Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD39005

The Accuracy of Linear Models in Stock Price Volatility Prediction: An Investigation and Comparison with Artificial Intelligence Approaches

Chin Yu Ching David

Shanghai High School International Division, Shanghai, China
yuqingq0303@gmail.com

Abstract. Stock volatility forecasting is always crucial for investors and stock owners to make financial decisions, portfolio allocation, and risk management. Even though there are a lot of popular volatility forecasting models existing in the stock market, it is hard to know which one was the best depending on different situations in the market. And so, this review examines the accuracy of linear models in forecasting stock price volatility, and evaluates the performance of these models compared with relatively complicated nonlinear and artificial-intelligence-based approaches. The review analyzes previous studies on major volatility forecasting models, including ARCH, GARCH, ARIMA, and HAR-RV. It compares findings from traditional linear models with more recent research on nonlinear models, machine learning, and artificial intelligence. In addition, the review also considers how forecasting performance changes across different markets, forecasting horizons, and market conditions. The reviewed evidence showed that linear models remain competitive because they are interpretable, computationally efficient, and effective when capturing persistent volatility patterns. Recent studies indicate that machine-learning models do not consistently outperform linear models in every situation, although AI approaches may provide advantages when financial relationships are more complex or when additional data are available. Overall, this review concludes that greater model complexity does not automatically lead to a higher forecasting accuracy. Future volatility prediction may benefit from hybrid and regime-aware models that combine the transparency of linear approaches with the flexibility of artificial intelligence.

Keywords: Stock price volatility, Volatility forecast, Linear models, HAR-RV model, artificial intelligence

1. Introduction

During modern days, the stock market is viewed as a complex platform for resource allocation, where movement of the stock price reflects the market situation and investor's view. The essential to understand these dynamic movements is regarded as the concept of "volatility". The stock market volatility is important for investors because investors use them as reference to predict the tendency of the stock. Poon and Granger had emphasized that the volatility forecasting received a

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD39005

considerable attention because it could provide useful information for investors and institutions [1]. However, predicting future volatility gradually became more difficult due to constant conflicts between countries, investor behaviors, economic campaigns, and unexpected information. Therefore, it is significantly important to develop a reliable method and model for forecasting stock price volatility.

Theres a variety of statistical models being developed for predicting financial volatility. Some of the early approaches include the Autoregressive Conditional Heteroskedasticity (ARCH) model proposed by Engle, which introduced a method for modeling changing conditional variance [2]. Later in 1986, Bollerslev also extended this approach by proposing the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, which combines the previous conditional variances into the prediction process [3]. Other time-series approaches, including the Autoregressive Integrated Moving Average (ARIMA) model have also been applied to financial volatility forecasting [4]. In addition, linear models provided a relatively simple approach to volatility forecasting. For instance, Corsi proposed the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV), that uses the volatility information from different time period to predict the future realized volatility [5]. The developed model has a simpler structure, while achieve an excellent forecasting performance. Furthermore, McAleer and Medeiros compared the linear and nonlinear approaches for forecasting the realized volatility of the S&P 500 and FTSE 100 indexes, demonstrating the importance of evaluating different modeling approaches [6].

Even though the linear models developed years ago were simple, easy to understand, interpret, and compute, their forecasting accuracy may be limited when the stock markets face complex and rapid changing conditions. Therefore, this review examines the accuracy of linear models in predicting stock price volatility. Section 2 reviews the main linear approaches used in volatility forecasting, with ARCH and GARCH, followed by ARIMA and the HAR-RV model, and explains their strengths and limitations identified by the previous studies. Next, Section 3 evaluates the evidence on forecasting accuracy by comparing the linear models with nonlinear and machine-learning approaches, and consider the future role of artificial intelligence in the stock volatility prediction. Lastly, the conclusion summarizes the findings of the review and explains the significance of combining accuracy, interpretability, and practical usefulness when evaluating volatility forecasting models.

2. Linear models in stock price volatility prediction

The development of volatility forecasting models has been strongly affected by the idea that market volatility is not constant. Instead, periods of relatively high volatility and low volatility can persist over time, and this means that the information from previous periods could be useful for forecasting future changes. Early econometric models provided an important foundation for this idea, and the later models attempted to improve the forecasting accuracy without making the model complicated. In addition, among these approaches, the HAR-RV model proposed by Corsi is particularly important to this review [5] because it provides an example of how a relatively simple linear structure can achieve strong forecasting performance.

