Journal of Mathematical Modeling
Vol. -, No. -, 2026, pp. -. Research Article

Enhancing implied volatility forecasting: multi-model approaches for the S&P500 index

Navideh Modarresi[†*], Reza Kazemi[†], Alireza Mousavi[†]

[†] *Department of Mathematics, Allameh Tabataba'i University*
Email(s): n.modarresi@atu.ac.ir, rkazemi453@gmail.com, ar.mousavie@gmail.com

Abstract. Implied volatility is a crucial indicator in financial markets, as it reflects market expectations of future volatility and serves as a cornerstone for option pricing, risk management, and asset allocation. Accurate tracking and forecasting of implied volatility are essential for investors and portfolio managers to optimize returns and manage risks effectively. This paper explores several modeling approaches for forecasting the implied volatility of the S&P 500 index, focusing on exponential autoregressive conditional heteroskedasticity (EGARCH), long short-term memory (LSTM) neural networks, and a non-linear autoregressive model with exogenous inputs (NARX). In addition, a rough fractional stochastic volatility (RFSV) model is also examined. The empirical study demonstrates that the LSTM model offers superior forecasting performance compared to EGARCH, NARX, and RFSV. These findings have important implications for practitioners and researchers aiming to enhance risk management and trading strategies.

Keywords: Implied volatility, LSTM neural network, NARX model, rough fractional model
AMS Subject Classification 2020: 91G20, 68T01, 68T07.

1 Introduction

Forecasting asset-return volatility is pivotal for effective option pricing, risk management, and trading strategies. Volatility defined as the variability of an asset's price over time is of particular importance in financial markets, where unexpected fluctuations can significantly affects pricing and portfolio decisions. Among various measures, implied volatility, derived from option prices, offers valuable forward-looking information by reflecting market participant's collective expectations of future price fluctuations [20].

Unlike historical volatility which focuses only on past price movements, implied volatility incorporates market's collective expectations and insights regarding future conditions. It anticipates the future

*Corresponding author
Received: 19 February 2025/ Revised: 30 July 2025/ Accepted: 15 September 2025
DOI: 10.22124/jmm.2025.29881.2667

movement of the underlying asset's price and predicts the extent of price fluctuation, aiding in determining the profitability potential options before expiry. This forward-looking nature allows implied volatility to adjust more rapidly to new information [3]. Thereby, it helps the practitioners and investors understand market movements and inform trading strategies [5].

However, forecasting implied volatility, which significantly affects option pricing, remains a major challenge in finance. Accurate forecasting is essential for making informed investment decisions and managing portfolio risk effectively. The reliance of implied volatility on various maturities and strike prices influences its accuracy in estimating future realized volatility. Goyal and Saretto [11] found that differences between historical and implied volatility are temporary, with one-month implied volatility effectively serving as a reliable measure for longer-term historical volatility. This finding highlights the importance of options implied volatility as a representation of realized volatility. Econometrics models like the generalized autoregressive conditional heteroskedasticity (GARCH) model [2] and its extension, the exponential GARCH (EGARCH) model [19], have been widely used to model time-varying volatility. However, these models often struggle to capture the non-linear and complex dynamics inherent in financial time series, limiting their effectiveness in tracking implied volatility [5]. Stochastic volatility models, such as Heston's model [13], also seek to account for aspects like mean reversion and the correlation between asset returns and volatility. Heston and Bates models have semi-closed form solutions for European option pricing, making them relatively easier to calibrate with market data. A new model for pricing American options has been developed, integrating stochastic volatility and jump-diffusion model, and has been shown to outperform traditional models in accuracy using S&P500 data [9].

In recent years, artificial neural networks (ANNs) have gained popularity due to their capacity to model complex non-linear relationships and detect detailed patterns within data [12,22]. Building on this trend, a novel leading moving average indicator based on a hybrid ANFIS-Wavelet approach has been introduced to enhance market trend prediction and trading decisions, showing effective performance on NASDAQ data [16]. The non-linear autoregressive model with exogenous inputs (NARX) is a type of an ANN used for time series forecasting. It is indeed notable for its effectiveness in modeling non-linear dynamic systems, including many financial applications. For example, D'Ecclesia and Clementi [5] demonstrated that NARX networks outperform traditional models like EGARCH and the Heston model for forecasting implied volatility. The EGARCH model is used to analyze and predict the volatility of time series data. It allows for asymmetric responses of volatility to shocks and captures the exponential dynamics of volatility where volatility clustering is observed. Similarly, Stokes and Abou-Zaid [21] showed the effectiveness of ANNs in forecasting exchange rates. The model's capability to handle complex relationships and time series data makes it particularly useful in this field, allowing analysts to better predict and understand financial trends.

