An Artificial Neural Networks Approach in Estimating Implied Volatility

Navideh Modarresi* Reza Kazemi† Alireza Mousavi‡

Abstract

Implied volatility is a crucial indicator in financial markets, representing the market's expectations of future volatility and serving 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 aiming to optimize returns and mitigate risks. This paper explores the effectiveness of different modeling approaches for tracking the implied volatility of the S&P500 index, focusing specifically on a comparison of exponential autoRegressive conditional heteroskedasticity (EARCH), long short-term memory (LSTM) neural networks and Nonlinear autoregressive with exogenous input (NARX) models, both types of artificial neural networks. Our empirical study shows that the LSTM model improves our estimation over NARX model.

Keywords: Implied volatility, LSTM neural network, NARX model

1 Introduction

Estimating the volatility of equity returns with high accuracy is pivotal for effective portfolio selection, risk management, and the development of trading strategies. Volatility, characterized by the level of price changes of an asset over time, is a critical factor in determining derivative pricing and evaluating market risk. Among various measures, implied volatility is particularly important as it reflects market expectations of future volatility derived from option prices [14].

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 movement of the underlying asset's price and predicts the extent of potential 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 [2]. Thereby, it helps the practitioners and investors anticipate market movements and inform trading strategies [3].

However, accurately calibrating implied volatility, which significantly affect option pricing remains, a major challenge in finance. Precise calibration 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 [8] 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.

Econometric models like the generalized autoregressive conditional heteroskedasticity (GARCH) model [1] and its extension, the exponential GARCH (E-GARCH) model [13], have been widely used to model time-varying volatility. However, these models often struggle to capture the nonlinear and complex dynamics inherent in financial time series, limiting their effectiveness in tracking implied volatility [3]. Stochastic volatility models, such as Heston's model [10], also seek to account for aspects like mean reversion and the correlation between asset returns and volatility. Heston as well as Bates model yields semi-closed form solutions for European option prices in terms of Fourier transforms making them relatively easy calibration to market data.

In recent years, artificial neural networks (ANNs) have gain popularity due to their capacity to model complex nonlinear relationships and detect detailed patterns within data [9], [16]. The nonlinear 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 nonlinear dynamic systems, including many financial applications. For example, Clementi [3] 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 [15] showed the effectiveness of ANNs in forecasting exchange rates. The models 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.

*Department of Statistics, Mathematics and Computer, Allameh Tabataba'i University, n.modarresi@atu.ac.ir
†rkazemi453@gmail.com
‡ar.mousavie@gmail.com

Building on NARX model successes, innovations in neural network architectures, most notably long short-term memory (LSTM) networks [11] have emerged. 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. DEcclesia 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 (RMSE). Despite these advancements, direct comparisons between EGARCH, NARX and LSTM models concerning the estimation of implied volatility are scarce in the literature, highlighting a critical research gap. This paper aims to fill that gap by identifying the most effective method for measuring equity returns and accurately tracking options implied volatility. Using S&P500 option data and equity price data from 2011 to 2018, this study seeks to provide insights valuable for both practitioners and researchers interested in enhancing risk management strategies and improving trading decisions.

2 Volatility estimation

In this section, we explore the estimation of equity returns volatility by evaluating various modeling techniques. The nonlinear autoregressive model with exogenous inputs (NARX) is assessed against established financial models, specifically the Heston model and the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model. Previous research, notably by Clementi [3], indicates that NARX outperforms both the Heston and EGARCH models in forecasting implied volatility. This finding highlights the NARX models 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:

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

where $r_t = \ln \frac{S_t}{S_{t-1}}$. In general market participant are used to deal with yearly volatility which is given by $\hat{\sigma_t} = \sqrt{\sigma_t^2 \cdot 252}$.

2.1 Historical rolling volatility

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

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

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.

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 (1982) introduced the AutoRegressive Conditional Heteroskedasticity (ARCH) model. Bollerslev's (1986) extension of this model, the Generalised ARCH (GARCH) model is often used for modelling stochastic volatility in financial time series. Although GARCH models give adequate fits for dynamics, these models often fail to perform well in modelling the volatility of stock returns because 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 (1991) introduced the Exponential GARCH (EGARCH) model [17].

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, $a_t = \sigma_t \epsilon_t$ in which σ_t is the conditional volatility at time t as

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i(|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}) + \sum_{j=1}^{q}\beta_j \ln \sigma_{t-j}^2. \quad (3)$$

where p is the order and $\alpha_0, \ldots, \alpha_p$ are the coefficient parameters of ARCH component and also q is the order and $\beta_0, \ldots, \beta_q$ are the coefficient parameters of

GARCH component. Moreover, $\{\epsilon_t\}$ 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 NARX model

Recent developments in the finance sector have sparked significant interest in a new class of nonlinear models inspired by the structure of the human brain, commonly referred to as Artificial Neural Networks (ANNs). ANN techniques have been extensively utilized for forecasting stock prices and historical volatility [6], [7]. In this study, we propose an innovative approach for modeling stock return volatility by leveraging machine learning and signal processing methodologies, particularly through the application of the Nonlinear Autoregressive model with Exogenous Inputs (NARX). The defining equation for the NARX model is

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-n}, u_{t-1}, u_{t-2}, \ldots, u_{t-n}). \quad (4)$$

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 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.