2.1. ARCH and GARCH: early approaches to changing volatility

One of the early developments in volatility modelling was the ARCH model, that was proposed by Engle [2]. The main contribution of the ARCH model was the idea that variance doesn't need to be constant over time. The conditional variance could be changed according to information from

DOI: 10.54254/2754-1169/2026.LD39005

previous periods. This was important because financial markets often experience changes of uncertainty and price variation. Therefore, treating volatility as a fixed value might not provide an accurate representation of the market's behavior.

A general ARCH(q) model can be written as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 \tag{1}$$

In this equation, σ_t^2represents the conditional variance at time t, while the squared error terms ε^2 represent the previous market shocks. The model uses past shocks in order to estimate current volatility [2]. One of the limitation of ARCH model is that a large number of lagged terms may be needed to describe the persistent volatility, which can make the model less explicit and simple.

Later, Bollerslev extended this approach by introducing the GARCH model [3]. The main development of this model was that the current conditional variance could depend on previous conditional variances, but not only previous shocks. A general GARCH(p, q) model can be written as:

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

GARCH provided a more flexible extension of the ARCH model, and demonstrated how information about the previous market conditions could be incorporated into volatility modeling. However, in this review, ARCH and GARCH are mainly important as foundations rather than as the central focus. They demonstrate the development from models that recognized changing volatility toward later approaches that attempted to model more persistent and obvious patterns in the financial volatility. These models provide an important foundation for later approaches that focus on time-series persistence and multiple forecasting periods [2, 3].

2.2. ARIMA as a linear time-series approach

Another important approach to the time-series forecasting is the Autoregressive Integrated Moving Average, or the ARIMA, model. ARIMA model combines three components, including autoregression, differencing, and moving averages. The autoregressive components uses previous values of a series, differencing could be used to make a non-stationary series more stable, and the moving average component uses previous forecast errors. This structure makes ARIMA model useful for identifying linear temporal patterns in the financial data.

An ARIMA(p, d, q) model could be represented conceptually as:

$$\varphi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t \tag{3}$$

Where B is the backshift operator, d is the number of differences used to obtain a more stable series, and the autoregressive and moving-average polynomials are represented by φ(B) and θ(B). For volatility forecasting, ARIMA is very useful when the time series contains patterns that can be represented through linear dependence on previous observations and errors.

Rubio, Palacio Pinedo, Mejía Castaño, and Ramos examined ARIMA and GARCH methods for volatility forecasting, and they emphasized that these traditional models have relatively low computational requirements [4]. Their study also noted a limitation of ARIMA: since its linear structure, it cannot fully capture nonlinear patterns in a volatility series. This limitations encouraged the development of hybrid approaches that combined linear time-series models with methods

DOI: 10.54254/2754-1169/2026.LD39005

designed to represent nonlinear behaviors [4]. Therefore, ARIMA remains useful as a clear and interpretable benchmark, while its performance depends on the characteristics of the data, as well as the complexity of the underlying volatility process.

2.3. The HAR-RV model as a relatively simple linear approach

A major development in volatility forecasting was the HAR-RV model by Corsi [5]. Unlike the approaches that focuses only on information from a single previous period, this model considers volatility that is measured over different time periods. This model was developed by the idea that the financial markets contained "heterogeneous" participants whose decisions were based on different investment horizons.

Based on this idea, the HAR-RV model uses daily, weekly, and monthly volatility components to forecast the future realized volatility. Its general structure could be represented as:

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w RV_t^{(w)} + \beta_m RV_t^{(m)} + \varepsilon_{t+1} \tag{4}$$

The model does not rely only on yesterday's volatility. Instead, it combined short-, medium-, and longer-term information. This is an important feature since financial volatility can demonstrate persistency, meaning that information from previous periods may continue to influence future volatility.

Corsi's main contribution was to show that information from different time periods could be incorporated into the simple autoregressive-type model. His paper describes HAR-RV as a model that considers volatility components realized over different periods. Since its relatively simple, the model was designed to reproduce important empirical features of financial returns, and this includes the persistent volatility behavior. Corsi's empirical results also showed remarkably good forecasting performance.

However, the simplicity of HAR-RV model may create limitations. The model represents the relationship between previous volatility measures and future volatility through a relatively stable linear structure. However, the financial markets can experience sudden changes caused by economic events, investor behavior changes, or other unexpected information. During these periods, the relationship between past and future volatility may become more complex. Therefore, the strong forecasting performance reported by Corsi does not mean that a linear model will always be the most accurate approach under every market condition.

Recent studies therefore continue to use HAR-type models as an important benchmarks when evaluating nonlinear and machine-learning methods [7-9].

2.4. The values and limitations of linear forecasting

The studies discussed above demonstrate why linear approach remain important in the volatility forecasting. First, these models are based on the idea that historical market information could provide useful information about future volatility. Both ARCH and GARCH established the importance of modeling changing conditional variance, while HAR-RV showed that information from different time periods can be incorporated into a relatively simple autoregressive framework.