Building on NARX model successes, innovations in neural network architectures, most notably long short-term memory (LSTM) networks [14] have emerged. The LSTMs address the vanishing gradient problem, inherent in traditional recurrent neural networks, enabling the capture of long-term dependencies in sequential data and making them suitable for financial time series forecasting. D'Ecclesia and Clementi [5] found that ANN models generally outperform traditional frameworks like Heston in effectively tracking implied volatility dynamics, particularly in terms of accuracy related to root mean squared error. Additionally, rough fractional stochastic volatility (RFSV) models have emerged to better capture the 'rough' character observed in financial volatility. Unlike traditional stochastic models and fractional Brownian motion approaches that typically assume a Hurst exponent $H \in (1/2, 1)$, the RFSV framework explicitly incorporates the observed roughness $H \in (0, 1/2]$ in volatility dynamics Gatheral et al. [10]. This method has proven promising for improving the modeling and forecasting volatility by addressing the nuanced roughness inherent in financial time series. Moreover, empirical studies underscore the universality of rough volatility. Bennedsen et al. [1] analyzed volatility time series for over five thousand individual US stocks, estimating rough volatility in each instance. A recent systematic review by Zhao et al. [24] provides a comprehensive overview of neural network-based financial volatility forecasting studies published after 2015, highlighting key trends and gaps. The review highlights historical volatility as the most frequently forecasted proxy, followed by realized volatility, while implied volatility is the least utilized, primarily due to limitations in data accessibility. In terms of methodological approaches, multilayer perceptrons (MLPs) are predominant, frequently integrated with autoregressive models such as GARCH. Recurrent neural networks (RNNs), particularly LSTM architectures, represent the second most commonly employed technique, whereas convolutional neural networks (CNNs) are comparatively rare. The review underscores a notable gap between state-of-the-art machine learning models and those currently employed in volatility forecasting, highlighting the limited adoption of deep learning techniques. It recommends addressing this disparity by exploring advanced architectures, including temporal convolutional networks and WaveNet, to enhance forecasting performance. It also emphasizes the importance of establishing shared tasks to facilitate meaningful comparisons across studies, thereby mitigating heterogeneity in volatility proxies, asset, and forecasting horizons. Recent advancements have also explored graph neural networks (GNNs) for volatility forecasting, leveraging relational dependencies in multivariate financial time series. For instance, Deng and Hooi [6] proposed a GNN-based approach for anomaly detection in multivariate time series, which has implications for capturing interconnected volatility patterns in financial data, extending beyond sequential models like LSTM. While Zhao et al. [24] underscore the potential of such emerging methods, they highlight that financial applications lag behind general machine learning advancements, motivating further integration of relational and deep learning models.

Despite significant progress in volatility forecasting techniques, direct comparisons among prominent models such as EGARCH, NARX, LSTM, and RFSV remain limited, particularly in the context of implied volatility. This gap hinders a comprehensive understanding of their relative performance and applicability in real world financial settings. To address this, the present study enhances prior research by introducing a novel set of input features and modifying the architecture of standard LSTM networks to better capture the complex dynamics of financial time series data. The primary aim of this study is to identify the most accurate approach for forecasting implied volatility. Using S&P 500 index data and its corresponding implied volatility from 2001 to 2024, the analysis offers valuable insights for both researchers and practitioners. These findings can inform the development of more effective risk management strategies and trading decisions in increasingly volatile markets.

To highlight the research gaps and motivations, Table 1 summarizes key prior works on volatility forecasting, evaluated across model types, datasets, performance metrics, strengths, limitations, and how they relate to our contribution. This comparative analysis highlights critical shortcomings: although econometric models (such as EGARCH) and stochastic approaches (like RFSV) effectively capture specific market dynamics, they tend to underperform in scenarios involving non-linear patterns and long-term dependencies. In contrast, advanced neural networks particularly LSTM architectures demonstrate superior performance, especially when direct multi-model comparisons are conducted using extended implied volatility datasets. Emerging methodologies such as GNNs offer valuable relational insights; however, they often fall short in integrating sequential forecasting capabilities, as observed in [24]. To address these limitations, our study presents a rigorous empirical comparison across modeling paradigms,

Table 1: Comparative summary of previous related works on volatility forecasting

Study	Model Types	Datasets	Performance Metrics	Strengths	Limitations	Relation to Our Study
DEcclesia and Clementi [5] (2021)	NARX, EGARCH, Heston	S&P 500 implied volatility (short-term)	RMSE, MAE	NARX captures non-linear dynamics better than parametric models	Limited to short-term data; no long-term dependencies. No RFSV	Compares with LSTM and RFSV on longer datasets, confirms NARX strengths but shows LSTM superiority
Gatheral et al. [10] (2018)	RFSV	Various financial time series	Roughness estimation (Hurst exponent)	Captures rough volatility universally	Assumes specific roughness; less focus on forecasting accuracy	Empirically test RFSV against neural models on S&P 500 data
Bennedsen et al. [1] (2022)	RFSV	5,000+ US stocks	Hurst exponent estimation	Demonstrate roughness in large-scale data	Descriptive rather than predictive focus	Incorporates RFSV with empirical Hurst estimation for forecasting
Zhao et al. [24] (2024)	Systematic review of NN methods (e.g., MLP, RNN, hybrids with GARCH)	Various (post-2015 studies)	N/A (review)	Identifies trends in NN volatility forecasting and gaps in deep learning adoption	Does not perform new empirical tests	Multi-model comparison on long-term implied volatility
Deng et al. [6] (2021)	Graph Neural Networks	Multivariate time series (general)	Anomaly detection accuracy	Handles relational dependencies in data	Not specifically tailored for volatility forecasting	Highlights emerging GNN trends; our study focuses on sequential models but notes potential for future hybrid approaches
Kim et al. [17] (2018)	Hybrid model integrating LSTM with GARCH	Financial time series (e.g., stock prices)	RMSE, MAE	Excels in long-term dependencies	Single-model focus; no comparison with econometric or rough models	Aligns with our LSTM findings; we provide multi-model validation
Poon and Granger [20] (2003)	GARCH variants	Various financial markets	Forecasting error	Comprehensive review of econometric models	Often underperforms in non-linear scenarios	Shows ML models outperform GARCH, addressing this gap

underscoring the practical need for hybrid architectures or advanced neural techniques in effective volatility forecasting.

The rest of the paper is organized as follows. In Section 2, we present some volatility forecasting models employed in our comparative study and also highlight the novel aspects that show how our approach differs from prior research. Section 3 is devoted to data analysis including data description and methodology. Section 4 is dedicated to performance evaluation, where we compare the models using statistical error metrics and assess cross-validation robustness. Additionally, this section includes a comparison with existing literature, situating our findings within the broader research landscape and highlighting consistencies and deviations relative to prior studies. Finally, Section 5 concludes the paper.

