Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28448

A Review of Stock Market Volatility Prediction Techniques Based on Machine Learning

Yangzhiping Chen

School of Mathematics and Economics, Sichuan University, Chengdu, China
chenyangzhiping@163.com

Abstract. Volatility is a crucial indicator for quantifying risk levels, guiding as- set allocation, and assisting in formulating macro policies, so realizing the prediction has important significance. Researchers are increasingly applying machine learning techniques, such as support vector machines, LSTM, and deep convolutional networks, to enhance the accuracy of volatility predictions in complex environments. Although numerous studies have proposed diverse methods that combine with machine learning for volatility forecasting, systematic reviews that sort out these methods remain scarce. To fill this gap, this paper systematically summarizes novel models developed by scholars for forecasting stock market volatility, aiming to provide valuable references for researchers to develop new forecasting technologies. By checking 19 empirical studies and 6 review articles, along with authoritative writings, this paper systematically organizes foundational methods related to stock market volatility prediction and demonstrates machine learning- based forecasting techniques. The review finds that machine learning methods such as support vector machine (SVM) and random forest (RF) enhance prediction accuracy by efficiently capturing the nonlinear characteristics of financial data through kernel functions and ensemble learning. Deep learning models, such as LSTM and GRU, excel in forecasting stock market volatility by capturing long-term dependencies and processing complex sequential data, significantly improving prediction accuracy compared to traditional models like GARCH. In addition, hybrid models (such as GARCH-LSTM and GARCH-MIDAS) further combine the advantages of econometrics and machine learning, and have demonstrated superiority in multiple empirical studies.

Keywords: Machine learning, Volatility prediction, Stock market

1. Introduction

Stock market volatility is an important reference index to measure the return and risk level of stock investment, and can affect individual asset allocation decisions, while reflecting the national macroeconomic trend. Relevant research has found that volatility risk helps to better understand the marginal changes in stock returns [1] and idiosyncratic volatility is positively correlated with the risk of a sudden and sharp decline in stock prices [2].Other studies have demonstrated that under certain conditions, the higher the volatility of stock returns, the higher the investability of individual stocks [3], and volatility can reduce the benefits of investment allocation or even lead to losses,

DOI: 10.54254/2755-2721/2025.LD28448

which can turn originally positive earnings into negative ones [4]. These empirical results reveal that stock market volatility significantly affects individual investment decisions. In terms of macroeconomics, increased stock price volatility is associated with a slower GDP growth trend [5]. Therefore, measuring and forecasting stock market volatility is important, and many traditional modeling methods have been developed to address this task.

Although the emergence of classical econometric models has facilitated predictive accuracy in stock markets, their methodological limitations have become increasingly evident with escalating market complexity. The ARCH and GARCH are unable to capture long-memory effects efficiently enough [6]. While the stochastic volatility model (SV) involves the intractable likelihood function and requires a lot of computation, especially the parameter estimation and inference under the actual data constraints [7]. In addition, heterogeneous autoregressive (HAR) models cannot capture nonlinear de- pendencies and market jumps [8]. To address these difficulties, volatility prediction models began to integrate different machine learning methods.

This review explores stock market volatility prediction models using different machine learning methods. The results show that volatility prediction models combining machine learning methods often have higher accuracy and thus are more valuable for practical applications. Empirical research shows that the stock market fluctuation prediction technology based on the long short-term memory model has a good predictive ability in complex markets because it can capture nonlinear features and remember the long-term dependence of historical data [9]. This conclusion also applies to the volatility prediction method combined with the support vector machine model [10]. In conclusion, the techniques for predicting volatility have evolved from traditional econometric models to hybrid models that integrate machine learning, thereby enhancing the accuracy of predictions.

2. The traditional methods for predicting stock market volatility

Forecasting volatility in the stock market involves estimating future fluctuations in stock prices and trading dynamics. Traditional stock market volatility forecasting methods often rely on statistical frameworks, econometric models, and market microstructure indicators. Below are some traditional methods commonly employed:

i.) Generalized Autoregressive Conditional Heteroskedasticity(GARCH)Model: The GARCH model assumes that current volatility depends on past error terms and past volatility, which enables it to capture more complex dynamics in market volatility. However, although the GARCH model is more flexible than the ARCH model, it still has difficulty simulating sharp fluctuating jumps and is limited in handling long memory effects [6].

