks such as sentiment analysis and spam detection in natural language processing.

The application of Gradient Boosting in predictive maintenance within engineering and manufacturing demonstrates its versatility. By analyzing time-series data on equipment performance, these models can forecast potential failures, allowing businesses to minimize downtime and reduce maintenance costs. This broad applicability highlights the value of Gradient Boosting and XGBoost in solving complex business problems and optimizing decision-making processes.

Advantages and Limitations

The popularity of Gradient Boosting stems from its numerous advantages. Its high predictive accuracy results from the iterative process of correcting residuals, which enables the model to capture intricate patterns in data. This flexibility allows it to adapt to custom loss functions, making it suitable for a wide range of use cases. Furthermore, the ability of Gradient Boosting models to rank feature importance provides valuable insights for decision-making, helping businesses identify key drivers of outcomes.

However, these advantages come with certain challenges. Training Gradient Boosting models can be computationally intensive, particularly for large datasets and complex hyperparameter grids. Careful tuning of parameters, such as learning rate, tree depth, and regularization terms, is necessary to achieve optimal performance. Without proper tuning, the model may overfit, especially in noisy datasets. Additionally, Gradient Boosting models, while capable of providing feature importance metrics, are often considered “black-box” algorithms, as their overall behavior is less interpretable than simpler models like linear regression.

Hyperparameter Optimization for Improved Performance

To address these challenges, hyperparameter tuning is essential for optimizing Gradient Boosting models. XGBoost provides a wide range of hyperparameters, such as the number of boosting rounds, learning rate, tree depth, and regularization terms. Systematic tuning of these parameters can significantly enhance model performance, with tools like GridSearchCV offering an efficient way to identify the best combination of settings.

For instance, the learning rate controls the contribution of each tree to the final model, balancing between faster convergence and the risk of overshooting the optimal solution. Parameters such as max depth and regularization terms (L1 and L2) manage model complexity, helping to prevent overfitting. Subsampling techniques, which involve using a fraction of the dataset or features for each tree, introduce randomness that can further improve the model’s generalization ability.

Gradient Boosting and XGBoost represent significant advancements in predictive modeling, combining high accuracy with robust performance. Their ability to capture complex patterns and adapt to diverse datasets has made them indispensable tools in business and management. Whether forecasting financial markets, optimizing marketing strategies, or enhancing operational efficiency, these models provide actionable insights that drive data-informed decision-making. While their computational demands and reliance on hyperparameter tuning require careful consideration, the benefits they offer in addressing complex analytical challenges far outweigh these limitations. By leveraging these advanced techniques, businesses can gain a competitive edge in navigating today’s data-driven landscape.

LSTM

Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber in 1997, represent a significant advancement in recurrent neural networks (RNNs), offering a robust solution for modeling sequential data. Unlike traditional RNNs, which struggle to capture long-term dependencies due to issues like vanishing gradients, LSTMs leverage a sophisticated gating mechanism to selectively retain or discard information. This ability to manage long-term dependencies makes LSTMs particularly well-suited for tasks such as time-series forecasting, volatility prediction, and sentiment analysis based on financial news.

The distinguishing feature of LSTMs lies in their architectural design, which includes a cell state and a system of three gates—input, forget, and output gates—that collectively control the flow of information through the network. The cell state acts as a long-term memory store, preserving information over extended periods. The gating mechanisms dynamically decide which information to add, retain, or discard from the cell state, enabling the model to focus on relevant patterns in the data while filtering out noise.

The forget gate determines what portion of the previous cell state should be retained or discarded. Using a sigmoid activation function, it generates values between 0 and 1, with 0 indicating complete forgetting and 1 indicating full retention of information. The input gate then decides which new information to store in the cell state. A combination of sigmoid and hyperbolic tangent (tanh) activations governs this process, ensuring that only relevant features are added. Finally, the output gate decides what information from the updated cell state should be passed to the next time step or used as output. Together, these mechanisms allow LSTMs to model both short-term variations and long-term dependencies effectively, a feature that is particularly valuable in sequential data analysis.

In financial forecasting, LSTMs have emerged as a powerful tool due to their ability to model the temporal dependencies inherent in financial markets. Tasks such as stock price prediction, volatility estimation, and trading volume forecasting benefit from the LSTM's capability to capture both recent trends and historical patterns. Beyond finance, LSTMs have demonstrated their versatility in a variety of domains, including natural language processing (NLP), where they are used for sentiment analysis, machine translation, and text generation, and in healthcare, where they support predictive analytics based on patient history. LSTMs also find applications in speech recognition, video analysis, and music composition, further illustrating their adaptability to diverse sequential data problems.

The advantages of LSTMs stem from their unique ability to overcome the limitations of traditional RNNs. By addressing the vanishing gradient problem, LSTMs enable the modeling of long-term dependencies that are often crucial in sequential data. Their dynamic memory management, facilitated by gating mechanisms, enhances their flexibility and robustness, particularly in noisy datasets where irrelevant information can be filtered out effectively. Additionally, their capacity to handle variable-length sequences makes LSTMs a versatile tool across different domains.

Despite their strengths, LSTMs are not without challenges. Their computational complexity is significant, especially when applied to large datasets or long sequences. Training LSTMs is often resource-intensive and time-consuming due to their iterative and sequential nature. Furthermore, their performance is highly sensitive to hyperparameters, such as learning rate, number of hidden units, and dropout rates, necessitating careful tuning for optimal results. Another drawback is their "black-box" nature, which makes it difficult to interpret their decision-making processes, a common concern in complex machine learning models.

LSTMs represent a critical advancement in sequential data modeling, offering unparalleled capabilities in capturing both short-term and long-term dependencies. Their unique design and adaptability have made them indispensable in financial forecasting and a range of other applications. However, their computational demands and interpretability challenges must be addressed to fully leverage their potential. As the field of machine learning continues to evolve, LSTMs remain a cornerstone technique, particularly for tasks that require a deep understanding of temporal dynamics in data.

Market Context and Challenges

Financial markets are inherently volatile, shaped by macroeconomic events, geopolitical developments, and systemic risks that continuously influence asset prices. Understanding and forecasting this volatility is critical for effective risk management, portfolio optimization, and financial decision-making. However, the unpredictable nature of markets and the limitations of both traditional and machine learning models pose significant challenges, particularly during periods of extreme market turbulence.

Historical Perspective on Market Volatility

The volatility of financial markets is most evident during periods of economic crisis, where conventional forecasting models often struggle to account for extreme fluctuations. For instance, the 2008 Global Financial Crisis exposed the deficiencies of traditional models such as GARCH, which relies on assumptions of normal distribution and stationarity. During this period, high levels of asset correlation and market risk led to significant forecasting inaccuracies. Dynamic conditional correlation (DCC) models, as highlighted by Pesaran and Pesaran (2010), revealed shortcomings in traditional value-at-risk (VaR) diagnostics, further demonstrating the limitations of these methods in extreme scenarios.

Similarly, the COVID-19 pandemic in 2020 presented unique challenges to both traditional and machine learning models. The sudden and unprecedented disruptions caused by global lockdowns and economic uncertainty led to volatility levels far beyond what existing models could predict. While machine learning models showed potential by adapting to rapidly changing market conditions, challenges such as overfitting and the need for extensive, high-quality data persisted. These events underscore the need for forecasting models capable of dynamically adjusting to volatile and high-stress environments.

Challenges in Volatility Forecasting

One of the key challenges in forecasting market volatility is the occurrence of "Black Swan" events. Coined by Nassim Nicholas Taleb, these events are rare but highly impactful, such as the 1987 stock market crash or the COVID-19 pandemic. Traditional models often fail to forecast these tail risks because they rely heavily on historical data and assume stationarity, which does not hold during such extraordinary events. While machine learning models leverage non-linear patterns and broader datasets, their effectiveness is still constrained by the sparsity of data associated with rare events.

The increasing complexity and interconnectedness of global financial markets present another challenge. As globalization deepens, correlations among assets rise, particularly during crises when shocks propagate rapidly across markets. This interconnectedness amplifies systemic risks, as seen during the 2008 crisis, where traditional models were unable to account for the cascading effects of financial distress. Furthermore, the rise of algorithmic and high-frequency trading adds another layer of complexity, with microsecond-level decisions contributing to flash crashes and heightened volatility.

Data-related challenges further complicate volatility forecasting. High-frequency data, while valuable, often introduces significant noise, making it difficult to distinguish meaningful trends from random fluctuations. For machine learning models, this challenge is particularly acute, as these models require clean, structured datasets for effective training. In emerging markets, the lack of historical and granular data exacerbates these difficulties, limiting the ability of models to provide reliable predictions.

Traditional and machine learning models each have their limitations. While traditional models like GARCH are effective in stable market conditions, they often fail during periods of extreme volatility due to their reliance on linear assumptions and fixed distributional properties. Machine learning models, on the other hand, are more flexible but prone to overfitting, particularly in noisy datasets. Their "black-box" nature also raises concerns about interpretability, which is critical for trust and adoption in financial applications.

Key Volatility Events (2014–2024)

Several major market events between 2014 and 2024 exemplify the complexity of forecasting volatility and the challenges faced by traditional and machine learning approaches.

The EUR/CHF blow-up in 2015 stands out as a particularly dramatic episode. The Swiss National Bank's abrupt decision to remove the floor on the EUR/CHF exchange rate caused the currency pair to plummet from 1.20 to as low as 0.68 within hours. Volatility surged from near-zero levels to nearly 100% instantaneously, highlighting how unexpected central bank policy decisions can lead to extreme market reactions. Although the currency quickly regained most of its losses, volatility remained elevated for months, underscoring the difficulty of preparing for rare but impactful events.

The Brexit referendum in 2016 marked another period of extreme market volatility. The unexpected "Leave" vote led to an immediate drop of 8% in the GBP/USD exchange rate and a surge in realized volatility to over 46%. Although volatility began to normalize within a month, the intermittent effects of geopolitical uncertainty, including the Pound flash-crash in October 2016, kept volatility elevated. This event illustrated how political decisions can have far-reaching implications for market behavior and forecasting models.

The "Volpocalypse" of 2018, triggered by the rapid unwinding of low-volatility trades, exemplified the risks associated with speculative strategies tied to volatility. The VIX spiked to an intra-day high of 50, driven by leveraged positions in volatility-linked instruments. This event revealed the vulnerability of financial systems to amplified responses during periods of stress, reinforcing the need for robust risk management strategies.

The onset of the COVID-19 pandemic in early 2020 caused unprecedented market disruptions. Equity markets experienced sharp declines, and the VIX reached a record high of 82.69, far surpassing levels seen during the 2008 crisis. Simultaneously, oil prices turned negative for the first time in history, reflecting extreme uncertainty and supply chain disruptions. The pandemic highlighted the difficulty of forecasting volatility in global crises with no historical precedent, emphasizing the importance of models capable of adapting to such conditions.

Geopolitical and economic shifts from 2022 to 2024 further contributed to market instability. Russia's invasion of Ukraine disrupted energy markets, leading to heightened commodity volatility. Simultaneously, aggressive interest rate hikes by central banks to combat post-pandemic inflation added uncertainty to equity and bond markets. These developments underscore the complexity of forecasting in an environment influenced by interconnected geopolitical, economic, and structural factors.

Despite the challenges, advancements in data analytics and computational power provide promising avenues for improving volatility forecasting. The integration of alternative data sources, such as sentiment analysis from news and social media, can complement traditional financial indicators. Hybrid models that combine statistical techniques with machine learning approaches hold potential for leveraging the strengths of both methodologies. Additionally, the development of explainable AI frameworks can enhance trust and adoption by providing interpretable insights into model behavior.