2 Volatility forecasting models

In this section, we explore the estimation of equity returns volatility by evaluating various modeling techniques. The NARX is assessed against established financial models, specifically the Heston model and the EGARCH model. Previous research, notably by D'Ecclesia and Clementi [5], indicates that NARX outperforms both the Heston and EGARCH models in forecasting implied volatility. This finding highlights the NARX model's strength in capturing the complexities of financial markets. Volatility is defined as the standard deviation of stock returns provided by the variable per unit of time when the return is expressed using continuous compounding. So given S_t, the stock price at the end of day t, the historical variance over a time horizon $[0,T]$ is given by:

$$\sigma_t^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2 \cong \frac{1}{T-1}\sum_{t=1}^{T} r_t^2, \tag{1}$$

where $r_t = \ln \frac{S_t}{S_{t-1}}$. In general, market participant are used to deal with annualized volatility which is given by $\hat{\sigma}_t = \sqrt{252}\sigma_t$.

2.1 Artificial neutral network approaches

Recent developments in the finance sector have sparked significant interest in a new class of non-linear models inspired by the structure of the human brain, commonly referred to as ANNs. The ANN techniques have been extensively utilized for forecasting stock prices and historical volatility [7, 18].

- **NARX**
  Now, we propose an innovative approach for modeling stock return volatility by leveraging machine learning and signal processing methodologies, particularly through the application of the NARX. The defining equation for the NARX model is

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-n}, u_{t-1}, u_{t-2}, \ldots, u_{t-n}). \tag{2}$$

  The NARX model employs a learning process similar to that of other neural network architectures. In the context of regression, model parameters are estimated using a training set comprised of input-output samples that represent the function we aim to approximate. To ensure that the model generalizes effectively, it is crucial to accurately estimate the function on data not included in the training set. In our study, the variable y_t corresponds to historical rolling volatility calculated over various rolling windows, specifically 20, 120, and 252 days. By selecting different window sizes, we can capture both short-term and long-term characteristics of stock return volatility.

  In addition, incorporating relevant supplementary information can enhance the training set. For example, trading volumes can provide valuable insights into market liquidity. Typically, increasing trading volumes are observed during bullish market conditions, where heightened enthusiasm among buyers drives prices higher. Conversely, if prices rise while trading volume declines, it may

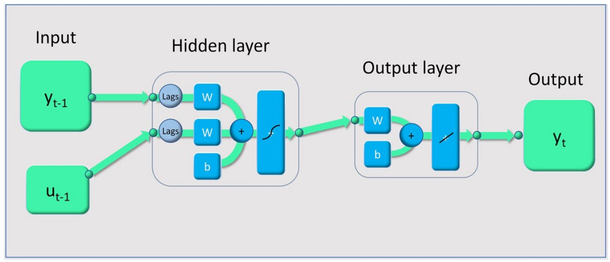

Figure 1: NARX neural network

indicate a lack of interest, suggesting a potential reversal in trend. Thus, price movements that occur on low volume are less significant, while changes on high volume may signal a fundamental shift in the stock, offering critical information for training the network. In our analysis, we trained the neural network using 70% of the available data for each price return series, with the objective of minimizing the sum of squared errors. The implementation was conducted using Python.

Historical rolling volatility is a method used to estimate stock return volatility, while NARX models are used for forecasting potentially of volatility indices. The simplest approach to measure time varying volatility is given by the historical rolling volatility estimated on log returns after choosing the right size of the rolling window. The historical yearly rolling window volatility, $\hat{\sigma}_{n,t}$ is given by

$$\hat{\sigma}_{n,t} = \sqrt{\frac{1}{n}\sum_{s=t-n-1}^{n}(r_s - \bar{r})^2 \cdot 252}, \tag{3}$$

where n is the window size, r_s the log-difference and $\bar{r}$ is the sample mean of the observations in each rolling window.

A fundamental challenge in this approach is determining the optimal window size. Ideally, the window size should be selected to minimize the volatility of $\hat{\sigma}_{n,t}$ providing the best estimate of true volatility. However, a primary criticism of this method is that it treats all observations with equal weight, failing to account for the greater influence that more recent prices have compared to older data. Consequently, an exponentially weighted moving average approach may yield more accurate estimates by placing more weight on recent observations. In this study, we estimate stock returns volatility using the historical rolling volatility approach, acknowledging both its simplicity and the limitations associated with the choice of window size.

- **LSTM**

  To process sequential data, we utilize LSTM networks, a sophisticated variant of RNNs. While RNNs are inherently designed to manage sequential data, they often encounter challenges, particularly the vanishing gradient problem. This issue complicates the learning of long-term dependencies within the data. The LSTMs present a robust solution to this limitation through the implementation of a more complex architecture that governs the flow of both historical memory and new inputs, effectively addressing the challenges associated with standard RNNs [14]. The LSTMs consist of several fundamental components known as gates, which employ activation functions to regulate the flow of information throughout the network. The primary gates and states involved in an LSTM layer include:

1. Cell State (also referred to as the memory cell) The Cell State retains information from previous LSTM cells, enabling it to capture and remember long-term relationships in the data. This information is protected by the Forget Gate and updated by the Input Gate.
2. Hidden State (also known as the output of the LSTM cell): The Hidden State reflects the output from prior LSTM cells and is utilized in conjunction with the Forget Gate, Input Gate, and Output Gate to generate a new Hidden State, serving as the output of the LSTM cell.
3. Forget Gate: This gate regulates how much information is retained from the Cell State, providing the model with the ability to discard irrelevant data.
4. Input Gate: The Input Gate determines the extent to which new information should be incorporated into the Cell State.
5. Output Gate: This gate controls the degree of information from the Cell State that is used to generate the output of the LSTM cell, also referred to as the Hidden State.