ii.) Stochastic Volatility(SV)Model: The SV model is an econometric paradigm used to describe the volatility of financial assets over time, which assumes that the volatility itself adheres to an unobserved stochastic process but can be estimated using methods such as Kalman filtering. Due to the SV model's more flexible representation of volatility dynamics, it can be applied to complex market environments. However, due to complex estimation technology, the model may be challenging in the calculation, and the hypothesis may not always be able to handle long memory effect or nonlinear dependencies in the data [7].

iii.) Heterogeneous Autoregressive(HAR)Model: The HAR (Heterogeneous Autoregressive) model is a statistical model used to predict volatility, which aims to capture long-term memory effects and volatility persistence. It achieves this by assuming that volatility is driven by the volatility observed over multiple time ranges in the past. However, it assumes linearity and stationarity, which makes it less suited to capture nonlinear dependencies or sudden shifts in market dynamics [8].

DOI: 10.54254/2755-2721/2025.LD28448

These traditional methods forecast stock market volatility from different perspectives and each method has its own strengths and limitations. In the past, using these methods could help predict stock market volatility; then, investors could make reasonable investment decisions. But, it is important to note that the practical application of these methods needs to be flexibly adjusted in combination with the market environment.

3. Machine learning techniques in volatility predicting

Since traditional forecasting methods are difficult to capture high-frequency data characteristics, nonlinear relations and sudden market jump behavior, and machine learning technology is increasingly applied in the financial field,many novel models and methods combined with machine learning have been developed to better predict stock market volatility. The following will present some machine learning methods and introduce the practical ways in which machine learning methods improve stock market volatility prediction technology.

3.1. Overview of machine learning methods

i.) Support Vector Machine(SVM): SVM is a supervised learning method that aims to achieve the optimal solution for classification or regression tasks by building an optimal hyperplane to separate data points and maximize the interval between classes. Through the introduction of kernel functions, it can deal with nonlinear problems in high dimensional space, which makes it have important application potential in stock market volatility prediction. By using historical stock market data, SVM can identify complex fluctuation patterns and make effective predictions [11]. Although SVM has the ability of nonlinear modeling, it should be pointed out that SVM may not perform well in the case of noisy or overlapping data. In the stock market, data can be affected by a variety of factors, making it difficult for the model to accurately capture changes in volatility [12].

ii.) Recurrent Neural Networks(RNN): RNN is a type of neural network designed for sequence data analysis, in which the output of neurons is fed back as input to the next time step, enabling the model to retain memory. This system enables RNNS to capture time-dependent and dynamic patterns, making them suitable for time series prediction, such as stock market fluctuations. Furthermore, special variants of RNNS, such as long short-term memory (LSTM) networks and gated recurrent units (GRUs), have addressed the challenge of long-term dependence. LSTM uses memory units and gating mechanisms (input, forget, and output gates) to regulate the information while GRU uses update gates and reset gates to simplify this process, thereby reducing computational complexity while still effectively capturing time dependencies. These design makes RNN, LSTM and GRU can more accurately predict the stock market volatility and time series data [13-15]. However, RNN cannot model multi-scale temporal features, and the volatility of the stock market is affected by multiple temporal scale factors, which makes traditional RNN models possibly unable to effectively model these multi-scale temporal features [16].

iii.) Random Forests(RF): RF is an ensemble learning method that enhances the stability and accuracy of the model by constructing multiple decision trees (Each decision tree is a learner) and making final predictions using majority voting or averages. Meanwhile, by training multiple trees on different sub-samples and feature subsets, the model can effectively identify the complex patterns of stock market, especially demonstrating strong robustness when dealing with noisy data. Therefore, it is able to handle high-dimensional data, with strong nonlinear modeling capability and good resistance to overfitting [17]. However, RF is composed of a large number of decision trees and its

DOI: 10.54254/2755-2721/2025.LD28448

internal mechanism is complex, which is prone to forming a “black box” effect and makes it difficult to explain the specific reasons for each prediction result [18].