Second, the forecasting accuracy should not be considered independently from the conditions under which a model is tested. Poon and Granger [1], in their review of volatility forecasting research, examined a wide range of method and it showed the diversity of approaches used to forecast financial volatility. Their review demonstrates that the forecasting performance can depend

DOI: 10.54254/2754-1169/2026.LD39005

on the volatility measure, the available data that can access, the forecasting horizon, and the method used to evaluate the accuracy. Therefore, it is relatively hard to conclude that one model will be superior in every situation.

Overall, the existing research provides evidence that the simple model can generate useful forecasts of financial volatility. In particular, Corsi's HAR-RV model provides strong evidence that a relatively simple autoregressive structure can achieve a good forecasting performance. However, the accuracy of linear models may depend on whether the relationships in the market can be sufficiently represented by linear patterns. The next section will evaluate the empirical evidence on forecasting accuracy and considers how linear models perform when compared with a nonlinear and artificial-intelligence-based approaches.

3. Discussion

The research reviewed in the Section 2 suggested that linear and other traditional econometric models continue to have an important role in stock price volatility forecasting. From the early ARCH and GARCH frameworks to ARIMA and HAR-RV model, researchers have shown that historical information about market volatility can be used to construct useful forecasts. Meanwhile, the findings also demonstrate that forecasting accuracy cannot be judged only by the complexity of a model. This section evaluates the accuracy of linear approaches by using empirical evidence, and compare them with nonlinear models, and considers the future role of artificial intelligence in the stock volatility prediction.

3.1. How accurate are linear models

The recent empirical evidence provides a measurable support for continued competitiveness of linear models. Branco, Rubesam, and Zevallos had evaluated several linear and nonlinear machine-learning models for realized volatility across ten global stock market indices from January 2000 to December 2021 [8]. Their results end up that no evidence that nonlinear machine-learning models could systematically outperform linear models. Linear models were almost always among the best-performing models across the daily, weekly, and monthly forecasting horizons after testing ten indices. The study also found that the additional predictors improved daily forecasts for all ten indices and weekly forecasts for nine of the ten indices, while the HAR-type models remained competitive at the monthly horizon [8].

Kilic provided further evidence by using rolling forecasts for S&P 500 realized volatility from 2006 to 2023 [7]. The study compared linear HAR and ARFIMA models with nonlinear regime-switching models, and several machine-learning methods, including Extreme Gradient Boosting, deep neural networks, LSTM models, and GRU models. For the full test period, the five models with the lowest prediction errors under both major loss measures were STHAR, HAR, THAR, MSHAR, and a deep neural network, with the nonlinear STHAR model achieving the strongest overall performance [7]. The placement of HAR ahead of several machine-learning models showed that a simple and interpretable linear model can still remain highly competitive even in a modern comparison involving advanced algorithms.

The most recent evidence have also showed that the relative performance of linear and AI models depends on the market conditions. Rahimikia and Poon trained and tested nearly seven million machine-learning model specifications for realized volatility forecasting, and they found that high-dimensional machine-learning models outperformed HAR models during approximately 90 percent of the out-of-sample period [9]. However, HAR-family models generally performed better during

DOI: 10.54254/2754-1169/2026.LD39005

periods that has extremely high volatility. These results support a more precise conclusion: linear models are not the best, but they can provide a stable forecast in important market regimes, especially when extreme volatility reduces the reliability of more flexible algorithms.

3.2. A linear, nonlinear, and machine-learning approaches

The nonlinear and machine-learning approaches are designed to capture interactions and changing patterns that may not be represented by a fixed linear relationship. McAleer and Medeiros compared linear and nonlinear models for forecasting realized volatility, and they provided an early framework for empirical comparison rather than assuming that complexity alone determines forecasting quality [6]. More recent studies have expanded this comparison by using neural networks, tree-based algorithms, regularization methods, and recurrent architectures.

The results of these comparisons remain mixed. Díaz, Hansen, and Cabrera found that machine-learning algorithms could outperform the HAR benchmark when forecasting S&P 500 realized volatility at a short one-month horizon, but this advantage did not continue at the longer six to twelve months horizons [10]. Their study also found that improved statistical prediction did not automatically produce a substantial economic gains. This distinction is very important because a model can reduce the statistical error measure without necessarily improving investment or portfolio outcomes [10].