By integrating these components, LSTMs effectively learn and remember patterns within sequential data, making them powerful tools for a range of tasks, including time series forecasting, natural language processing, and various applications in finance and economics. In our study, we apply LSTM networks to model the implied volatility of return.

2.2 EGARCH model

In recent years, much attention has been focused on modelling financial-market returns by processes other than simple Gaussian white noise. To capture the property of time varying volatility, Engle introduced the AutoRegressive Conditional Heteroskedasticity (ARCH) model [23]. Bollerslev's extension of this model, the GARCH model is often used for modelling stochastic volatility in financial time series [23]. Although GARCH models give adequate fits for dynamics, these models often fail to perform well in modelling the volatility of stock returns since GARCH models assume that there is a symmetric response between volatility and returns. Therefore, they are not able to capture the leverage effect of stock returns. In order to model asymmetric variance effects between positive and negative asset returns, Nelson introduced the EGARCH model [23].

Let $x_t = \mu + a_t$ be the time series value at time t, where μ is the mean of the GARCH model and a_t is the model's residual at time t. Additionally, $a_t = \sigma_t \varepsilon_t$ in which σ_t is the conditional volatility at time t that satisfies in

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2, \tag{4}$$

where p is the order and $\alpha_0, \ldots, \alpha_p$ are the coefficient parameters of ARCH component. Also q is the order and $\beta_0, \ldots, \beta_q$ are the coefficient parameters of GARCH component. Moreover, $\{\varepsilon_t\}$ is an iid sequence of residuals that approximates the measurement error sequence under the assumption that they are normally distributed with zero mean and constant variance.

2.3 RFSV model

Recent developments in volatility modeling have introduced RFSV as a paradigm-shifting approach to capturing the intrinsic roughness observed in financial time series. Unlike traditional stochastic volatility models that assume smooth sample paths, RFSV models employ fractional Brownian motion with

Hurst exponent $H < 0.5$ to characterize the irregular, fractal-like behavior of volatility dynamics [10]. This framework fundamentally differs from Heston-type models by recognizing that volatility exhibits roughness at all time scales, as demonstrated empirically across thousands of US equities [1]. The RFSV model is a simple model under the form

$$\log \sigma_t \approx \nu W_t^H + C. \tag{5}$$

The forecasting method is based on the following formula

$$\mathbb{E}[W_{t+\Delta}^H | \mathscr{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^{t} \frac{W_s^H}{(t-s+\Delta)(t-s)^{H+1/2}} ds, \tag{6}$$

where $\mathscr{F}_t$ is the filtration generated by W_t^H. By construction over the reasonable timescale of interest, the forecasting formula for log-variance of the RFSV model follows as

$$\mathbb{E}[\log \sigma^2{}_{t+\Delta} | \mathscr{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^{t} \frac{\log \sigma^2{}_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds. \tag{7}$$

Building on the models described above, this study offers the following key contributions to implied volatility forecasting, with an emphasis on novel aspects that distinguish our approach from prior research:

- Direct multi-model empirical comparison: Unlike studies focusing on single models or limited pairings, we provide a rigorous evaluation of EGARCH, NARX, LSTM, and RFSV using a long-term S&P500 implied volatility dataset (2001–2024). This underexplored timeframe facilitates a robust evaluation across varied market regimes, notably encompassing periods of extreme volatility such as the 2008 financial crisis and the 2020 pandemic-induced downturn.

- Model enhancements and novel integrations: We modify standard architectures to better suit financial data. For LSTM, a multilayer design incorporates implied volatility lags and historical rolling volatility inputs, surpassing basic applications in capturing non-linear, long-term dependencies. The NARX innovatively includes trading volume as an exogenous input for liquidity effects, while RFSV features empirical Hurst parameter estimation, extending descriptive analyses to predictive contexts.

- Addressing literature gaps from systematic reviews: Given the scarcity of implied volatility datasets and the limited exploration of advanced neural architectures in volatility modeling, our study addresses these gaps by prioritizing implied volatility, an inherently rarer and more informative metric than historical proxies. We combine LSTM with RFSV and EGARCH approaches to establish a comprehensive benchmark for shared tasks. The results demonstrate LSTM's advantage in capturing non-linear patterns, outperforming traditional GARCH-MLP hybrids in complex market environments.

- Practical insights for risk management: Identifying LSTM as the superior model yields actionable guidance for option pricing and hedging, contrasting theoretical validations with empirical applicability on a major index like the S&P500.

These contributions underscore the novelty of our comparative framework, validating trends while pioneering methodological advances for more precise volatility forecasting.