2.4 LSTM model

To process sequential data, we utilize Long Short-Term Memory (LSTM) networks, a sophisticated variant of recurrent neural networks (RNNs). While RNNs are

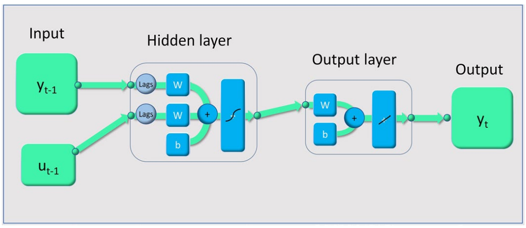

Figure 1: NARX neural network

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. 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 [11]. 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 too model the implied volatility of return.

3 Data analysis

In this section, we detail the data sources utilized for the analysis and the methodologies implemented to track volatility and evaluate the effectiveness of the EGARCH, NARX and LSTM models. In Figures 2 and 3 the daily stock index quotes and the corresponding option implied volatilities are reported. 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 the stock prices fall, volatility typically increase and vice versa.

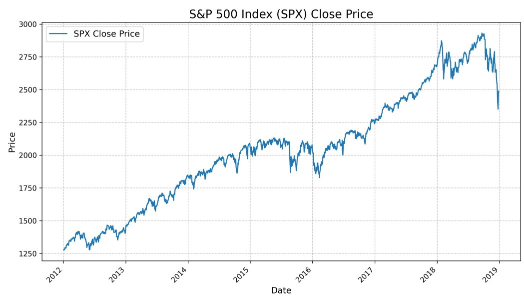

Figure 2: Sample path of price

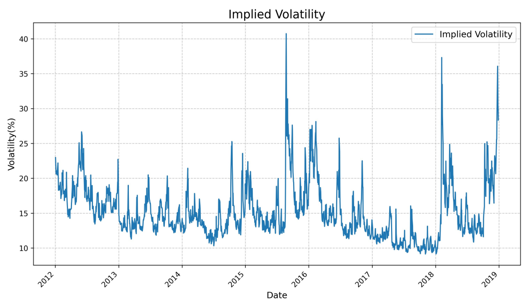

Figure 3: Implied volatility

3.1 Methodology

This study utilizes S&P500 option data spanning from 2011 to 2018 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 models' predictive capabilities. For the Long Short-Term Memory (LSTM) model, a multi-layer architecture was employed. The model's input features included implied volatility (IV) lags and historical rolling volatility (HRV). In contrast, the Nonlinear Autoregressive Exogenous (NARX) model was constructed with 23 neurons. The inputs for this model consisted of IV lags, 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 predictive performance. Both the LSTM and NARX models were trained using appropriate loss functions and optimization techniques [12] to ensure accurate forecasting of the S&P500 options ATM with 30 days expiry during the testing phase. Figure 5 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 nonlinear prediction. Moreover, Figure 6 depicts the performance of the LSTM in forecasting implied volatility and shows the fitted values on the training ant 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.

3.2 Performance evaluation

To evaluate the performance of the EGARCH, NARX and LSTM 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), and Root Mean Squared Error (RMSE). These metrics provide a comprehensive understanding of model performance by quantifying the errors in the predictions relative to the actual observed values.