The financial markets between 2014 and 2024 were marked by significant volatility events, each underscoring the limitations of existing forecasting models and the complexity of market dynamics. From the EUR/CHF blow-up and Brexit referendum to the COVID-19 pandemic and geopolitical tensions, these episodes provide a valuable framework for analyzing the effectiveness of volatility forecasting models. Addressing the multifaceted drivers of volatility requires models that are adaptable, interpretable, and capable of integrating diverse data sources, paving the way for more robust and reliable financial forecasting.

Summary for Literature Review

Forecasting volatility is highly relevant for financial analysis, enabling market participants to assess risks and optimize portfolio strategies effectively. While extensive research exists on various methods for volatility forecasting, the literature often focuses on either traditional statistical models like GARCH or modern machine learning approaches like LSTM. However, there is a noticeable gap in studies that directly compare the performance of these models across different asset classes. This dissertation aims to bridge this gap by conducting a comparative analysis of four distinct models: EWMA, GARCH, XGBoost, and LSTM.

The study focuses on three major stocks—AAPL, MSFT, and JPM—and three indices—S&P 500, NASDAQ 100, and NIFTY 50—spanning a decade of market data from 2014 to 2024. By employing these diverse models, the research evaluates their strengths and weaknesses in forecasting volatility under varying market conditions, including periods of heightened uncertainty like the COVID-19 pandemic and geopolitical crises such as the 2022 Russia-Ukraine conflict.

The motivation for this research lies in the absence of a comprehensive framework that delineates the comparative effectiveness of these models. Traditional methods like EWMA and GARCH, known for their simplicity and robustness in capturing historical volatility patterns, are often contrasted with machine learning models such as XGBoost and LSTM, which leverage data-driven techniques to account for complex, non-linear relationships. This thesis specifically investigates whether the increased computational complexity and adaptability of machine learning models translate into significantly better performance compared to traditional methods.

The findings aim to provide actionable insights for both academics and practitioners. By assessing metrics like Mean Squared Error (MSE), R-squared (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), the dissertation identifies the contexts in which each model excels. For instance, EWMA and GARCH may be more suitable for stable market conditions, while XGBoost and LSTM could demonstrate superior performance during volatile periods due to their ability to adapt to evolving data patterns.

This study's contribution lies in offering a clear comparative framework, enabling a deeper understanding of volatility forecasting methodologies. It not only informs future research but also equips market participants with knowledge to select the most appropriate model for specific market scenarios, thereby enhancing decision-making in risk management and investment strategies.

Methodology

The methodology employed in this research was meticulously designed to ensure robust and comprehensive forecasting of stock market volatility. By integrating traditional statistical models with advanced machine learning techniques, this study leverages the strengths of diverse modeling approaches to analyze complex financial time series data. The selection of Exponentially Weighted Moving Average (EWMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) models provides a balanced framework that captures both historical patterns and non-linear relationships in market behavior.

The data collection and preparation phases laid a strong foundation for the analysis, ensuring data completeness and enhancing its predictive power through feature engineering. Statistical models like EWMA and GARCH were employed to exploit their proven track record in financial econometrics, while machine learning models like XGBoost and LSTM were selected for their adaptability and capacity to uncover intricate market dynamics. Comprehensive evaluation metrics tailored to the unique characteristics of each model ensured rigorous assessment of their predictive performance. This methodological design offers a multifaceted perspective on volatility forecasting, bridging traditional and contemporary techniques to provide actionable insights.

Data Collection

The data for this research was sourced from the Bloomberg Terminal, a widely recognized platform for accessing comprehensive financial datasets. Bloomberg's reliability and robustness made it the optimal choice for this study, ensuring that the data used for analysis was both accurate and complete. The terminal's export-to-Excel functionality was utilized to extract data seamlessly across multiple features for the selected stocks, indices, and exchange-traded funds (ETFs). This process not only facilitated efficient data handling but also minimized the risk of errors during the acquisition phase.

Access to the Bloomberg Terminal was provided through Strathclyde Business School, which offers free access to its master's students. While alternative platforms, such as S&P Capital IQ, were initially considered, they were ultimately deemed unsuitable due to limitations in data completeness and quality. Bloomberg's datasets, verified through Python-based checks, exhibited no missing values or inconsistencies, further reinforcing its selection as the primary data source for this study.

The dataset covers the period from January 1, 2014, to December 15, 2024. This extended timeframe was deliberately chosen to capture long-term market behaviors and trends, providing a comprehensive basis for machine learning applications. The range includes significant economic events that introduced notable market volatility, such as the COVID-19 pandemic, Brexit, and U.S.-China trade tensions. These events offer critical insights into market dynamics during periods of uncertainty and stress, thereby enhancing the relevance and applicability of the dataset.

To ensure a diverse and representative analysis, the study focused on a combination of major market indices and leading stocks. The selected indices include the S&P 500, NASDAQ-100, and NIFTY-50, which collectively represent global and regional market trends. For additional comparison, the INDY-US index was included to evaluate the NIFTY-50 in U.S. dollar terms. Alongside the indices, three prominent stocks—Apple Inc., Microsoft, and JPMorgan Chase—were chosen for their leadership in the technology, finance, and large-cap sectors. The inclusion of these instruments provided a holistic view of market behavior across different geographies and sectors, with ETFs further enriching the analysis by capturing investor sentiment and fund performance relative to their underlying indices.

Table 1: Explanation of Features

Name	Explanation
OPEN	The price of a stock or index at the beginning of the trading session.
HIGH	The highest price reached by a stock or index during the trading session.
LOW	The lowest price reached by a stock or index during the trading session.
ADJ CLOSE	The closing price of a stock or index adjusted for corporate actions like splits or dividends.
1D CLOSE	The closing price of a stock or index at the end of the trading day.
VOLUME	The total number of shares or contracts traded during a specific period.
TURNOVER	The total monetary value of the trades executed during a specific period.
10 DAY VOLATILITY	The measure of price fluctuations over the past 10 days.
30 DAY VOLATILITY	The measure of price fluctuations over the past 30 days.
90 DAY VOLATILITY	The measure of price fluctuations over the past 90 days.
AVG BID-ASK SPREAD	The average difference between the highest price buyers are willing to pay and the lowest sellers will accept.
CURRENT MARKET CAP	The total market value of a company or index calculated using the latest stock price.
HISTORICAL MARKET CAP	The total market value of a company or index at previous points in time.
AVERAGE DIVIDEND YIELD	The average percentage of a company's share price returned as dividends to shareholders over time.
P/E Ratio	The ratio of a company's current share price to its earnings per share, reflecting valuation.
Basic EPS	The portion of a company's profit allocated to each outstanding share of common stock.

The dataset comprises several key features that are critical for modeling stock market volatility. Price-related variables, such as Open, High, Low, Adjusted Close, and 1-Day Close prices, were included to offer immediate insights into market behavior. Trading activity variables, such as Volume, Turnover, and Average Bid-Ask Spread, provide additional dimensions to understand market liquidity and activity levels. Volatility measures, including 10-day, 30-day, and 90-day realized volatility, serve as direct inputs for the predictive models, capturing market fluctuations over varying time horizons. Additionally, fundamental metrics, such as Current Market Cap, Historical Market Cap, Average Dividend Yield, Price-to-Earnings Ratio (P/E), and Basic Earnings Per Share (EPS), contribute a deeper understanding of the underlying economic conditions influencing market movements.

The primary target variables for this study were 10-day volatility measure. These were selected to capture short-term, medium-term, and long-term trends, providing a comprehensive view of market fluctuations. Short-term volatility reflects immediate market movements, often influenced by news or events, while medium-term trends capture broader market conditions over weeks. Long-term volatility, on the other hand, encompasses patterns that emerge over months, offering insights into the overarching stability or instability of the markets.

The rationale for selecting these features and target variables lies in their ability to represent market dynamics comprehensively. Price and trading activity variables offer real-time insights into market behavior, while volatility measures directly inform predictive models by quantifying market fluctuations. Fundamental metrics enrich the analysis by linking observed market movements to economic fundamentals, providing a more holistic perspective. Together, these carefully selected features form a robust foundation for the analysis of stock market volatility, aligning with the objectives of this dissertation.

Data Preparation

The preparation of data for this study was designed to ensure the highest quality inputs for the predictive models, with particular attention to preserving the integrity of the dataset and enhancing its analytical value through thoughtful feature engineering. This phase was critical in laying the groundwork for robust modeling and meaningful comparative analysis, encompassing data cleaning, feature engineering, and data splitting.

Data Cleaning

The dataset sourced from Bloomberg was of exceptional quality, with no missing values. This eliminated the need for imputation or additional data integrity checks related to incomplete records. Furthermore, the dataset included entries for every calendar day from January 1, 2014, to December 15, 2024, rather than being limited to trading days. This comprehensive inclusion of data ensured that no gaps existed in the time series, providing a consistent foundation for analysis.

Outliers, often candidates for removal in financial datasets, were retained in this study due to their significance in forecasting market volatility. In the context of volatility analysis, outliers are not anomalies to be dismissed but rather critical indicators of market behavior, particularly during periods of extreme stress. Retaining these data points allowed the study to focus on high-volatility periods, which are pivotal for understanding market risk and evaluating the performance of forecasting models under extreme conditions. The decision to include outliers ensured that the models would be tested comprehensively, providing a more realistic assessment of their utility in financial risk analysis.

Feature Engineering

To enhance the predictive power of the dataset, additional derived features were engineered. These features were selected to capture nuanced market behaviors and patterns, particularly to improve the performance of machine learning models like XGBoost. Key features included measures such as log returns, moving averages, and volatility bands, which provided critical insights into price trends, momentum, and volatility.

Log returns, calculated as the natural logarithm of the ratio of successive adjusted closing prices, offered a normalized measure of price movement that accounts for percentage changes over time. Moving averages, including simple moving averages (SMA) and exponential moving averages (EMA), were also introduced. SMAs, calculated over short- (5-day) and medium-term (20-day) windows, smoothed price fluctuations to reveal trends, while EMAs prioritized recent price changes to better capture current momentum.

Additional features such as the Relative Strength Index (RSI) and Bollinger Bands added valuable dimensions to the dataset. RSI measured the speed and change of price movements, identifying overbought or oversold conditions, while Bollinger Bands defined upper and lower thresholds based on standard deviations from a moving average, offering a clear indication of price volatility. These derived features were instrumental in enriching the dataset, particularly for machine learning models, which leverage feature importance to detect patterns in financial time series effectively.

Data Splitting

To evaluate the models' predictive performance, the dataset was split into training and testing subsets using an 80:20 ratio. This approach ensured sufficient data for training while reserving a portion for independent evaluation, adhering to standard practices in predictive modeling. The rationale behind this split was threefold. First, a larger training set enabled the models to learn underlying patterns more effectively, reducing the risk of overfitting. Second, a reserved testing set ensured an unbiased assessment of the models on unseen data, providing a realistic measure of generalization performance. Finally, maintaining a consistent split across all models—EWMA, GARCH, XGBoost, and LSTM—facilitated a fair comparison under identical conditions.

Python's train_test_split function was used to conduct the random splitting, with a fixed random state to ensure reproducibility. For time-series models like LSTM, the temporal sequence of the data was preserved to prevent data leakage, ensuring that the testing set consisted only of future data relative to the training set. This temporal structure was crucial for maintaining the integrity of predictions in a sequential context.

The data preparation process ensured that the dataset retained its completeness and relevance for volatility forecasting. By retaining outliers, the analysis maintained its focus on high-stress market conditions, preserving the ability to evaluate models under realistic scenarios. Feature engineering introduced critical variables that enriched the dataset, enhancing its value for machine learning applications. Finally, a rigorous data-splitting methodology ensured robust evaluation and fair comparison across all approaches. This meticulous preparation laid a strong foundation for the subsequent modeling and analysis phases, ensuring the reliability and robustness of the research findings.