Gunnarsson reviewed research on AI and machine learning for realized and implied volatility forecasting, and concluded that AI methods are highly promising and they often achieve results comparable to or better than the traditional econometric models [11]. During the same time, the review showed that traditional models remain highly relevant and that hybrid approaches combining econometric models with machine learning represents an important research direction [11]. Overall, the evidence supports a comparative approach in which the linear models are treated as strong benchmarks rather than outdated methods. The most suitable model depends on different conditions, including the forecasting horizon, the available predictors, the market regime, and the purpose of the forecast.

3.3. Future directions: AI and stock volatility prediction

Based on the studies reviewed, the future of stock volatility prediction is likely to involve a closer integration of artificial intelligence with traditional econometric models. Modern AI algorithms can process large number of predictors, and they can incorporate information that is difficult to include in a simple linear equation. The potential inputs include high-frequency trading data, limit-order-book information, financial news, and measures of market sentiment. These broader information environment gives AI the potential to identify patterns that the traditional models may not capture [9, 11].

Deep-learning methods such as LSTM and GRU are particularly relevant because they are designed to learn temporal relationships in sequential data. Other attention-based architecture may further improve the ability of forecasting systems to process long sequences and combine multiple types of financial information. However, higher flexibility also increases the risk of overfitting, unstable performance, and limited interpretability. The mixed empirical evidence in recent studies suggests that future progress should not focus only on replacing linear models with larger algorithms [8-10].

In addition, future stock volatility prediction may become more regime-aware. Instead of relying on one model for every market condition, an AI system could identify whether the market is in a

DOI: 10.54254/2754-1169/2026.LD39005

relatively normal or highly volatile regime and then place different weights on linear, nonlinear, and machine-learning forecasts. Recent evidence indicates that machine-learning models may perform strongly during normal periods while HAR-family models can remain more robust during extreme volatility [9]. A regime-aware ensemble could combine the strengths of both approaches rather than requiring one model to dominate every forecasting situation.

4. Conclusion

To summarize, this review examined the forecasting accuracy of linear models in stock price volatility prediction by tracing the development of several important approaches, and compare their performance with nonlinear and artificial-intelligence-based methods. Section 2 reviewed ARCH, GARCH, ARIMA, and HAR-RV as major approaches for modeling changing and persistent volatility. These models demonstrate that historical information can provide important predictive value, without requiring highly complicated algorithms. In particular, the HAR-RV framework shows how information from daily, weekly, and monthly horizons can be combined within a relatively simple and interpretable structure.

The evidence reviewed in Section 3 shows that linear models remain competitive, rather than automatically being replaced by other advanced methods. Recent studies across multiple stock indices and forecasting horizons found no consistent evidence that non linear machine-learning can improve forecasts under particular conditions, while the HAR-family models can remain more robust during periods of extreme volatility [9]. These findings demonstrate that forecasting accuracy depends on market conditions, forecasting horizons, predictor availability, and evaluation criteria.

The main contribution of this review is to show that model complexity should not be treated as the only indicator of forecasting quality. A useful volatility forecasting model must balance accuracy with interpretability, computational efficiency, robustness, and practical value. This conclusion is very meaningful for future financial research because artificial intelligence is expanding the range of available forecasting tools. Rather than viewing AI as a complete replacement for the traditional econometric models, future work can combine the strengths of both approaches, improving the stock volatility prediction while maintaining the transparency needed for financial decision-making.

References

[1] Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature, 41*, 478-539.
[2] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*, 987-1007.
[3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307-327.
[4] Rubio, L., et al. (2023). Forecasting volatility by using wavelet transform, ARIMA and GARCH models. *Eurasian Economic Review, 13*, 803-830.
[5] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*, 174-196.
[6] McAleer, M., & Medeiros, M. C. (2009). Forecasting realized volatility with linear and nonlinear models. CARF Working Paper, No. CARF-F-189, Center for Advanced Research in Finance, University of Tokyo.
[7] Kılıç, R. (2025). Linear and nonlinear econometric models against machine learning models: Realized volatility prediction. *Finance and Economics Discussion Series, 2025-061*, Board of Governors of the Federal Reserve System.
[8] Branco, R. R., Rubesam, A., & Zevallos, M. (2024). Forecasting realized volatility: Does anything beat linear models? *Journal of Empirical Finance, 78*, 101524.

DOI: 10.54254/2754-1169/2026.LD39005

[9] Rahimikia, E., & Poon, S. H. (2026). Machine learning for realised volatility forecasting. *Journal of Empirical Finance, 88*, 101739.
[10] Díaz, J. D., Hansen, E., & Cabrera, G. (2024). Machine-learning stock market volatility: Predictability, drivers, and economic value. *International Review of Financial Analysis, 94*, 103286.
[11] Gunnarsson, E. S., et al. (2024). Prediction of realized volatility and implied volatility indices using AI and machine learning: A review. *International Review of Financial Analysis, 93*, 103221.