3.2. The application of machine learning

i.) Using SVM Forecasting Stock Market Volatility: In one volatility prediction method, SVM as a framework for predicting stock market volatility, maps high-dimensional market data (such as historical returns on multiple timescales) to feature space through a kernel function, and automatically captures nonlinear features of volatility (such as long-term memory effect and multi-scale correlation). Compared with the traditional GARCH model, which needs to preset the form of equations, SVM directly learns the internal rules of data, effectively solving the problem of “dimensional disaster”. In the S&P 500 index experiment, SVM (10th-order lag input) significantly outperformed the naive model, and its performance was comparable to that of the optimal GARCH variant, demonstrating its ability to efficiently extract high-dimensional data information and providing a flexible framework for volatility prediction [10].

ii.) Using RF Forecasting Stock Market Volatility: Some scholars have proposed using RF to predict the volatility of the South African stock market, which effectively captures the nonlinear characteristics of financial data by constructing multiple decision trees for ensemble learning. In addition, in this model, RF adopts a feature random selection and out-of-bag sampling mechanism, which avoids overfitting and enhances the model’s robustness. Experiments show that when RF predicts the realized volatility of the JSE Financial Index (JFIN) and the Basic Materials Index (JBIND), the R2 is as high as 97.1%, significantly outperforming artificial neural networks(ANN), especially maintaining stable prediction performance during the COVID-19 pandemic when volatility intensified. Verify the advantages of handling high-dimensional market variables in predicting volatility [19].

iii.) Using RNN Forecasting Stock Market Volatility: In a study, LSTM recurrent neural networks were applied to predict financial fluctuations in indices such as the S&P 500 and Apple. When dealing with time series data, such as past returns and volatility, LSTM outperforms traditional models like GARCH in large-interval predictions by leveraging its ability to capture long-term dependencies through gate control mechanisms [20]. Meanwhile, another study used the GRU network to predict the trading signals of stock indices such as the Hang Seng Index, the DAX Index, and the S&P 500 Index. GRU simplifies the structure while effectively processing sequence data, and uses reset and update gates to filter information, which improves the classification accuracy of signals such as multi-head or short-head [21]. These two models have significantly enhanced the reliability of predicting volatility and other related financial data by learning complex patterns from historical data.

iv.) Using Hybrid Model Forecasting Stock Market Volatility: Some scholars have proposed a hybrid model, GARCH-LSTM, which combines the GARCH-type model with the LSTM network to predict stock market fluctuations. It addresses the highly skewed distribution of volatile data by introducing the Volume-Up (VU) strategy, which uses a root-type function to transform the input distribution, moving it to the right to reduce the concentration close to zero. This method enhances the accuracy of prediction by inputting GARCH’s output into the LSTM input, especially for extreme events, resulting in a 21.03% increase in the RMSE of the S&P 500 index volatility compared to traditional hybrid models [22]. In addition, other scholars have adopted the GARCH-MIDAS model, which decomposes volatility into short-term and long-term components based on GARCH. The latter includes low-frequency macroeconomic and financial variables. For long-term components, the model integrates low-frequency macro financial variables through MIDAS filtering

DOI: 10.54254/2755-2721/2025.LD28448

and then uses the Adaptive-Lasso variable selection within a penalized likelihood framework for variable screening. This improvement reduces overfitting and enhances the out-of-sample prediction accuracy of long-term stock market fluctuations [23]. Both of these models utilize machine learning to enhance traditional econometric methods: GARCH-LSTM improves distribution processing to better detect anomalies, while GARCH-MIDAS optimizes variable inclusion to achieve robust long-term predictions.

To sum up, the SVM, RNN and hybrid models such as GARCH-LSTM and GARCH-MIDAS, by addressing the limitations of traditional models, significantly enhance the ability to predict stock market fluctuations. These methods enhance prediction accuracy by capturing nonlinear relationships, long-term dependencies, and integrating macroeconomic variables. However, there are still challenges in data quality, interpretability, model stability, generalization ability and computational complexity.

4. Challenges and future research directions

Machine learning faces some challenges in predicting stock market volatility. Noise or incomplete data in the financial market may lead to overfitting and undermine the robustness of the model. Moreover, many machine learning models lack interpretability, which limits their practical application in financial decision-making, and on the other hand, these models struggle to take into account both stability and generalization under different market conditions, which reduces their ability to adapt to sudden market changes. Finally, the high computational cost of training advanced models poses a significant obstacle to their scalability, especially for real-time prediction of large datasets [24].

In terms of future directions, cross-market forecasting and multi-asset collaborative modeling will become the focus of future research to enhance the robustness of predictions. In addition, integrating high-frequency and multimodal data, including alternative sources such as news and sentiment, can enhance the accuracy of predictions by capturing more comprehensive market dynamics [25].

5. Conclusion

In this paper, we systematically reviewed various machine learning techniques applied to stock market volatility prediction. Traditional econometric models, such as GARCH and ARCH, are useful for understanding fluctuation dynamics, but they are limited in capturing nonlinear dependencies and sudden market changes. By integrating machine learning methods, including models such as Support Vector Machine (SVM), Recurrent Neural Network (RNN), GARCH-LSTM and GARCH-MIDAS, researchers have significantly improved the prediction accuracy. These models effectively model complex market behaviors, such as long-term dependencies and nonlinear relationships, making them particularly powerful tools for predicting stock market fluctuations.