Modelling Approach

The methodology for volatility forecasting employed a combination of traditional statistical models and advanced machine learning techniques. The selected models—Exponentially Weighted Moving Average (EWMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM)—were chosen to provide a comprehensive perspective on both established and modern approaches to volatility modeling. Each model was tailored to specific characteristics of the dataset and the dynamics of financial time series, ensuring robust and meaningful results.

Traditional Statistical Models

Exponentially Weighted Moving Average (EWMA)
The EWMA model was utilized to forecast 10-day market volatility due to its capacity to adapt to recent changes in financial data while preserving a memory of historical trends. The model prioritizes more recent observations by assigning exponentially decreasing weights to past data, effectively capturing the clustering effect of volatility often observed in financial markets.

The dataset was preprocessed by transforming adjusted closing prices into logarithmic returns, a step necessary to normalize price changes and prepare them for volatility modeling. Any missing values introduced by the calculation of log returns were removed to ensure consistency. The data was sorted by date and divided into training (80%) and testing (20%) subsets. This split allowed the model to be trained on historical data while its predictive performance was evaluated on unseen data, reflecting real-world forecasting scenarios.

The EWMA model was fitted to the training dataset to estimate historical volatility. Using Python's ewm method, the exponentially weighted standard deviation was computed with a smoothing factor derived from a span parameter (\lambda = 30). This span was chosen to balance the model's sensitivity to recent data against its incorporation of historical information. Volatility predictions for the test dataset were generated recursively, starting with the last estimated volatility from the training set. Each prediction was calculated using the equation:

$$\sigma_t = \sqrt{}\{(1 - \alpha) \cdot \sigma^2_{\{t-1\}} + \alpha \cdot r_t^2 \}$$

Where:

- σ_t : Predicted volatility at time t,
- $\sigma_{\{t-1\}}$: Volatility from the previous step,
- r_t : Log return at time t.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH)
The GARCH model was implemented to forecast financial market volatility by modeling its time-dependent and clustering behavior. GARCH's ability to estimate conditional variances made it particularly suited for time-series data like stock returns. The model was applied to historical adjusted closing prices, transformed into log returns to ensure stationarity and adherence to model assumptions.

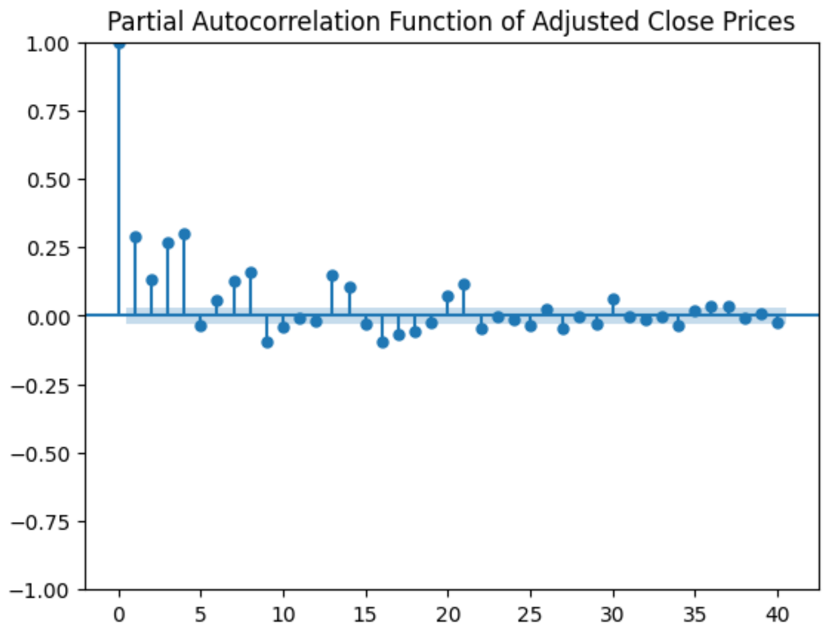

Figure 2: PACF Plot for S&P500

To determine the optimal order of the GARCH model, the squared returns were analyzed using a Partial Autocorrelation Function (PACF) plot. The PACF revealed significant lags at orders 1 and 1, leading to the selection of a GARCH (1,1) specification. The GARCH model was implemented using Python's arch library, with parameters estimated through maximum likelihood estimation. The model specification included one lag for both autoregressive (p = 1) and moving average (q = 1) terms in the conditional variance equation. The fitted model produced conditional volatility estimates that effectively captured the time-varying nature of volatility, as evidenced by plots of conditional volatility against actual returns.

The model also generated 30-step-ahead forecasts for conditional variance, calculated as:

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

Where:

- σ_t^2 : Conditional variance at time t,
- ω , α_i , β_j : Model coefficients,
- $r_{\{t-i\}}^2$: Squared returns.

The GARCH model's performance was benchmarked against a simple 10-day rolling mean of returns, which provided a baseline for comparison. GARCH consistently outperformed the benchmark across all evaluation metrics, particularly during periods of high volatility, demonstrating its robustness and practical utility in volatility modeling.

Machine Learning Models

Extreme Gradient Boosting (XGBoost)

XGBoost was employed to forecast 10-day market volatility, leveraging its efficiency and ability to capture non-linear relationships in structured data. The dataset included both raw features, such as adjusted closing prices and trading volumes, and engineered features like moving averages, Bollinger Bands, and log returns. These features provided rich insights into market trends, volatility, and momentum.

The data was split into training (80%) and testing (20%) subsets using random stratified sampling to maintain representative distributions of volatility patterns. GridSearchCV was then used to optimize XGBoost's hyperparameters, including the number of trees, tree depth, learning rate, subsampling, column sampling, and regularization terms. The final model configuration balanced complexity and generalization, minimizing variance between training and test performance metrics.

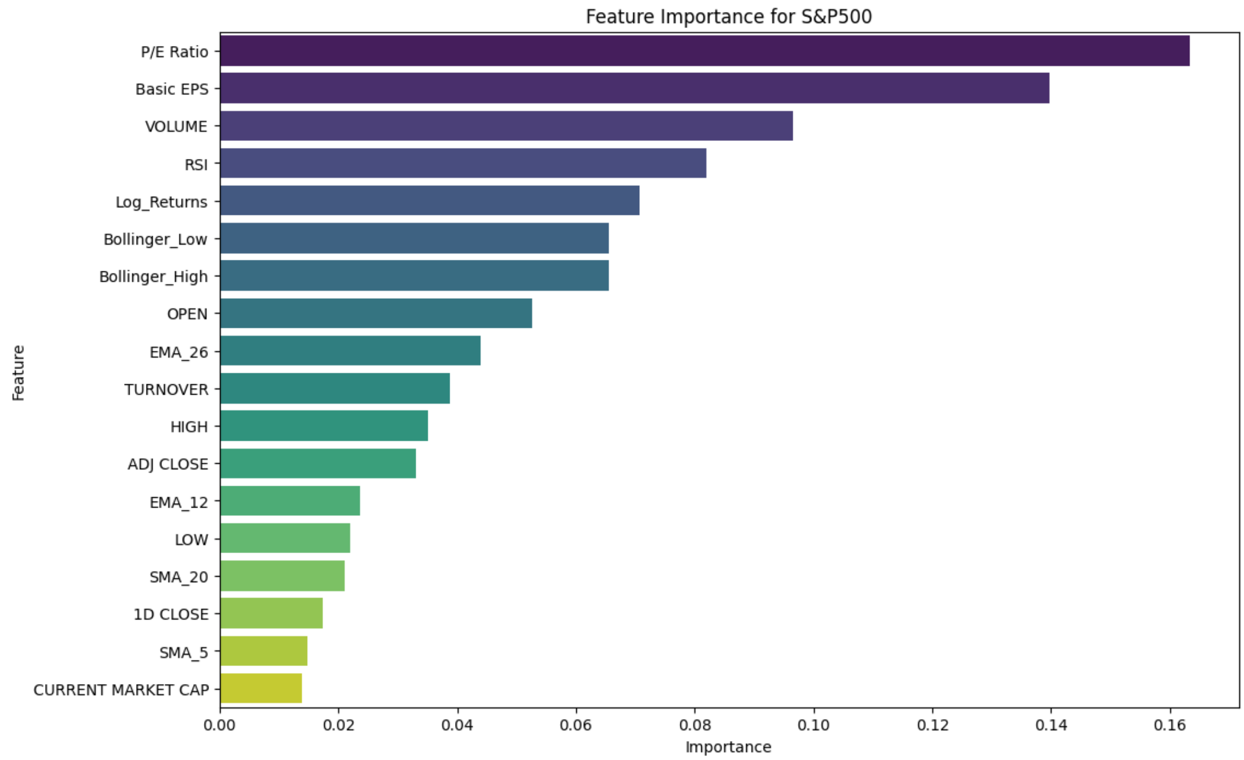

Figure 3: Feature Importance generated by XGBoost

Each bar in the chart shows the model's relative reliance on a given feature, with the total of all bar lengths summing to 1. This means XGBoost is essentially dividing up the 'importance pie' among the features according to how much each one contributes to reducing the model's prediction error. The taller bars indicate features that the model finds more influential, while shorter bars have a smaller (but still potentially meaningful) impact on the outcome. By looking at these proportions, we can see which variables are most critical for the model's predictive power.

XGBoost's scalability and feature importance analysis made it particularly effective in identifying key drivers of volatility. Its predictions provided actionable insights, emphasizing its relevance in financial forecasting.

Long Short-Term Memory (LSTM)

LSTM, a type of recurrent neural network, was selected for its ability to capture temporal dependencies in sequential data, making it ideal for financial time series forecasting. The dataset included features such as adjusted closing prices, trading volumes, market capitalization, and P/E ratios, with 10-day volatility as the target variable.

The data was scaled using a MinMaxScaler to normalize values between 0 and 1, preventing numerical instabilities during model training. A sliding window approach was used to create 3D input data, where each sample included a defined number of past observations to predict future volatility. The temporal order of the data was preserved to prevent leakage and ensure the integrity of predictions. In most deep learning frameworks (such as Keras), an LSTM layer expects a 3D input shaped as (samples, timesteps, features). The first dimension (samples) corresponds to how many sequences (or sliding windows) you have in your dataset, the second dimension (timesteps) corresponds to the length of each sequence, and the third dimension (features) is how many variables are being fed in at each time step.

If you have only one feature, you still need to include this as a third dimension of size 1—for example, (samples, timesteps, 1). This is why we often refer to LSTM inputs as "3D data," because it includes three dimensions: the number of samples, the time length of each sample, and the number of features measured at each time step.

To optimize the LSTM architecture, Keras Tuner was employed for hyperparameter tuning. The best configuration included two LSTM layers with 150 and 100 units, dropout rates of 0.2 and 0.3, respectively, and a learning rate of 0.0005. The final architecture used the Adam optimizer and minimized mean squared error (MSE) to ensure robust convergence. The model was trained for 50 epochs with a batch size of 32, achieving effective learning without significant overfitting. Each LSTM layer is a 'stage' of the model that learns to recognize patterns in the sequence data. Adding more layers helps the model pick up on increasingly complex relationships. 'Dropout' is a safeguard that randomly turns off parts of the model during training, preventing it from becoming too specialized in ways that don't generalize well to new data. The 'learning rate' controls how quickly the model updates its knowledge after seeing each batch of data. The 'Adam optimizer' is the algorithm that guides these updates so the model steadily improves its predictions. By measuring performance with mean squared error (MSE), the model tries to keep its predictions close to real values. 'Convergence' simply means the model reliably settles into a good level of accuracy over time, rather than wandering aimlessly or getting stuck in a poor fit.