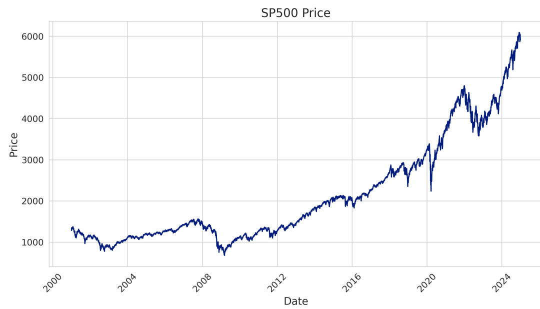

Figure 2: Sample path of price

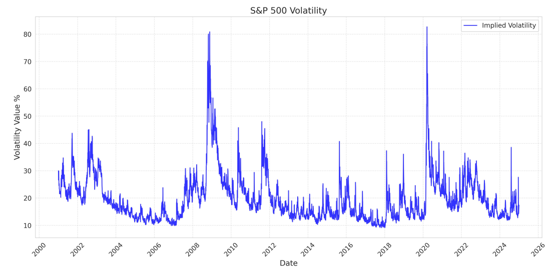

Figure 3: Implied volatility

3 Data analysis

In this section, we begin by providing a detailed description of the dataset, outlining its key characteristics. To illustrate the distributional properties of implied volatility, we present both histogram and kernel density plots, offering visual insights into its behavior across the sample period. Then, we provide a comprehensive overview of the data sources employed in the analysis, along with a detailed explanation of the methodologies used to monitor volatility dynamics. Furthermore, we assess the performance and predictive capabilities of four distinct modeling approaches: EGARCH, NARX, LSTM, and RFSV, highlighting their respective strengths and limitations in capturing the complex behavior of implied volatility.

3.1 Data description

The dataset used in this analysis was obtained from the Federal Reserve Economic Data (FRED) database, specifically the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) series available at https://fred.stlouisfed.org/series/VIXCLS. This index reflects the market's expectation of near-term volatility as implied by S&P 500 stock index option prices. The daily frequency series capturing closing values, spans from January 2, 2001 to July 24, 2024, and is not seasonally adjusted.

Figures 2 and 3 present the daily closing values of the stock index and the corresponding option-implied volatilities, respectively. These visualizations offer a comparative view of market price movements alongside investor expectations of future volatility. For each index, the volatility clustering effect is confirmed as well as the well-documented leverage effect [4]. This asymmetry highlights the non-linear relationship between volatility and market returns. When stock prices fall, volatility typically increases and vice versa.

To offer a comprehensive overview of the dataset's characteristics, Table 2 presents key descriptive statistics. The implied volatility data exhibits a mean of 19.73% and a standard deviation of 8.58%, reflecting moderate variability over the sample period. These metrics suggest a dynamic but not excessively volatile market environment, consistent with historical patterns observed in the VIX index. The dataset comprises 6,036 daily observations, spanning approximately 23 years of trading activity after accounting for market holidays. The distribution of implied volatility exhibits a positive skewness of 2.25, indicating a right-tailed structure with a tendency toward higher than average volatility spikes. Additionally, the excess kurtosis of 8.08 reflects the presence of fat tails, suggesting a heightened probability of extreme volatility events. These characteristics are consistent with well-documented behaviors in financial time

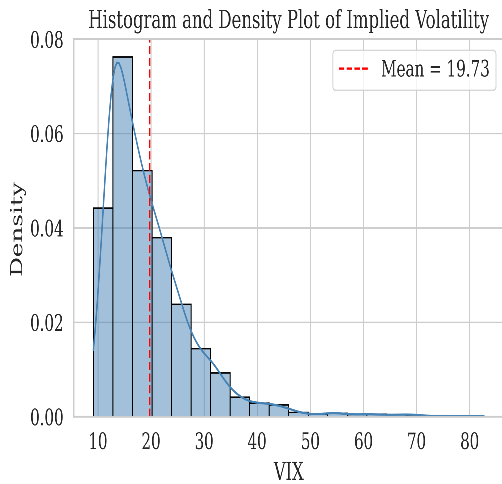

Figure 4: Histogram and density plot of implied volatility

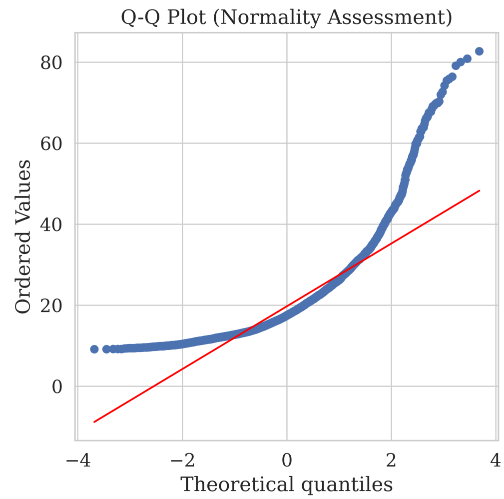

Figure 5: Q-Q plot of implied volatility (Normality check)

series, particularly in markets which have sudden regime shifts and crisis-driven dynamics.

Table 2: Descriptive statistics of S&P500 VIX dataset

Number of observations	Mean	Standard deviation	Skewness	Excess kurtosis
6036	19.73	8.58	2.25	8.08

Figure 4 displays the histogram and kernel density plot of implied volatility values, revealing a right-skewed distribution with a pronounced peak near the mean of 19.73%. Figure 5 presents the Q-Q plot against a theoretical normal distribution, where noticeable deviations in the tails underscore the data's non-normality and fat-tailed characteristics. These distributional features are critical for volatility modeling, as they reflect the presence of extreme events and challenge assumptions underlying traditional Gaussian-based approaches.