The application of machine learning techniques such as LSTM and GRU has been proven to be particularly valuable for time series data, as they can capture long-term dependencies and dynamic patterns that are difficult to model with traditional methods. In addition, a hybrid approach that combines machine learning with econometric models such as GARCH-LSTM and GARCH-MIDAS can provide more robust and accurate predictions. These advancements offer new opportunities to enhance the reliability of stock market predictions and guide investment decisions in an increasingly complex financial environment.

DOI: 10.54254/2755-2721/2025.LD28448

However, there are still significant challenges in applying machine learning to volatility prediction. Problems such as overfitting, lack of interpretability and high computational costs have hindered the practical deployment of these models. Future research should focus on enhancing the robustness of the model, incorporating high-frequency data and exploring cross-market predictions to improve the prediction accuracy under different market conditions. All in all, machine learning has great potential in improving volatility prediction and perfecting stock market prediction models.

References

[1] Pati, P. C., Rajib, P., & Barai, P. (2019). The role of the volatility index in asset pricing: The case of the indian stock market. The Quarterly Review of Economics and Finance, 74, 336–346.
[2] Cao, J., Wen, F., Zhang, Y., Yin, Z., & Zhang, Y. (2022). Idiosyncratic volatility and stock price crash risk: Evidence from china. Finance Research Letters, 44, 102095.
[3] Bae, K.-H., Chan, K., & Ng, A. (2004). Investibility and return volatility. Journal of financial Economics, 71(2), 239–263.
[4] Cavallo, E., Galindo, A., Izquierdo, A., & León, J. J. (2013). The role of relative price volatility in the efficiency of investment allocation. Journal of International Money and Finance, 33, 1–18.
[5] Bhowmik, D. (2013). Stock market volatility: An evaluation. International Journal of Scientific and Research Publications, 3(10), 1–17.
[6] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307–327.
[7] Jacquier, E., Polson, N. G., & Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. Journal of Bsiness & Economic Statistics, 20(1), 69–87.
[8] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of financial econometrics, 7(2), 174–196.
[9] Fraz, T. R., Fatima, S., & Uddin, M. (2022). Modeling and forecasting stock market volatility of cpec founding countries: Using nonlinear time series and machine learning models. JISR management and social sciences & economics (JISR-MSSE), 20(1), 1–20.
[10] Gavrishchaka, V. V., & Banerjee, S. (2006). Support vector machine as an efficient framework for stock market volatility forecasting. Computational Management Science, 3(2), 147–160.
[11] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273–297.
[12] Muhammad, D., Ahmed, I., Naveed, K., & Bendechache, M. (2024). An explainable deep learning approach for stock market trend prediction. Heliyon, 10(21).
[13] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533–536.
[14] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735–1780.
[15] Cho, K., Van Merri¨enboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078.
[16] Challet, D., & Ragel, V. (2024). Multi-timescale recurrent neural networks beat rough volatility for intraday volatility prediction. Risks, 12(6), 84.
[17] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5–32.
[18] Haddouchi, M., & Berrado, A. (2024). A survey and taxonomy of methods interpreting random forest models. arXiv preprint arXiv: 2407.12759.
[19] Lamine, D., & Brijlal, P. (2024). Forecasting stock market realized volatility using random forest and artificial neural network in south africa. International Journal of Economics and Financial Issues, 14(2), 5.
[20] Liu, Y. (2019). Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. Expert Systems with Applications, 132, 99–109.
[21] Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. Procedia computer science, 131, 895–903.
[22] Koo, E., & Kim, G. (2022). A hybrid prediction model integrating garch models with a distribution manipulation strategy based on lstm networks for stock market volatility. IEEE Access, 10, 34743–34754.
[23] Fang, T., Lee, T.-H., & Su, Z. (2020). Predicting the long-term stock market volatility: A garch-midas model with variable selection. Journal of Empirical Finance, 58, 36–49.

[24] Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics, 10(21), 2717.

[25] Palaniappan, V., Ishak, I., Ibrahim, H., Sidi, F., & Zukarnain, Z. A. (2024). A review on high-frequency trading forecasting methods: Opportunity and challenges for quantum based method. IEEE Access, 12, 167471–167488.