LSTM's ability to process sequential data and retain long-term dependencies allowed it to respond effectively to market dynamics. The model's performance was evaluated against traditional statistical methods and XGBoost, showcasing the potential of deep learning in capturing complex temporal patterns in financial data.

Model Evaluation

To evaluate the predictive capabilities of the models employed in forecasting 10-day market volatility, a range of performance metrics was adopted, tailored to the specific characteristics of both traditional statistical models and machine learning frameworks. This approach ensured a comprehensive evaluation by addressing absolute, relative, and scale-dependent prediction errors while considering the inherent properties of the datasets and models. Each metric was chosen to highlight critical aspects of model performance, ranging from accuracy to sensitivity in high-volatility environments.

The traditional models, GARCH and EWMA, operate on the log-return scale, a transformation that normalizes price changes and reduces the influence of extreme values. For these models, **Mean Absolute Error (MAE)** and **Mean Squared Error (MSE)** were employed to quantify the average magnitude of errors. MAE, as a straightforward measure of average deviation, offered an interpretable metric that revealed the typical error magnitude between predicted and observed values. MSE, on the other hand, emphasized larger deviations by squaring the error terms, making it especially useful in identifying periods of heightened volatility where prediction errors tend to amplify. Complementing these, **Root Mean Squared Error (RMSE)** provided a scale-consistent measure of error, allowing for easier interpretation in the context of the original data range. These metrics collectively offered a solid foundation for assessing the overall accuracy of traditional models.

For the GARCH model, a specialized metric, **Heteroskedastic Mean Squared Error (HMSE)**, was included to account for the time-varying variance, or heteroskedasticity, inherent in financial data. GARCH explicitly models conditional variance, making HMSE a particularly relevant metric. By weighting squared errors based on predicted variances, HMSE ensured that errors during high-volatility periods were appropriately scaled, preventing undue penalization during periods of market stability or underweighting during periods of rapid fluctuations. This metric, widely adopted in econometric studies, added an additional layer of rigor to the evaluation of GARCH, ensuring its performance was accurately contextualized in dynamic market conditions.

The EWMA model, while simpler in design compared to GARCH, was evaluated using a similar set of metrics. In addition to MAE, MSE, and RMSE, **R-squared (R^2)** was introduced to measure the proportion of variance in observed volatility explained by the model. The high R^2 values observed for EWMA across different datasets underscore its strength in capturing general volatility trends, despite its limitations in adapting to rapid market changes. This aligns with EWMA's assumption of volatility clustering, which allows it to perform well in steady market conditions but limits its responsiveness during extreme events.

In contrast, machine learning models such as XGBoost and LSTM were evaluated using an extended set of metrics, reflecting the differences in their operational scale and model architecture. Both XGBoost and LSTM operated on the raw scale of volatility data, preserving the original range and magnitude of fluctuations. This introduced greater sensitivity to extreme values, necessitating the use of **Mean Absolute Percentage Error (MAPE)** alongside MAE, MSE, and RMSE. MAPE offered a scale-independent measure by calculating the average percentage error relative to the observed values, making it particularly suitable for comparing performance across datasets with varying volatility levels. For instance, the lower MAPE values observed for XGBoost highlight its robustness in providing consistent predictions across different stocks and indices, even during volatile periods.

The inclusion of R^2 further complemented the evaluation of machine learning models, offering insights into their ability to explain variance in the target variable. XGBoost demonstrated consistently high R^2 values across datasets, reflecting its ability to model complex, non-linear relationships between input features and volatility. In comparison, LSTM, while still effective, displayed slightly lower R^2 scores and higher MAPE, particularly during periods of high volatility. These results underscore the challenges LSTM faces in capturing abrupt changes in financial time series, despite its strength in learning temporal dependencies.

The significant differences observed between traditional and machine learning models can be partially attributed to the scale of operation. Traditional models like EWMA and GARCH, operating on log returns, normalized volatility and dampened extreme values, resulting in more stable predictions. In contrast, machine learning models retained the raw data scale, amplifying the influence of outliers and extreme market events. This fundamental difference in data processing highlights the importance of selecting appropriate metrics, such as HMSE for GARCH and MAPE for machine learning models, to align evaluations with the specific properties of each model.

To ensure consistency and fairness in evaluation, all models were tested using the same 80:20 train-test split and were benchmarked against the actual volatility values in the test set. The combination of absolute metrics like MAE and RMSE, relative metrics like MAPE, and variance-explaining metrics like R^2 provided a holistic framework for performance assessment. This approach not only highlighted the accuracy of predictions but also contextualized performance across varying levels of market volatility.

Businesses across various industries rely on regression metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and others to quantify how close their model predictions are to actual outcomes and to tie that accuracy back to real financial or operational impact. In e-commerce and retail, for instance, MSE or RMSE is often closely tracked to catch large forecast errors that could lead to overstock or stock-outs, both of which carry significant costs. Finance teams frequently prefer percentage-based metrics such as MAPE (Mean Absolute Percentage Error) because it translates easily into proportional financial risk, helping them justify budget allocations or revenue forecasts. Manufacturing and supply chain organizations also keep a close eye on MSE or RMSE to ensure that resource planning and capacity decisions are optimally balanced and do not incur waste or production bottlenecks. Healthcare and insurance companies, where even a single large deviation can be critical, often favor metrics that heavily penalize outliers—again pointing to MSE or RMSE—because such errors might translate into severe costs or, in the case of healthcare, directly affect patient outcomes. In the digital marketing space, MAE (Mean Absolute Error) can be more interpretable and is often useful for understanding the typical magnitude by which predictions deviate in everyday terms, such as the average dollars or leads off-target. By embedding these metrics in dashboards, setting thresholds (e.g., keeping RMSE under a certain level), and continuously monitoring performance, companies can ensure that predictive models deliver tangible value and that stakeholders maintain confidence in forecasts guiding strategic decisions.

The selected metrics provided a comprehensive lens through which the performance of EWMA, GARCH, XGBoost, and LSTM was evaluated. The inclusion of HMSE for GARCH underscored its suitability for high-volatility conditions, while MAPE and R^2 effectively captured the strengths of machine learning models in modeling broader volatility patterns. This rigorous evaluation framework ensures that the models' predictive capabilities are interpreted accurately, setting the stage for a detailed comparative analysis in the Results and Discussion section.

Tools and Techniques

The analysis was conducted using a combination of tools and libraries to ensure efficient data processing, modeling, and evaluation. **Microsoft Excel** was used for initial data collection, specifically to extract historical stock data from Bloomberg Terminal, ensuring consistency and reliability in the raw dataset. For all subsequent tasks, **Python** was the primary programming language due to its vast ecosystem of libraries that simplify complex data operations and ensure reproducibility.

Key libraries and packages used include **pandas** and **numpy** for data manipulation and numerical computations. These libraries facilitated efficient handling of large datasets, allowing for smooth integration of operations like calculating log returns, normalizing data, and managing missing values. Visualization was an integral part of the analysis, with **matplotlib** and **seaborn** used to generate insightful plots that highlighted patterns in the data and displayed model performance effectively.

For modeling, specialized libraries like **statsmodels** and **arch** were used for implementing traditional statistical models such as EWMA and GARCH, enabling precise calculations of conditional variances and time-series patterns. Machine learning frameworks like **xgboost** and **tensorflow** were leveraged for advanced predictive models. **XGBoost**, with its exceptional handling of structured data, provided fast and accurate predictions, while **TensorFlow** powered the implementation of LSTM, allowing for sequential data modeling. To optimize the machine learning models, tools like **GridSearchCV** for XGBoost and **Keras Tuner** for LSTM were employed, streamlining hyperparameter tuning for maximum predictive accuracy.

A dedicated **virtual environment** was created on the system to install the required libraries. This ensured a clean and conflict-free workspace, preventing compatibility issues that might arise when managing multiple dependencies.

The structured workflow—spanning data collection, preprocessing, feature engineering, model training, and evaluation—was efficiently executed using these tools, ensuring both accuracy and scalability. These techniques underscore the adaptability and robustness of Python's ecosystem in conducting advanced financial analyses.

Summary for Methodology

This research employed a rigorous methodology, integrating traditional statistical and machine learning models to forecast 10-day market volatility. Data collection was conducted using Bloomberg Terminal, ensuring high-quality and comprehensive datasets covering major indices and stocks from 2014 to 2024. The dataset was carefully prepared, including feature engineering to derive log returns, moving averages, and volatility measures, and data splitting into training and testing subsets to maintain integrity and prevent leakage.

Traditional models like EWMA and GARCH leveraged their econometric foundations to capture volatility clustering and time-varying variance. Machine learning models, including XGBoost and LSTM, utilized engineered features and advanced hyperparameter tuning to enhance predictive accuracy. EWMA and GARCH operated on normalized log returns, ensuring stability in predictions, while XGBoost and LSTM retained the raw data scale, capturing complex market dynamics.

Model evaluation employed diverse metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2), tailored to each model's scale and structure. GARCH's performance was further assessed using Heteroskedastic Mean Squared Error (HMSE), while MAPE provided insights into the scale-independent accuracy of machine learning models.

Results and Discussions

The methodology was applied to a comprehensive dataset encompassing various indices and stocks, including NASDAQ-100, NIFTY-50, and leading companies such as Apple Inc., Microsoft, and JPMorgan Chase. However, this section focuses exclusively on the S&P 500 index. The decision to prioritize the S&P 500 stems from its status as a benchmark for the U.S. equity market, representing a broad cross-section of the economy, and its significant global relevance in financial analysis. Additionally, the other indices and stocks analyzed exhibited similar patterns and trends to those observed in the S&P 500. As such, focusing on the S&P 500 ensures clarity and conciseness while maintaining the broader applicability of the findings. While detailed discussions for other indices and stocks are not included here due to word count limitations, their results were calculated as part of the study and are consistent with the patterns observed in the S&P 500.