3.2 Methodology

This study utilizes S&P500 implied volatility data spanning from 2001 to 2024 to develop and evaluate time series forecasting models. The dataset was divided into training and testing sets, with 70% of the data allocated for training purposes and the remaining 30% reserved for testing the model's forecasting capabilities. For the LSTM model, a multilayer architecture was employed. The model's input features included implied volatility lags and historical rolling volatility (HRV). In contrast, the NARX model was constructed with 23 neurons. The inputs for this model consisted of implied volatility lag, and trading volume as an exogenous variable. The inclusion of trading volume was intended to provide additional market context, potentially enhancing the model's understanding of the underlying dynamics and improving its forecasting performance. Both the LSTM and NARX models were trained using appropriate loss functions and optimization techniques [15] to ensure accurate forecasting of the implied volatility

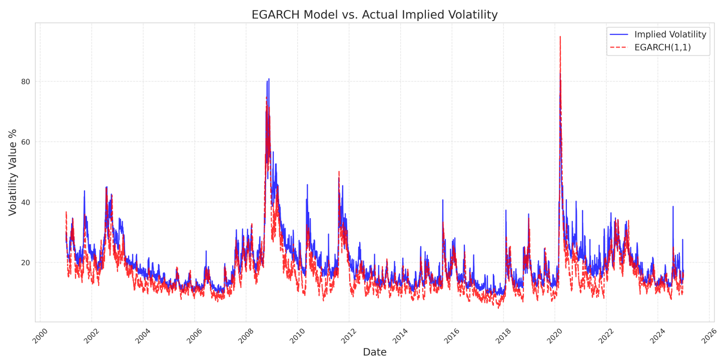

Figure 6: Tracking implied volatility using the EGARCH model

of S&P500 options ATM with 30 days expiry during the testing phase. Figure 7 illustrates the fitted NARX model's performance in forecasting implied volatility, showcasing both the training set and test set results. The model successfully captures the underlying trends and fluctuations in implied volatility across the time series, demonstrating its effectiveness in non-linear prediction. Moreover, Figure 8 depicts the performance of the LSTM in forecasting implied volatility and shows the fitted values on the training and test set, illustrating the model's capability to capture complex patterns and temporal dependencies in the volatility data. The training and evaluation of the models were conducted using established machine learning libraries and frameworks. Additionally, for the RFSV model, it is necessary to estimate the Hurst parameter from the time series data. There are numerous methods to estimate the Hurst parameter. As suggested by Gatheral [10], we employ linear regression to estimate $H = 0.29$ for our dataset. Figure 9 displays the accuracy of the RFSV model in forecasting volatility compared to EGARCH model. It shows that although its parametric characteristics, its performance is comparable to that of other non-parametric models.

4 Performance evaluation

To evaluate the performance of the EGARCH, NARX, LSTM and RFSV models, we employed several statistical metrics that are commonly used for assessing forecasting accuracy. The selected metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage error (MAPE). These metrics provide a comprehensive understanding of model performance by quantifying the errors in the predictions relative to the actual observed values.

The MAE measures the average absolute differences between predicted and actual values, providing

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t|, \tag{8}$$

where Y_t is the observed implied volatility (actual value) and $\hat{Y}_t$ is the estimated volatility (predicted value) by the models. As another error measure, we present MSE that squares these differences before averaging:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2. \tag{9}$$

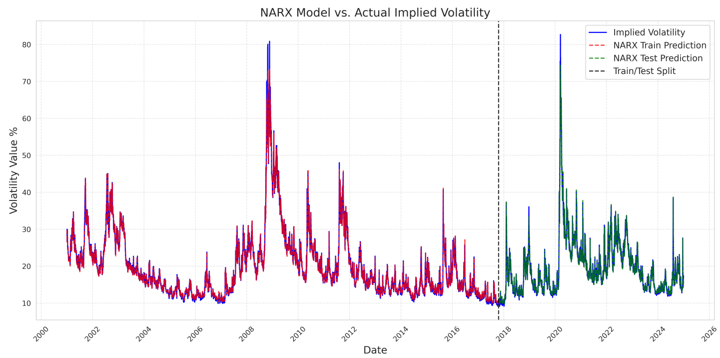

Figure 7: Tracking implied volatility using the NARX model

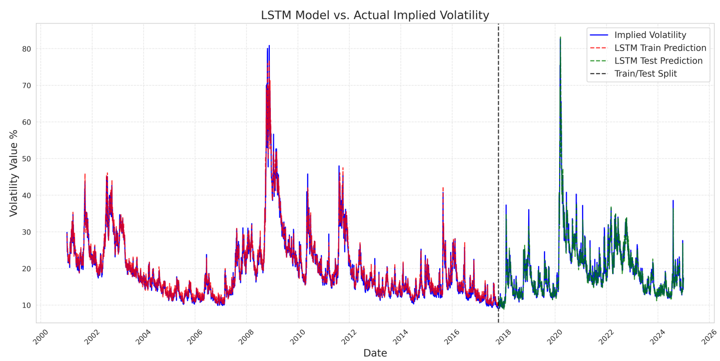

Figure 8: Tracking implied volatility using the LSTM model

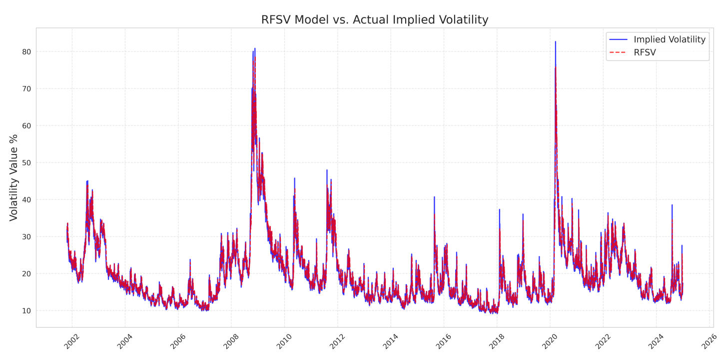

Figure 9: Tracking implied volatility using the RFSV model

That means, it penalizes larger errors more heavily, making it useful for assessing model performance when outliers are present.

The RMSE is simply the square root of MSE as $RMSE = \sqrt{MSE}$, offering the error metric in the same units as the original data, thus making it more interpretable.