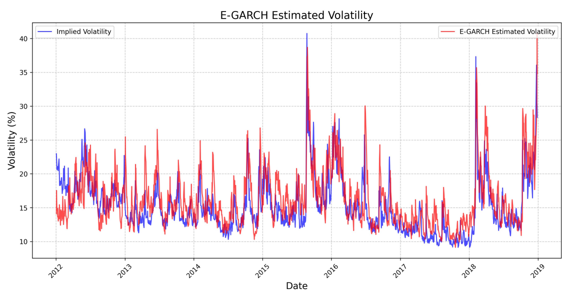

Figure 4: Tracking implied volatility using the EGARCH model

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

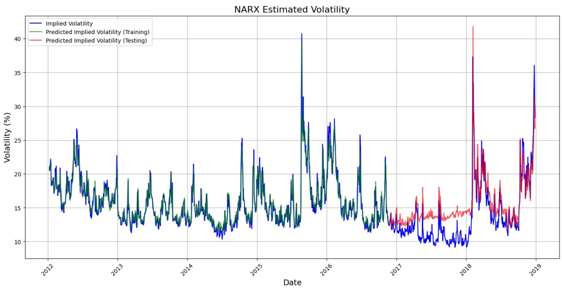

Figure 5: Tracking implied volatility using the NARX model

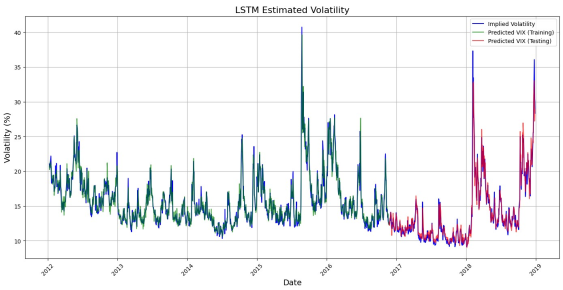

Figure 6: Tracking implied volatility using the LSTM model

a straightforward interpretation of error magnitude as

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \tag{5}$$

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

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2, \tag{6}$$

and 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 and NARX models in tracking the implied volatility of the S&P500 index. As presented in Table 1, 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 nonlinear patterns inherent in financial time series data, more effectively than the NARX model. The lower

	RMSE	MSE	MAE
EGARCH	1.491	2.223	1.440
NARX	0.018	0.033	0.012
LSTM	0.007	0.005	0.005

Table 1: Comparison of error measures for EGARCH, NARX and LSTM models in forecasting implied volatility

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 conclusion

In conclusion, this study has highlighted the effectiveness of the LSTM model in comparison to the EGARCH and NARX models for time series forecasting tasks. The performance metrics obtained demonstrate that the LSTM model outperforms the other models across all evaluated criteria. These findings indicate that the LSTM model's architectural advantages, particularly its ability to effectively capture temporal dependencies, contribute to its superior predictive accuracy. This research shows the significance of model selection in time series analysis and suggests that deep learning approaches, such as LSTM, can enhance forecasting capabilities. Future studies may explore integrating additional variables and testing alternative models to further improve predictive performance and expand the understanding of time series dynamics.

References

[1] T. Bollerslev Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3): 307-327, 1986.

[2] B.J. Christensen and N.R. Prabhala The Relation between Implied and Realized Volatility. *Financial Management*, 27(1): 25-45, 1998.

[3] F. Clementi Neural Networks for Forecasting Implied Volatility in Financial Markets. *International Journal of Forecasting*, 35(3): 1-15, 2019.

[4] E. Daal, A. Naka and J.S. Yu Volatility clustering, leverage effects, and jump dynamics in the US and

emerging Asian equity markets. *Journal of Banking & Finance*, 31(9): 2751-2769, 2007.

[5] R. DEcclesia and D.M. Clementi Volatility in the stock market: ANN versus parametric models. *Annals of Operations Research*, 299(1): 1101-1127, 2021.

[6] R.G. Donaldson and M. Kamstra An Artificial Neural NetworkGARCH Model for International Stock Return Volatility. *Journal of Empirical Finance*, 4: 17-46, 1997.

[7] F.G. Miranda and N. Burgess Modelling market volatilities: the neural network perspective. *European Journal of Finance*, 3(2): 137-157, 1997.

[8] A. Goyal and A. Saretto Cross-sectional predictability of implied volatility. *Journal of Financial Markets*, 12(4): 523-550, 2009.

[9] D.D. Hawley, J.D. Johnson and D. Raina Artificial Neural Systems: A New Tool for Financial Decision Making. *Financial Analysts Journal*, 46(6): 63-72, 1990.

[10] S.L. Heston A closed-form solution for options with stochastic volatility. *The Review of Financial Studies*, 6(2): 327-343, 1993.

[11] S. Hochreiter, J. Schmidhuber Long short-term memory. *Neural Computation*, 9(8): 17351780, 1997.

[12] X. Huang, H. Cao and B. Jia Optimization of Levenberg Marquardt Algorithm Applied to Nonlinear Systems. *Processes*, 11(6): 1794, 2023.

[13] D.B. Nelson Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica: Journal of the Econometric Society*, 59: 347-370, 1991.

[14] S.H. Poon and C.W.J. Granger Forecasting volatility in financial markets: A review. *Journal of Forecasting*, 22(1): 55-76, 2003.

[15] A. Stokes and A.S. Abou-Zaid Forecasting foreign exchange rates using artificial neural networks: a trader's approach. *International Journal of Monetary Economics and Finance, Inderscience Enterprises Ltd*, 5(4): 370-394, 2012.

[16] P. Thakkar and K. Chaudhari A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications*, 177: 114800, 2021.

[17] R.S. Tsay Analysis of Financial Time Series *John Wiley & SONS*, 2nd edition, ISBN: 0-471-690740, 2005.