GARCH

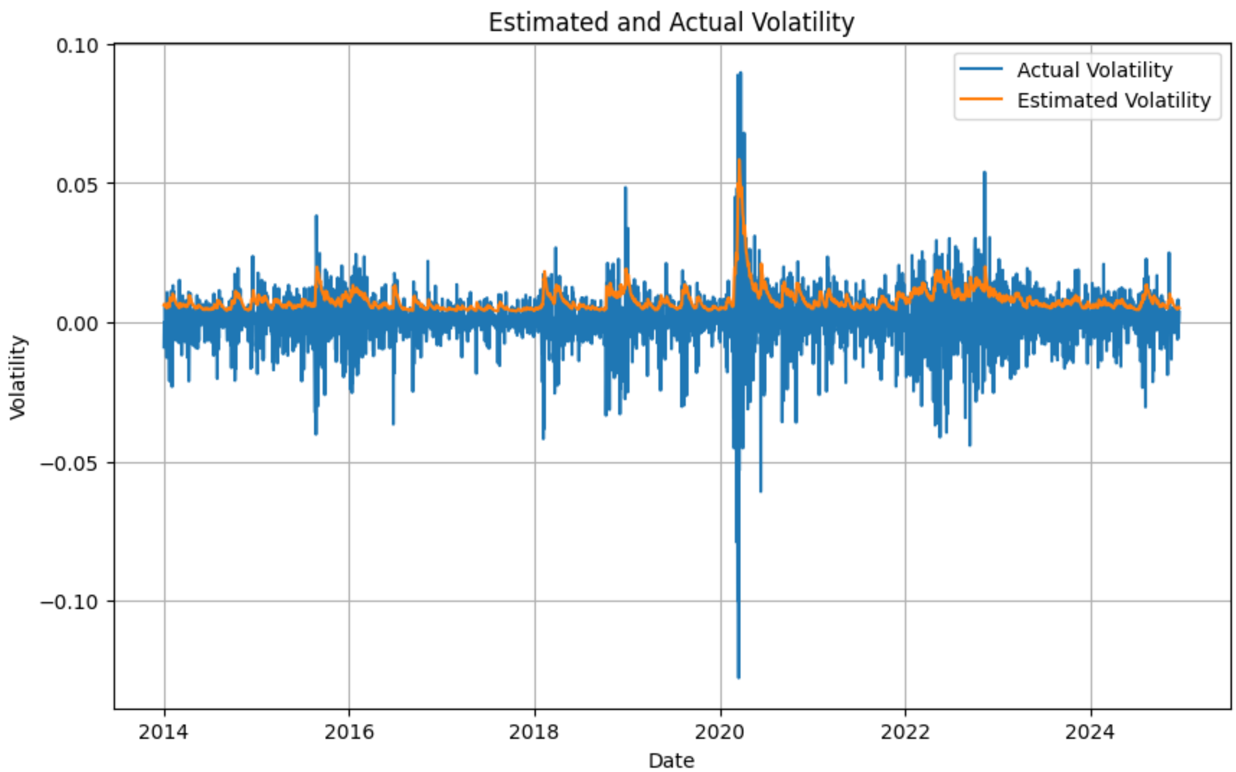

Figure 4: GARCH Volatility Forecast vs Actual

The performance of the GARCH model in forecasting the 10-day volatility of the S&P 500 index was assessed using several key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Heteroskedasticity-Adjusted Mean Squared Error (HMSE). For the S&P 500, the GARCH model achieved an MAE of 0.00897, an RMSE of 0.01269, and an HMSE of 0.05478. These results indicate the model's ability to capture time-varying volatility and its clustering behavior effectively.

Table 2: GARCH Results

S.No	Name	MAE	MSE	RMSE	HMSE	MODEL
1	MSFT	0.014765	0.000380	0.019503	0.043163	1,1
2	AAPL	0.015884	0.000412	0.020301	0.047661	1,1
3	JPM	0.014765	0.000380	0.019503	0.043163	1,1
4	S&P500	0.008951	0.000161	0.012674	0.054780	1,1
5	NASDAQ100	0.011518	0.000243	0.015590	0.037200	1,1
6	NIFTY50	0.008777	0.000138	0.011730	0.046027	1,1

A comparative analysis was conducted between the GARCH model and a simpler benchmark model, which employed a rolling mean approach to predict volatility. The rolling mean model, designed to calculate a moving average of returns over a 10-day window, yielded an MAE of 0.00521, an RMSE of 0.00866, and an HMSE of 0.00906 for the S&P 500. When compared, the GARCH model demonstrated higher errors across all metrics. Specifically, the GARCH model's MAE was 72% higher, its RMSE was 46% higher, and its HMSE was approximately 5.7 times worse than the rolling mean model. These discrepancies suggest that while GARCH captures intricate volatility patterns, its complexity may introduce overfitting or prevent it from adapting rapidly to sudden market shifts.

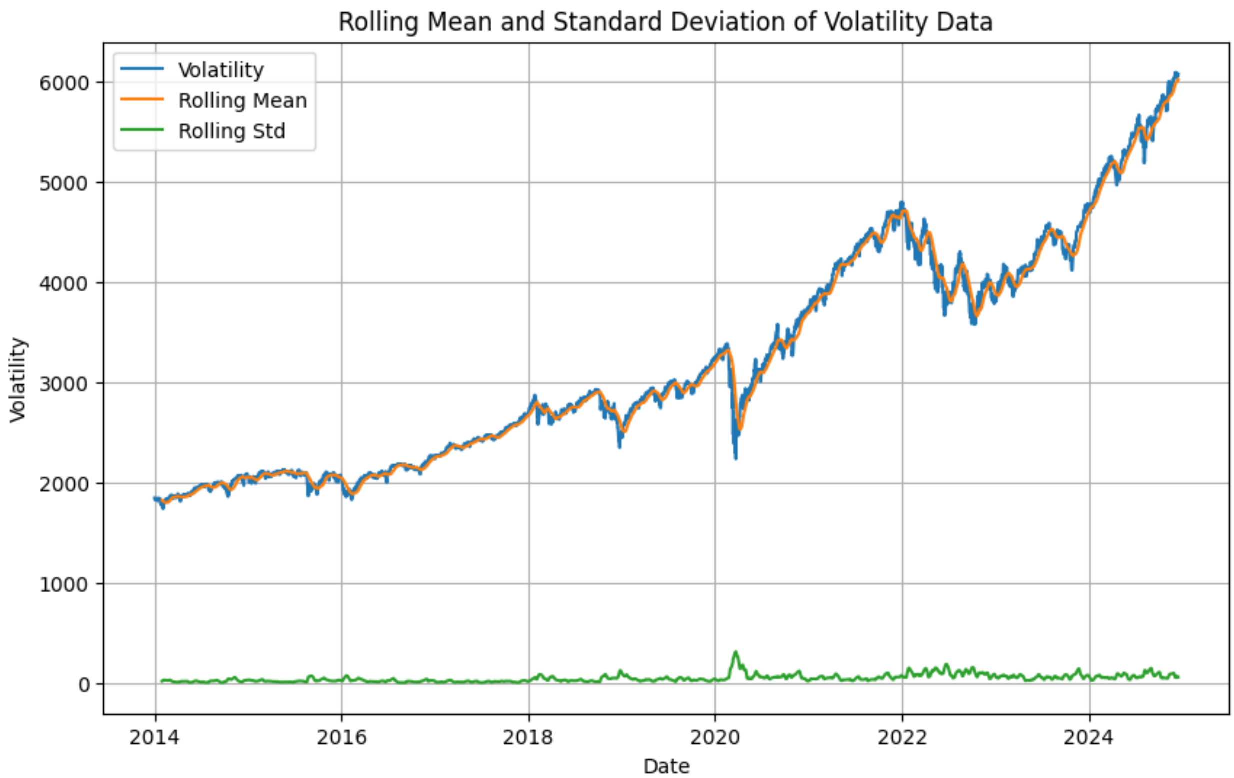

Figure 5: Rolling Mean vs Rolling Standard Deviation S&P500

The benchmark rolling mean model's simplicity provides certain advantages. By smoothing daily fluctuations in returns and offering a steady estimate of volatility, it adapts effectively to gradual market changes. However, the model's inherent assumption of equal weighting for all past observations within the rolling window underrepresents recent market dynamics and fails to account for volatility clustering. GARCH, in contrast, explicitly models conditional variance and volatility clustering, enabling it to capture the temporal dependencies that are characteristic of financial markets.

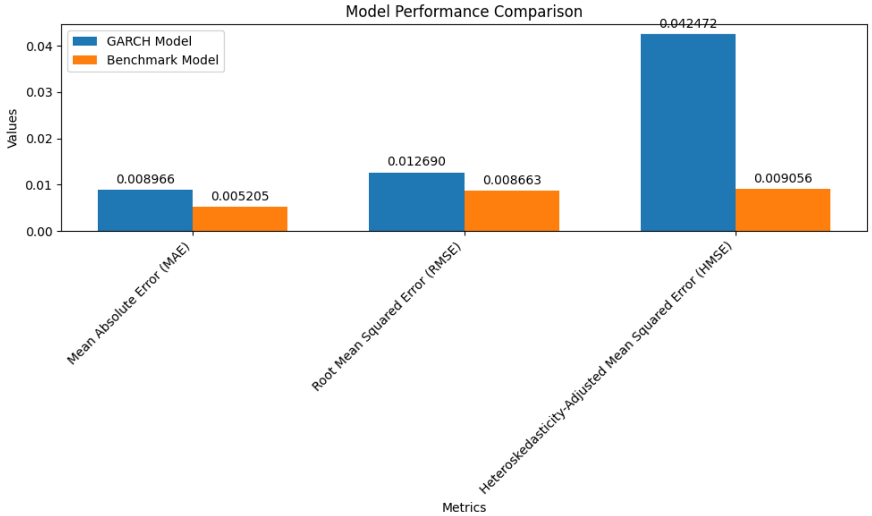

Figure 6: Baseline Model Comparison for GARCH

The significantly higher HMSE for the GARCH model reflects its challenge in adequately scaling errors during periods of extreme volatility, which are more pronounced in financial time series. The rolling mean model, despite its limitations in modeling volatility clustering, benefits from a more straightforward structure that avoids these issues. This result highlights a fundamental trade-off between the simplicity of baseline models and the sophistication of advanced techniques like GARCH.

Overall, the GARCH model provided valuable insights into the dynamic nature of S&P 500 volatility but demonstrated weaknesses in comparative error metrics. These results underline the importance of balancing complexity and adaptability in volatility forecasting and set the stage for further exploration of alternative modeling approaches such as machine learning techniques.

EWMA

Table 3: EWMA Results

S.No	Name	R^2	MSE	RMSE	MAE
1	MSFT	0.88450571	2.14E-06	0.0014631	0.00095871
2	AAPL	0.83710719	2.94E-06	0.00171389	0.00106328
3	JPM	0.84633956	3.05E-06	0.00174681	0.00117785
4	S&P500	0.8112929	1.25E-06	0.00111838	0.00063309
5	NASDAQ100	0.84049227	1.64E-06	0.00127945	0.000752
6	NIFTY50	0.79486391	8.51E-07	0.0009226	0.00054276

The performance of the Exponentially Weighted Moving Average (EWMA) model in forecasting the 10-day volatility of the S&P 500 was evaluated using key metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). For the S&P 500, the model achieved an MAE of 0.00063, an MSE of $1.25 \times 10^{\{-6\}}$, an RMSE of 0.00112, and an R^2 value of 0.811. These results demonstrate the EWMA model's ability to explain 81.1% of the variance in actual volatility, highlighting its strength in capturing general volatility patterns.

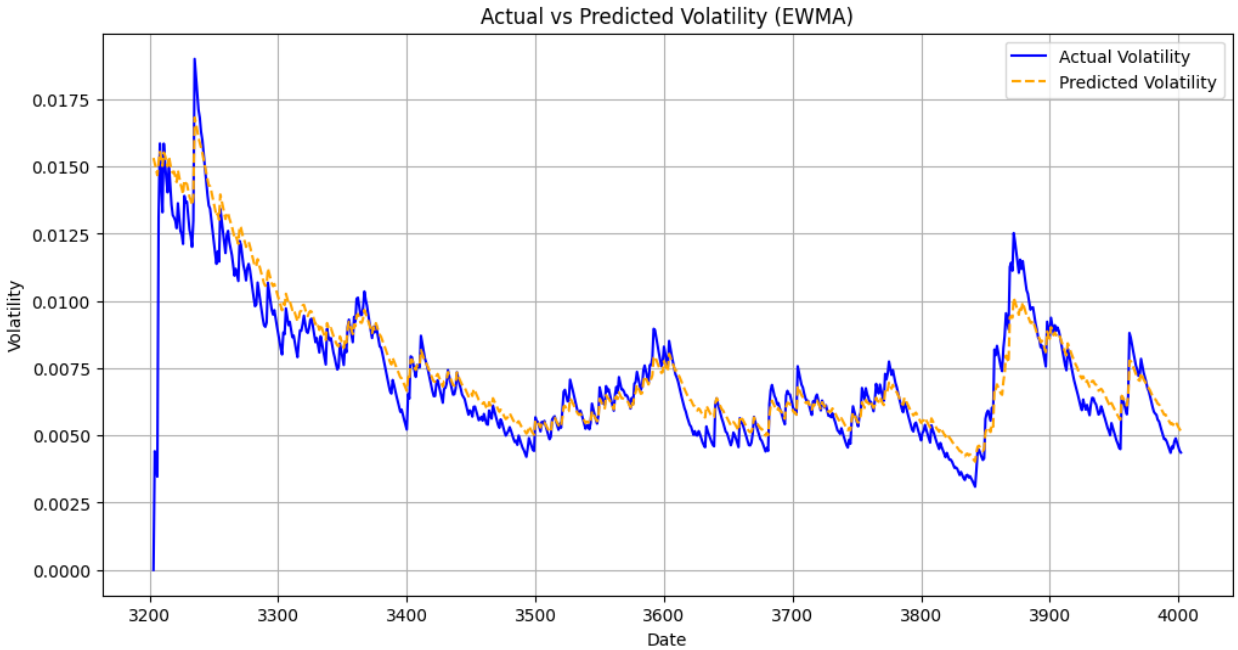

Figure 7: EWMA Volatility Forecast for S&P500

When benchmarked against two baseline models—a rolling standard deviation model and a constant volatility model—the EWMA model showed mixed performance. Compared to the rolling standard deviation model, which achieved an MSE of $4.67 \times 10^{\{-7\}}$, the EWMA model's MSE was approximately 2.7 times worse. This indicates that while EWMA is effective at capturing market trends, its weighting mechanism may not adapt as quickly to sudden changes in market dynamics as the rolling standard deviation baseline. However, when compared to the constant volatility model, which had an MSE of $1.18 \times 10^{\{-5\}}$, EWMA significantly outperformed, with an error approximately 9.4 times lower, emphasizing its superiority in modeling real-time fluctuations over a fixed volatility estimate.