Our empirical analysis provides clear evidence that the LSTM model outperforms the EGARCH,

Table 3: Comparison of error measures for different models in forecasting implied volatility

Model	MSE	RMSE	MAE	MAPE
EGARCH	25.33	5.03	3.90	0.19
RFSV	5.72	2.39	1.51	0.07
LSTM	**1.11**	**1.05**	**0.68**	**0.03**
NARX	2.81	1.67	1.01	0.05

Table 4: Comparison of average error measures from 5-fold cross- validation

Model	MSE	RMSE	MAE	MAPE
EGARCH	36.23	5.81	4.78	0.399
RFSV	6.02	2.43	1.51	0.072
LSTM	**5**	**1.92**	**0.92**	**0.03**
NARX	6.06	2.29	1.28	0.039

NARX and RFSV models in tracking the implied volatility of the S&P500 index. As presented in Table 3, the LSTM model achieved lower error metrics across all measures. These results highlight the LSTM model's superior ability to capture the complex temporal dependencies and non-linear patterns inherent in financial time series data, more effectively than the NARX model. The lower error rates indicate that the LSTM network provides a more accurate and reliable tool for forecasting implied volatility, which is crucial for making informed decisions in financial markets.

The implications of these findings are significant for practitioners and researchers in finance. By adopting advanced deep learning techniques like LSTM networks, market participants can enhance their volatility forecasting capabilities, leading to improved risk management and more strategic investment decisions. This study highlights the potential of leveraging cutting-edge neural network architectures to gain a competitive edge in the dynamic landscape of financial markets.

4.1 Cross-validated performance for robustness

To further validate our initial findings and ensure the results are robust, we conduct a more rigorous evaluation using a 5-fold expanding window cross-validation, as detailed in Subsection 3.2. This method provides a more reliable estimate of model performance by testing on multiple, distinct periods of the time series. The aggregated performance metrics, averaged across the five test folds, are presented in Table 4. These results confirm the conclusions drawn from the initial train-test split: the LSTM model demonstrates superior forecasting accuracy.

As shown in Table 4, the LSTM model achieves the lowest average RMSE of 1.92, MAE of 0.92, and MAPE of 3.93%. This consistent outperformance across multiple error metrics underscores its reliability. While its average MSE is slightly higher than some models, its overall performance, particularly on metrics less sensitive to outliers like MAE and RMSE, is clearly the strongest. This robust cross-validation process provides stronger evidence that LSTM networks offer a superior approach for forecasting implied volatility compared to the EGARCH, RFSV, and NARX models.

A deeper analysis of the performance on each individual fold reveals interesting learning dynamics between the models. Data intensive models like LSTM and NARX show a clear trend of improvement as the training dataset expands. For instance, their MSE is highest on the first fold and decreases substantially in subsequent folds as more historical data becomes available for training. This suggests that these complex models are highly effective at leveraging larger datasets to improve their predictive accuracy. In contrast, the RFSV models do not exhibit a similar learning curve. Their performance across the folds is more varied and does not show a consistent trend of improvement with more data. While the LSTM model's superior average performance makes it the best overall choice, this fold-wise analysis indicates its advantage is most pronounced in data-rich environments. This highlights the scalability of deep learning approaches for financial forecasting tasks where large historical datasets are often available.

4.2 Comparison with existing literature

The results in Table 3 align with established trends in the literature, where machine learning models like LSTM and NARX often outperform traditional econometric methods such as EGARCH in volatility prediction tasks. For instance, DEcclesia and Clementi [5] demonstrated that ANN-based models, including LSTM variants, yield lower RMSE in tracking implied volatility dynamics compared to Heston or EGARCH frameworks, attributing this to their ability to model non-linear dependencies. Similarly, Kim and Won [17] demonstrated the superior performance of LSTM in financial time series forecasting, reporting RMSE improvements of 10–20% over traditional GARCH models. This advantage is attributed to LSTM's ability to capture long-term sequential dependencies, a factor that aligns with our findings, where the multilayer LSTM architecture with lagged inputs shows strong predictive capability on the extended S&P 500 dataset.

The inclusion and moderate performance of RFSV reflect current methodological trends, corroborating studies like Bennedsen et al. [1], who found RFSV effective in capturing roughness ($H < 0.5$) in large scale stock volatility data, leading to better medium-term forecasts than smooth stochastic models. Our empirical Hurst estimation enhances RFSV's predictive utility, aligning with Gatheral et al. [10], who showed rough volatility's universality in improving model fit. However, our outcomes also deviate in nuanced ways from some prior work. While Poon and Granger [20], in their comprehensive review, noted that GARCH variants, including EGARCH, can occasionally match or outperform simpler machine learning models in short-horizon forecasts, particularly under stable market conditions, our results reveal a more pronounced advantage for advanced machine learning approaches. Specifically, LSTM achieves a 67% reduction in RMSE compared to EGARCH, underscoring its effectiveness in capturing complex volatility dynamics.

This deviation likely stems from our longer dataset (2001–2024), which includes multiple crisis periods (e.g., 2008 financial crisis, 2020 COVID-19 volatility spikes), amplifying non-linearities and long dependencies that econometric models struggle with, as noted in Zhao et al. [24]. In contrast, studies on shorter horizons, like D'Ecclesia and Clementi [5], report closer competitions between NARX and EGARCH, but our extended timeframe reveals greater LSTM gains, possibly due to regime shifts and volatility clustering not fully captured by parametric assumptions. Challenging perspectives are offered by Euch and Rosenbaum [8], who highlight the limitations of rough volatility models in highly non-stationary environments. They suggest that, without methodological refinements or hybrid integrations, such models may underperform relative to data-driven approaches. This insight partially accounts for the observed underperformance of RFSV compared to LSTM in our evaluation metrics. Overall, our findings reinforce the growing dominance of machine learning techniques in non-linear forecasting [24], while also highlighting performance variations influenced by dataset length and model-specific enhancements. These results point to promising directions for future research, particularly the development of hybrid approaches such as LSTM-RFSV combinations that may offer improved robustness across diverse market conditions.