The rolling standard deviation baseline calculates historical volatility over a fixed 30-day window, dynamically adapting to recent market changes. This method's simplicity allows it to track short-term market trends effectively. However, it assigns equal weights to all observations within the window, potentially underestimating the importance of more recent data that may better reflect current volatility conditions.

In contrast, the constant volatility model assumes stable market conditions by using the overall standard deviation of log returns as a fixed estimate. While computationally efficient, this approach fails to account for time-varying volatility, rendering it less effective in capturing sudden shifts or clustering behavior in financial time series.

Although the EWMA model's performance fell short of the rolling standard deviation baseline in terms of MSE, its ability to explain a significant portion of volatility variance (as reflected by its high R^2 value) underscores its effectiveness in modeling general volatility patterns. The exponential weighting mechanism allows EWMA to prioritize recent data, balancing responsiveness with the integration of historical trends. This makes it a valuable tool for financial forecasting, particularly in environments where capturing long-term patterns is as important as adapting to short-term fluctuations.

In summary, the EWMA model provides a nuanced approach to volatility forecasting, outperforming the constant volatility baseline by a wide margin and holding its own against the rolling standard deviation baseline despite its limitations in rapid market adjustments. These results highlight its utility as a balance between computational simplicity and dynamic responsiveness, particularly for long-term market analysis.

XGBoost

Table 4: XGBoost Results

S.No	Name	R^2	MSE	RMSE	MAPE
1	MSFT	0.83	25.05	5.01	19.18%
2	AAPL	0.84	25.9	5.09	17.91%
3	JPM	0.89	16.41	4.05	17.00%
4	S&P500	0.88	9.23	3.04	18.20%
5	NASDAQ100	0.88	14.07	3.75	17.98%
6	NIFTY50	0.83	10.54	3.25	16.98%

The XGBoost model demonstrated robust predictive performance in forecasting the 10-day volatility of the S&P 500 index. The model achieved an R^2 score of 0.88, an MSE of 9.23, an RMSE of 3.04, and a Mean Absolute Percentage Error (MAPE) of 18.20%. These metrics underscore the model's ability to capture 88% of the variance in the observed volatility, signifying its efficacy in modeling complex patterns and relationships in the dataset.

When comparing the model's performance to a baseline model, which simply predicts the mean of the target variable across all observations, the XGBoost model exhibited significantly better results. The baseline model yielded an MSE of 79.75, substantially higher than XGBoost's MSE of 9.23. This comparison indicates that XGBoost achieved nearly ten times lower error than the baseline, emphasizing its advanced capability to model volatility dynamics beyond simple averages.

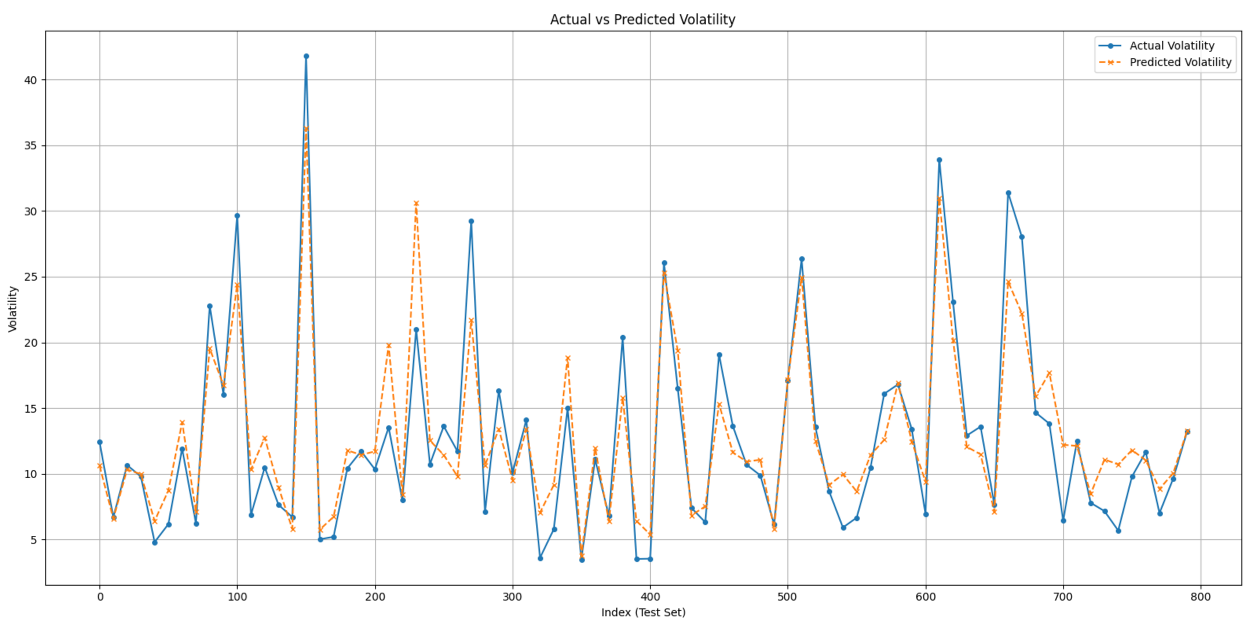

Figure 8: Predicted Vs Actual Volatility for S&P500 by XGBoost

The R^2 score of 0.88 further highlights the effectiveness of XGBoost in explaining the variability in the test dataset. In contrast, the baseline model operates under the assumption that all variance in the target variable is due to deviations from the mean, leading to significantly higher errors. By incorporating engineered features such as moving averages, Bollinger Bands, and log returns, the XGBoost model was able to capture temporal trends and nuanced market dynamics that the baseline model inherently overlooks.

A key observation is that the XGBoost model's MSE of 9.23 is significantly smaller than the variance of the target variable, which stands at 79.85. This result indicates that the model captures the variability in 10-day volatility effectively, outperforming simpler approaches. Furthermore, its relatively low MAPE of 18.20% demonstrates strong scale-independent accuracy, making it well-suited for forecasting purposes.

The baseline model, while computationally efficient and providing a straightforward reference point, lacks the sophistication to account for temporal trends or clustering effects in financial time series data. Its reliance on a single constant prediction (the mean) limits its practical application in forecasting scenarios. The substantial gap in MSE between the baseline and XGBoost reinforces the added value of employing advanced machine learning techniques.

In conclusion, the XGBoost model's ability to outperform the baseline by a significant margin, coupled with its strong R^2 score and low MAPE, highlights its utility in capturing and predicting complex volatility patterns. These results underscore the potential of machine learning models to enhance forecasting accuracy in financial markets, particularly when leveraging thoughtfully engineered features and robust evaluation frameworks.

LSTM

Table 5: LSTM Results

S.No	Name	R^2	MSE	RMSE	MAPE
1	MSFT	0.65	61.68	7.85	30.07%
2	AAPL	0.70	53.68	7.33	26.91%
3	JPM	0.76	47.39	6.88	24.77%
4	S&P500	0.72	29.80	5.46	36.35%
5	NASDAQ100	0.72	37.32	6.11	31.84%
6	NIFTY50	0.65	34.84	5.90	27.89%

The performance of the Long Short-Term Memory (LSTM) model in forecasting 10-day volatility for the S&P 500 index demonstrated mixed results, highlighting its ability to capture temporal dependencies while also revealing some challenges with accuracy. The LSTM model achieved an R^2 score of 0.72, an MSE of 29.80, an RMSE of 5.46, and a MAPE of 36.35%. While these metrics indicate that the model explains 72% of the variance in observed volatility, the relatively high MAPE suggests that the model struggles with certain aspects of volatility prediction, particularly during periods of extreme market fluctuations.

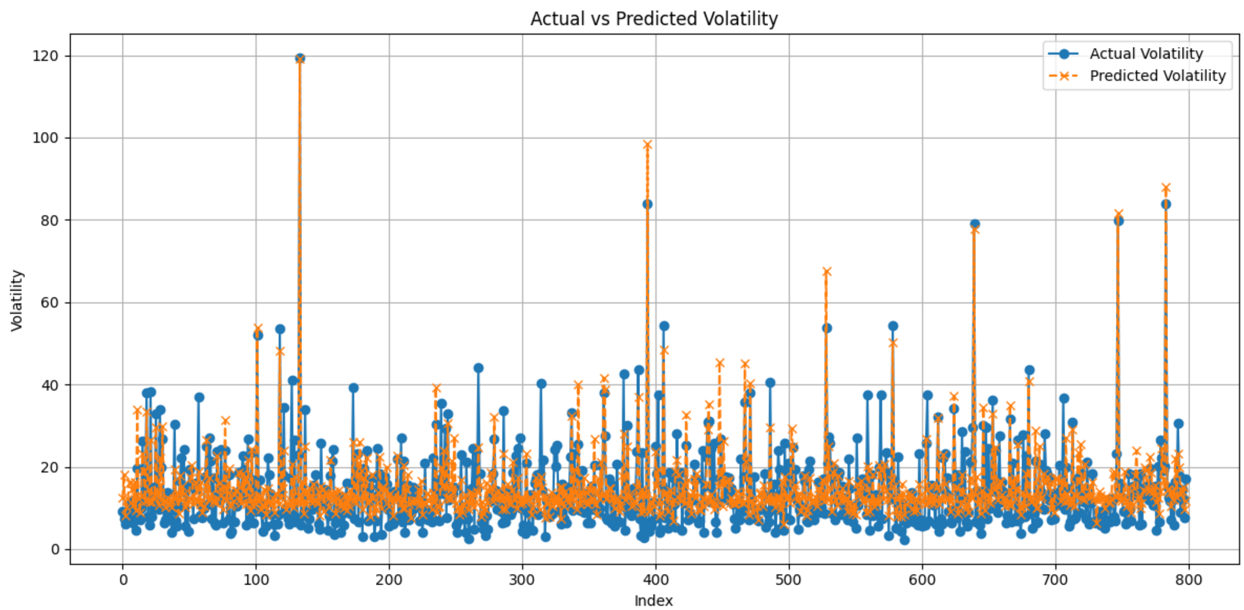

Figure 9: Predicted Vs Actual Volatility for S&P500 by LSTM

In comparison to the other models discussed, LSTM showed greater difficulty in maintaining low error rates. The MSE of 29.80, while still indicative of the model's ability to capture trends in volatility, is higher than that of the XGBoost model, suggesting that LSTM's reliance on sequential data and temporal dependencies may limit its adaptability to abrupt changes in market conditions. The higher MAPE of 36.35% further reflects the model's reduced accuracy in capturing scale-independent predictions, particularly in volatile periods.