5 Conclusion

This study conducted a comparative analysis of EGARCH, NARX, RFSV, and LSTM models for forecasting the implied volatility of the S&P 500 index. Our empirical results, validated through a robust 5-fold expanding window cross-validation, demonstrate that the LSTM model provides the most accurate and reliable forecasts. It consistently outperformed the other models across key error metrics, including RMSE, MAE, and MAPE. These findings highlight the architectural advantages of LSTM networks in effectively capturing the complex, non-linear temporal dependencies inherent in financial volatility data. This research underscores the significance of model selection in financial time series analysis and showcases the potential for deep learning approaches to enhance forecasting capabilities, leading to improved risk management and more strategic investment decisions. While this study provides valuable insights, it is important to acknowledge its limitations, which also present clear avenues for future research. First, our analysis is focused on four specific modeling paradigms. Future work could expand this comparison to include other advanced architectures. For instance, Transformers may be better suited to capture longer-term dependencies in the data, while GNNs could model volatility spillovers between correlated assets, a dimension not explored in this study. Second, the scope of our research is confined to the implied volatility of the S&P 500 index. The findings may not be directly generalizable to other asset classes, such as individual stocks, commodities, or cryptocurrencies, which exhibit different volatility dynamics. These limitations highlight promising directions for subsequent research. Future studies may explore integrating additional variables and testing these alternative deep learning architectures to further improve predictive performance and expand the understanding of time series dynamics in the ever-evolving landscape of financial markets.

Acknowledgements

The authors are grateful to the editors and referees for their helpful comments and valuable remarks that help to improve the paper.

References

[1] M. Bennedsen, A. Lunde, M.S. Pakkanen, *Decoupling the short-and long-term behavior of stochastic volatility*, J. Financ. Econ. **20** (2022) 961–1006.

[2] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, J. Econom. **31** (1986) 307–327.

[3] B.J. Christensen, N.R. Prabhala, *The relation between implied and realized volatility*, J. Financ. Econ. **50** (1998) 125–150.

[4] E. Daal, A. Naka, J.S. Yu, *Volatility clustering, leverage effects, and jump dynamics in the US and emerging Asian equity markets*, J. Bank. Finance **31** (2007) 2751–2769.

[5] R. D'Ecclesia, D.M. Clementi, *Volatility in the stock market: ANN versus parametric models*, Ann. Oper. Res. **299** (2021) 1101–1127.

[6] A. Deng, B. Hooi, *Graph neural network-based anomaly detection in multivariate time series*, Proceedings of the AAAI Conference on Artificial Intelligence **35** (2021) 4027–4035.

[7] R.G. Donaldson, M. Kamstra, *An artificial neural network—GARCH nodel for international stock return volatility*, J. Empir. Finance. **4** (1997) 17–46.

[8] O.E. Euch, M. Rosenbaum, *The characteristic function of rough Heston models*, Math. Financ. **29** (2019) 3–38.

[9] S. Fallah, F. Mehrdoust, *A study of American options under stochastic volatility and double exponential jumps*, J. Math. Model. **13** (2025) 467–484.

[10] J. Gatheral, T. Jaisson, M. Rosenbaum, *Volatility is rough*, Quant. Finance **18** (2018) 933–949.

[11] A. Goyal, A. Saretto, *Cross-section of option returns and volatility*, J. Financ. Econ. **94** (2009) 310–326.

[12] D.D. Hawley, J.D. Johnson, D. Raina, *Artificial neural systems: A new tool for financial decision making*, Financ. Anal. J. **46** (1990) 63–72.

[13] S.L. Heston, *A closed-form solution for options with stochastic volatility*, Rev. Financ. Stud. **6** (1993) 327–343.

[14] S. Hochreiter, J. Schmidhuber, *Long short-term memory*, Neural Comput. **9** (1997) 1735–1780.

[15] X. Huang, H. Cao, B. Jia, *Optimization of Levenberg Marquardt algorithm applied to nonlinear systems*, Processes **11** (2023) 1794.

[16] M. Kazemi, H. M. Peyhani, M. Akbari, *A novel leading moving average indicator using ANFIS-Wavelet hybrid method for financial market forecasting*, J. Math. Model. **13** (2025) 775–788.

[17] H.Y. Kim, C.H. Won, *Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models*, Expert Syst. Appl. **103** (2018) 25–37.

[18] F.G. Miranda, N. Burgess, *Modelling market volatilities: the neural network perspective*, Eur. J. Financ. **3** (1997) 137–157.

[19] D.B. Nelson, *Conditional heteroskedasticity in asset returns: a new approach*, Econometrica **59** (1991) 347–370.

[20] S.H. Poon, C.W.J. Granger, *Forecasting volatility in financial markets: A review*, J. Econ. Lit. **41** (2003) 478–539.

[21] A. Stokes, A.S. Abou-Zaid, *Forecasting foreign exchange rates using artificial neural networks: a trader's approach*, Int. J. Monet. Econ. Finance **5** (2012) 370–394.

[22] P. Thakkar, K. Chaudhari, *A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions*, Expert Syst. Appl. **177** (2021) 114800.

[23] R.S. Tsay, *Analysis of Financial Time Series*, John Wiley & SONS, 2005.

[24] P. Zhao, H. Zhu, W.S.H. NG, D.L. Lee, *From GARCH to neural network for volatility forecast*, Proceedings of the AAAI Conference on Artificial Intelligence **35** (2024) 16998–17006.