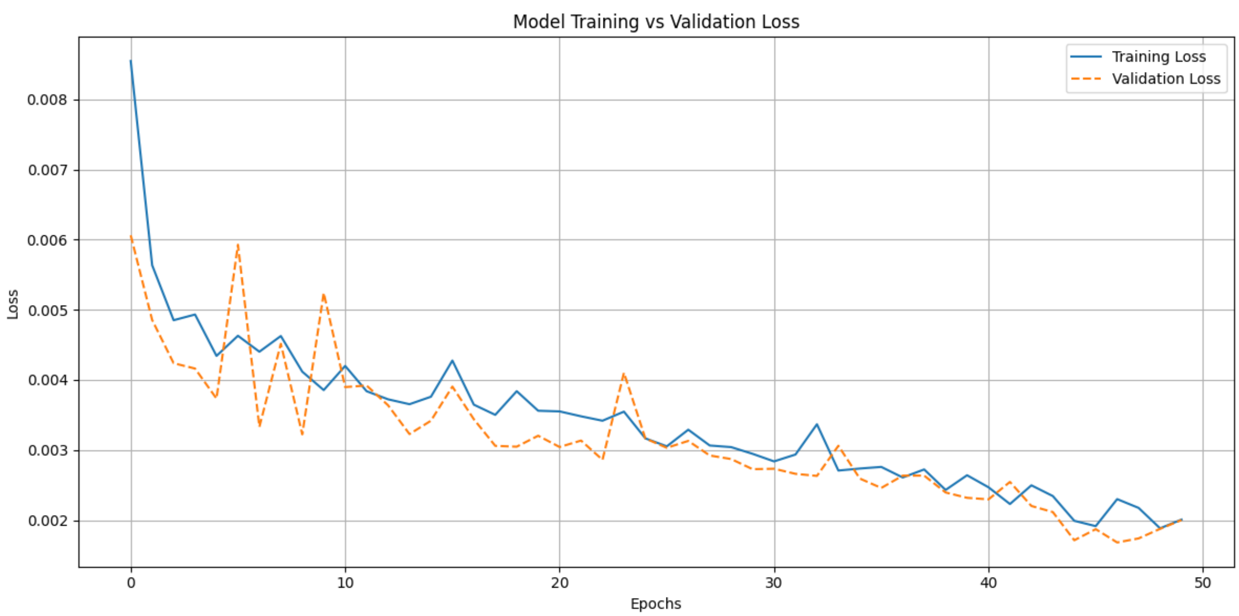

Figure 10: Training vs Validation Loss

The learning curve analysis, as visualized in the training versus validation loss graph, provides valuable insights into the model's behavior. Both the training and validation losses decrease steadily over epochs, with convergence indicating that the model effectively learns underlying patterns without significant overfitting. This observation underscores the stability of the LSTM model, which is a crucial characteristic for real-world financial forecasting applications. However, the fluctuations in validation loss during certain epochs suggest sensitivity to noise in the data, which could explain the relatively higher error metrics.

The ability of the LSTM model to capture temporal dependencies is evident from the alignment between predicted and actual volatility trends, as shown in the prediction graph. However, discrepancies between the two during periods of rapid market movement highlight the model's challenge in adjusting to sudden shifts, which are characteristic of financial time series data.

In summary, while the LSTM model demonstrates strength in modeling sequential dependencies, its higher error metrics, particularly MAPE, indicate room for improvement in accurately capturing rapid volatility changes. The convergence of the learning curves, however, emphasizes the model's stability and reliability when applied to unseen data, making it a valuable tool for long-term trend analysis in financial markets. Nonetheless, its performance in the S&P 500 context suggests the need for further refinement to enhance its responsiveness to extreme market dynamics.

Comparative Analysis of Models

The comparative analysis of the models employed in this study highlights their varied strengths, limitations, and suitability for different volatility forecasting scenarios. XGBoost emerged as the most robust model, delivering superior performance across all key metrics. It achieved a test MSE of 9.23, an RMSE of 3.04, and an R^2 of 0.88, demonstrating its ability to explain 88% of the variance in actual volatility. The MAPE of 18.2% further underscores its precision in capturing nuanced market behavior. Unlike traditional models such as EWMA and GARCH, XGBoost does not require transformations to address scale differences, making it highly adaptable to complex datasets. This adaptability proved particularly advantageous during periods of extreme volatility, such as the 2020 COVID-19 market shock, where rapid fluctuations overwhelmed simpler approaches. The XGBoost model also demonstrated remarkable alignment between predicted and actual volatility during sharp market spikes, showcasing its ability to adapt to sudden disruptions effectively.

However, the computational intensity of XGBoost is a notable drawback, as training the model requires significant processing power. This limitation could pose challenges for real-time applications or organizations with limited computational resources. Additionally, its "black box" nature reduces interpretability, which can be a concern for stakeholders requiring transparency in financial decision-making. While its predictive accuracy is undeniable, careful validation is essential to avoid overfitting and ensure reliable results.

The LSTM model, designed to capture temporal dependencies, produced mixed results. It achieved a test MSE of 29.80, RMSE of 5.46, and R^2 of 0.72, lagging behind XGBoost but outperforming traditional models in certain scenarios. LSTM's strength lies in its ability to model sequential patterns effectively, which proved beneficial during stable market periods where gradual volatility trends prevailed. The learning curve analysis validated the model's generalization capability, with training and validation losses converging steadily over 50 epochs. This convergence indicates that LSTM avoided overfitting and maintained reliability when applied to unseen data.

Despite these strengths, LSTM struggled during periods of extreme volatility, such as the 2018 Volpocalypse. The model exhibited delays in responding to rapid market shifts, as seen in discrepancies between actual and predicted volatility. This limitation arises from LSTM's reliance on past patterns, which may fail to account for sudden structural changes in the market. Additionally, LSTM's complexity and sensitivity to hyperparameter tuning make it less accessible for practitioners in scenarios requiring quick deployment or interpretability.

EWMA, a simpler and more traditional approach, offered a balance between traditional and machine learning methodologies. With an MSE of and an R^2 of 0.811, EWMA captured over 81% of the variance in actual volatility, highlighting its ability to dynamically adjust to market trends. The EWMA graph showed a tight correlation between predicted and actual volatility during steady market conditions, demonstrating the model's responsiveness to evolving dynamics. Unlike the constant volatility baseline, EWMA assigns exponentially decreasing weights to older data points, emphasizing recent market movements.

However, EWMA's reliance on a decay factor introduces sensitivity to parameter calibration. While it significantly outperformed the constant volatility baseline by a factor of 9.4, it fell short of the rolling standard deviation baseline, which achieved a lower MSE. This finding suggests that EWMA's additional complexity does not always translate to superior performance, particularly in environments with minimal volatility clustering. Its inability to capture extreme spikes, as seen during the 2016 Brexit vote, limits its utility in highly volatile markets.

The GARCH model, despite its theoretical robustness, was the weakest performer in this study. For the S&P 500, GARCH achieved an MSE of 0.000161 and an RMSE of 0.01267 but consistently underperformed relative to the rolling mean baseline. This discrepancy was particularly evident in its HMSE score, which was nearly 5.7 times worse than the benchmark. GARCH struggled to adapt to rapid market changes, as observed during the 2015 EUR/CHF blow-up and other high-volatility events. Its assumption of symmetric responses to shocks further limited its effectiveness, as it failed to account for the leverage effect, where negative shocks disproportionately impact volatility.

Nevertheless, GARCH's interpretability remains a strength. Its explicit modeling of volatility clustering and conditional heteroskedasticity provides transparency valued by stakeholders. This makes GARCH a useful tool in scenarios prioritizing theoretical rigor and interpretability over predictive accuracy. However, its high computational demands and sensitivity to parameter selection render it less practical for real-time forecasting in dynamic market conditions.

From a business perspective, the study underscores the importance of aligning model selection with specific use cases and market conditions. XGBoost's high accuracy and adaptability make it ideal for applications requiring granular predictions, such as risk management and portfolio optimization in fast-paced environments. While LSTM was less accurate overall, it excelled in capturing temporal patterns, making it valuable for longer-term forecasting or stable market conditions. EWMA's simplicity and responsiveness make it a practical choice for real-time assessments, while GARCH's theoretical foundation serves applications requiring interpretability and consistency.

These comparisons emphasize that no single model universally outperforms others across all scenarios. XGBoost dominated in terms of accuracy and adaptability but posed challenges related to computational intensity and complexity. LSTM excelled in stable markets but struggled during abrupt market shifts. Meanwhile, EWMA and GARCH traded off simplicity and interpretability for predictive power, demonstrating the importance of a tailored approach to volatility forecasting.

Ultimately, this study highlights the value of combining multiple models to gain a comprehensive understanding of market dynamics. A hybrid approach, integrating machine learning techniques with traditional frameworks, allows stakeholders to leverage the strengths of advanced predictive models while retaining the interpretability and theoretical rigor of established methodologies. This strategy provides a robust solution for navigating the complexities of financial markets and delivering actionable insights in an evolving landscape.

Practical Implications of Volatility Forecasting Models in Business Decision-Making

The application of volatility forecasting models extends far beyond academic interest, holding immense value for businesses, individual investors, portfolio managers, and policymakers. These models enable stakeholders to anticipate market fluctuations, evaluate risks, and optimize investment strategies. By offering data-driven insights, they enhance decision-making in a field often characterized by uncertainty and rapid change. Each model, from traditional approaches like GARCH and EWMA to advanced techniques like XGBoost and LSTM, provides distinct advantages suited to different use cases.

GARCH: Interpretable Insights for Institutional Decision-Making

The GARCH model, with its focus on conditional heteroskedasticity and volatility clustering, offers businesses a theoretically grounded framework for understanding market behavior. While its predictive power lagged behind other models in this study, its interpretability remains a significant strength. For institutional investors, such as pension funds and hedge funds, this interpretability is vital. The ability to explicitly model the conditional variance of returns allows decision-makers to quantify risk more transparently, aligning with compliance requirements and risk management protocols.

For example, during high-volatility periods like the 2008 financial crisis, GARCH models can be used to identify patterns of volatility clustering, helping firms anticipate prolonged market instability. This insight can guide decisions on whether to increase cash reserves, rebalance portfolios toward safer assets, or implement hedging strategies using derivatives like options and futures. The model's clear structure also makes it a valuable tool for risk officers who need to communicate volatility trends to stakeholders without overwhelming them with overly technical explanations.

EWMA: Adaptable and Efficient for Dynamic Markets

The EWMA model provides a middle ground between simplicity and responsiveness, making it a practical choice for businesses operating in fast-changing markets. Its ability to assign greater weight to recent data allows it to adapt to evolving conditions, which is particularly useful for firms engaged in short-term trading strategies or daily risk management.

Consider a retail trading firm managing a portfolio of high-volatility assets, such as technology stocks or cryptocurrency. Using EWMA, the firm can dynamically adjust portfolio weights based on the most recent volatility trends, reducing exposure to risky assets during turbulent periods. For example, during the COVID-19 pandemic, when market volatility spiked, EWMA could have provided timely adjustments to portfolio allocations, minimizing potential losses while maintaining exposure to growth opportunities.

Furthermore, EWMA's computational efficiency makes it accessible for smaller firms or individual investors without the resources to implement more complex models. An individual investor, for instance, could use EWMA to monitor the volatility of their investment portfolio and make informed decisions about asset allocation. By focusing on recent market conditions, they could reduce their stake in volatile stocks during periods of uncertainty and reinvest in safer options, such as bonds or ETFs, when the market stabilizes.

XGBoost: Precision and Scalability for Diverse Applications

XGBoost stands out as a robust and adaptable tool for volatility forecasting, particularly in complex, data-rich environments. Its ability to incorporate non-linear relationships and interactions among features makes it a preferred choice for businesses seeking highly accurate and actionable insights.

For instance, a multinational investment bank managing diversified portfolios across global markets can benefit from XGBoost's predictive capabilities. By leveraging data on price movements, trading volumes, and macroeconomic indicators, the model can forecast short-term volatility with exceptional precision. During events like the Brexit referendum or the U.S.-China trade tensions, where markets exhibited extreme fluctuations, XGBoost could help identify high-risk assets and guide reallocation strategies to mitigate potential losses.

Moreover, XGBoost's feature importance analysis offers additional value by highlighting the most influential factors driving volatility. This can inform targeted decision-making, such as identifying which sectors or regions are most susceptible to market shocks. For example, during the 2020 market crash, the model could have pinpointed technology stocks as particularly volatile, prompting businesses to hedge their exposure in this sector.

The scalability of XGBoost also enables its application across various financial instruments, from equities and ETFs to derivatives and commodities. For instance, an energy trading firm could use the model to forecast volatility in oil prices, guiding procurement and inventory decisions. By anticipating price swings, the firm can optimize purchase schedules, reducing costs and improving profitability.

LSTM: Temporal Analysis for Long-Term Forecasting

LSTM models, with their ability to capture temporal dependencies, are particularly suited for tasks requiring an understanding of sequential patterns in financial data. While they may not excel during extreme volatility, their strength lies in modeling gradual trends, making them valuable for long-term forecasting and planning.

Consider a real estate investment trust (REIT) planning its investment strategy over the next five years. Using an LSTM model, the firm could analyze historical market data to forecast long-term volatility trends, helping it identify periods of relative stability or heightened risk. This information could guide decisions on property acquisitions, debt financing, and portfolio diversification. For example, during periods of anticipated low volatility, the REIT might allocate more resources to high-growth assets, whereas in high-volatility periods, it could focus on stable, income-generating properties.

Similarly, policymakers and regulatory bodies can leverage LSTM models to anticipate systemic risks in financial markets. For instance, by analyzing historical patterns of bank stock volatility, regulators could identify early warning signs of financial instability, prompting preemptive measures to stabilize the sector.

Practical Examples and Implications

The implications of these models extend to numerous other domains. For algorithmic traders, advanced models like XGBoost and LSTM offer opportunities to optimize trading strategies by predicting short-term price movements with high accuracy. A quantitative trading firm, for instance, could use these models to design automated trading algorithms that execute trades based on predicted volatility spikes, capturing opportunities for profit while minimizing downside risk.

Insurance companies, on the other hand, can use models like EWMA and GARCH to price products such as catastrophe bonds or market-linked life insurance policies. By accurately forecasting market volatility, insurers can adjust premiums and reserve requirements, ensuring financial stability even during economic downturns.

Individual investors also stand to benefit from these tools, particularly as user-friendly platforms increasingly integrate machine learning models into their offerings. For example, a retail investor using an app equipped with XGBoost-based forecasting could receive personalized recommendations to rebalance their portfolio, reducing exposure to high-risk assets during volatile periods.

Reshaping Decision-Making with a Tailored Approach

The diverse capabilities of these models highlight the importance of aligning model selection with specific business objectives and market conditions. No single model is universally superior, but rather, their value lies in their targeted application. Businesses can leverage XGBoost's precision for granular, short-term predictions, EWMA's adaptability for dynamic market conditions, LSTM's temporal analysis for long-term forecasting, and GARCH's interpretability for compliance and transparency.

Ultimately, these models have the potential to reshape decision-making in financial markets by providing stakeholders with actionable insights grounded in data. By integrating advanced forecasting tools into their operations, businesses and investors can navigate uncertainty more effectively, optimize resource allocation, and gain a competitive edge in an increasingly complex financial landscape. The key to success lies in understanding the strengths and limitations of each model and applying them strategically to address specific challenges and opportunities.

Conclusion

This dissertation provides a comprehensive analysis of traditional and machine learning models for forecasting stock market volatility, focusing on their applicability in varying market conditions. The findings reveal that while traditional models like GARCH and EWMA offer interpretability and are well-suited for steady markets, their limitations become apparent in periods of extreme volatility. In contrast, machine learning approaches, particularly XGBoost, demonstrate exceptional predictive accuracy, effectively capturing non-linear patterns and abrupt market changes. However, the computational complexity of XGBoost and the interpretability challenges of LSTM highlight the trade-offs inherent in advanced methodologies.

For practitioners, the study emphasizes the need to align model selection with specific business objectives. GARCH and EWMA remain valuable for their simplicity and ability to model conditional variance, making them suitable for applications requiring transparency and computational efficiency. On the other hand, machine learning models excel in dynamic environments, providing granular forecasts critical for high-frequency trading, risk management, and portfolio rebalancing.

The research underscores the potential of hybrid approaches, combining the interpretability of traditional models with the adaptability of machine learning techniques. Such integration offers a pathway to more robust, context-sensitive forecasting, addressing the complexities of modern financial markets. Future studies could explore the inclusion of alternative data sources, such as sentiment analysis and economic policy uncertainty, to further enhance predictive performance. Ultimately, this dissertation highlights the transformative potential of advanced analytics in reshaping financial decision-making, paving the way for more informed, data-driven strategies.

References

Alfeus, M. and Nikitopoulos, C.S. (2022) 'Forecasting volatility in commodity markets with long-memory models', Journal of Commodity Markets, 28 pp. 100248. Available.

Amirshahi, B. and Lahmiri, S. (2023) 'Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies', Machine Learning with Applications, 12 pp. 100465. Available.

Bergsli, L.Ø. et al. (2022) 'Forecasting volatility of Bitcoin', Research in International Business and Finance, 59 pp. 101540. Available.

Capelli, P., Ielasi, F. and Russo, A. (2021) 'Forecasting volatility by integrating financial risk with environmental, social, and governance risk', Corporate Social Responsibility and Environmental Management, 28 (5), pp. 1483-1495. Available.

Cumby, R., Figlewski, S. and Hasbrouck, J. (1994) 'International asset allocation with time varying risk: an analysis and implementation', Japan and the World Economy, 6 (1), pp. 1-25. Available.

Day, T.E. and Lewis, C.M. (1993) 'Forecasting futures market volatility', The Journal of Derivatives, 1(2), pp. 33–50.

Gong, X. and Lin, B. (2018) 'Structural breaks and volatility forecasting in the copper futures market', Journal of Futures Markets, 38 (3), pp. 290-339. Available.

Guo, X. et al. (2022) 'Forecasting volatility of EUA futures: New evidence', Energy Economics, 110 pp. 106021. Available.

Hamid, A. and Heiden, M. (2015) 'Forecasting volatility with empirical similarity and Google Trends', Journal of economic behavior & organization, 117 pp. 62-81. Available.

Haugom, E. et al. (2014) 'Forecasting volatility of the US oil market', Journal of Banking & Finance, 47 pp. 1-14. Available.

Jia, F. and Yang, B. (2021) 'Forecasting Volatility of Stock Index: Deep Learning Model with Likelihood-Based Loss Function', Complexity, 2021 (1), pp. 5511802. Available.

Kang, S.H., Kang, S.-M. and Yoon, S.-M. (2009) 'Forecasting volatility of crude oil markets', Energy Economics, 31 (1), pp. 119-125. Available.

Kim, J.-M., Jun, C. and Lee, J. (2021) 'Forecasting the volatility of the cryptocurrency market by GARCH and Stochastic Volatility', Mathematics, 9 (14), pp. 1614. Available.

Kuen, T.Y. and Hoong, T.S. (1992) 'Forecasting volatility in the Singapore stock market', Asia Pacific Journal of Management, 9 pp. 1-13. Available.

Lang, Q. et al. (2021) 'Is Baidu index really powerful to predict the Chinese stock market volatility? New evidence from the internet information', China Finance Review International, 13 (2), pp. 263-284. Available.

Lehrer, S., Xie, T. and Zhang, X. (2021) 'Social media sentiment, model uncertainty, and volatility forecasting', Economic Modelling, 102 pp. 105556. Available.

Liu, J. et al. (2021) 'Forecasting the volatility of EUA futures with economic policy uncertainty using the GARCH-MIDAS model', Financial Innovation, 7 pp. 1-19. Available.

Liu, R. and Gupta, R. (2022) 'Investors' uncertainty and forecasting stock market volatility', Journal of Behavioral Finance, 23 (3), pp. 327-337. Available.

Marfatia, H.A., Ji, Q. and Luo, J. (2022) 'Forecasting the volatility of agricultural commodity futures: The role of co-volatility and oil volatility', Journal of Forecasting, 41 (2), pp. 383-404. Available.

Megaritis, A., Vlastakis, N. and Triantafyllou, A. (2021) 'Stock market volatility and jumps in times of uncertainty', Journal of International Money and Finance, 113 pp. 102355. Available.

Poon, S.-H. and Granger, C. (2005) 'Practical issues in forecasting volatility', Financial analysts journal, 61 (1), pp. 45-56. Available.

Poon, S.-H. and Granger, C.W.J. (2003) 'Forecasting volatility in financial markets: A review', Journal of economic literature, 41 (2), pp. 478-539. Available.

Ramos-Pérez, E., Alonso-González, P.J. and Núñez-Velázquez, J.J. (2019) 'Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network', Expert Systems with Applications, 129 pp. 1-9. Available.

Raza, S.A. et al. (2023) 'Forecasting the volatility of precious metals prices with global economic policy uncertainty in pre and during the COVID-19 period: Novel evidence from the GARCH-MIDAS approach', Energy Economics, 120 pp. 106591. Available.

Roh, T.H. (2007) 'Forecasting the volatility of stock price index', Expert Systems with Applications, 33 (4), pp. 916-922. Available.

Takahashi, M., Watanabe, T. and Omori, Y. (2024) 'Forecasting daily volatility of stock price index using daily returns and realized volatility', Econometrics and Statistics, 32 pp. 34-56. Available.

Verma, S. (2021) 'Forecasting volatility of crude oil futures using a GARCH–RNN hybrid approach', Intelligent Systems in Accounting, Finance and Management, 28 (2), pp. 130-142. Available.

Walsh, D.M. and Tsou, S.C. (1998) 'Forecasting index volatility: Some observations on the range-based method', International Review of Financial Analysis, 7(2), pp. 79–95.

Wang, L. et al. (2020) 'Forecasting stock price volatility: New evidence from the GARCH-MIDAS model', International Journal of Forecasting, 36 (2), pp. 684-694. Available.

Wen, F., Gong, X. and Cai, S. (2016) 'Forecasting the volatility of crude oil futures using HAR-type models with structural breaks', Energy Economics, 59 pp. 400-413. Available.

Xia, Y. et al. (2023) 'The role of uncertainty index in forecasting volatility of Bitcoin: fresh evidence from GARCH-MIDAS approach', Finance Research Letters, 52 pp. 103391. Available.

Xu, Y. and Lien, D. (2022) 'Forecasting volatilities of oil and gas assets: A comparison of GAS, GARCH, and EGARCH models', Journal of Forecasting, 41 (2), pp. 259-278. Available.