Vol. 10 (2025): Proceedings of the 2nd International Conference on Business, Management and Sustainability (ICBMS 2025)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icbms.2502.329334

Research on Stock Realized Volatility Prediction Using Multi-Models Incorporating International Macro Factors

Jiarui He*

School of International Business, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China

**Corresponding author: Jiarui He*

Abstract

Predicting realized volatility (RV) is crucial for asset allocation and risk management. Existing multi-factor research on predicting realized volatility in stocks has rarely incorporated international macroeconomic factors, and the comparison of machine learning models remains limited. This paper constructs a multi-factor framework combining technical and international macroeconomic factors using monthly data from the 50 most liquid stocks in the CSI 300 Index from 2018 to 2024. Three models-ridge regression, extreme gradient boosting (XGBoost), and a multilayer perceptron (MLP)-are used to predict the next period's RV. Results show that incorporating international macroeconomic factors improves the forecast accuracy of all models, with machine learning performing even better on the test set ($R^2 \approx 0.43$). The study finds that incorporating international macroeconomic factors significantly improves model forecast accuracy, with machine learning outperforming linear regression models. This research provides a reference for selecting methods for volatility forecasting in different market scenarios and offers theoretical insights for investors, asset pricing, and risk management.

Keywords

realized volatility, multi-factor model, linear regression, machine learning, volatility forecasting

1. Introduction

Realized Volatility (RV) has important applications in asset pricing, risk management, derivatives pricing, and portfolio optimization. RV is constructed from the sum of squares of high-frequency returns. It is a non-parametric and testable volatility measure used to represent the intensity of volatility after the event. It is a mature estimation and prediction framework (Andersen et al., 2003). In the study of RV prediction, heterogeneous autoregressive (HAR) and generalized autoregressive conditional heteroskedasticity (GARCH) are the most classic time series frameworks, among which HAR-RV is widely used as a common baseline for low-frequency prediction (Corsi, 2009). In recent years, with the development of machine learning algorithms, a large number of scholars have tried to use RV to train machine learning models to predict volatility, and their prediction results are basically better than the linear regression HAR model (Gunnarsson et al., 2024). For example, Bouri et al. used the random forest algorithm in 2021 to construct a Bitcoin realized volatility prediction model. The results showed that the random forest algorithm has better prediction results than the linear model, HAR, and other traditional models in the Bitcoin trading market (Bouri et al., 2021).

At the same time, statistics from the Ministry of Commerce and the State Administration of Foreign Exchange show that China's total outward direct investment in all industries in 2024 will be RMB 1,159.27 billion, an increase of 11.3% year-on-year (Department of Cooperation of the Ministry of Commerce and State Administration of Foreign Exchange, 2025). With the integration of the global economy, overseas investment has gradually become an effective means for domestic enterprises to obtain innovative resources and advanced technologies. In this context, the fluctuations of international macro factors (such as exchange rates) have an increasingly greater impact on the Chinese stock market. Girardin and Joyeux found that macro fundamentals can explain long-term stock fluctuations based on research on the Chinese market (Girardin and Joyeux, 2013). Liu et al. found that global stock market information can significantly improve the out-of-sample prediction of China's stock market fluctuations (Liu et al., 2019). However, in the current research on multi-factor stock prediction, the introduction and verification of international macro factors are not extensive, and the model still has certain limitations on the accuracy of earnings prediction for multinational companies.

Drawing on the application of realized volatility and existing research on the impact of international macroeconomic factors on A-share volatility, this paper conducts an empirical analysis, selecting monthly data from the 50 most liquid stocks in the CSI 300 Index from 2018 to 2024 as a sample, with the next period's realized volatility as the forecast target. Two control groups, one with and one without international macroeconomic factors, are used to construct three models: ridge regression, XGBoost, and MLP. Model performance is then compared using rolling forward validation and metrics such as R^2, MAE, and RMSE. This study explores the incremental predictive value of international macroeconomic factors, providing a new perspective for individual stock volatility research. It also validates the advantages of machine learning over linear benchmarks, provides a reference framework for volatility forecasting, and provides effective input for investment strategies such as target volatility and risk parity.

1.1 Data and Methods

1.1.1 Data processing and feature engineering

The data for this study were extracted from the Tushare financial database API (Tushare, 2025). The study used the 50 most liquid constituent stocks in the CSI 300 Index from 2018 to 2024 as a sample, extracted monthly frequency data, and the target variable was the realized volatility of the next month. Due to the obvious skewness of the original distribution, this paper first took the logarithm of the RV and fitted each model; after the prediction, the resulting index was restored to the RV, and the indicators were calculated in the original RV space to ensure the comparability of different models and feature settings.

The basic feature set includes volatility features, price trading features, and fundamental features. Volatility features include the previous month's short-term realized volatility, the trailing twelve-month long-term realized volatility, and the ratio of the two. For price trading features, this study uses the previous month's return as a momentum indicator, log market capitalization to characterize scale, and log turnover rate to characterize liquidity. For fundamental features, return on equity (ROE) is used to characterize profitability, and asset growth rate is used to characterize corporate growth.

Based on existing research demonstrating the predictive power of macroeconomic fundamentals and cross-market information for Chinese stock market volatility, this paper selects the monthly fluctuations of the RMB against the US dollar, the monthly fluctuations of the US 10-year Treasury yield, and the monthly fluctuations of the federal funds rate as features in constructing international macroeconomic factors. These represent the state of foreign exchange, interest rates, and global financial conditions, respectively. All features are lagged one month before modeling to prevent future information leakage.

Data preprocessing is an indispensable and key link in the modeling process. With the help of this link, data quality can be effectively improved. The quality of data is directly related to the reliability of the final output results of the model.

This paper first conducted data testing and cleaning, including addressing missing values and outliers. Secondly, continuous variables were truncated at the 1% and 99% percentiles to mitigate the influence of extreme values on the results. Furthermore, cross-sectional factors were z-score standardized by month, while macroeconomic factors were time series standardized. To ensure the robustness of the linear model estimates, this paper tested for correlation and multicollinearity between features. Since the results were all within the normal range, no features were removed.

1.1.2 Modeling and Evaluation

The study compared the prediction results of three models: Ridge Regression, XGBoost, and a shallow MLP, based on a unified training partition, on two feature sets. Model selection followed a logic of increasing model capacity gradients, comprehensively comparing the suitability of models of varying complexity for the study scenario. Ridge Regression, a representative linear baseline, offers high interpretability; XGBoost, a representative ensemble tree model in machine learning, can capture high-order interactions within an additive ensemble framework and excels at capturing nonlinear changes; and MLP, a shallow neural network, can capture smooth and complex structures under moderate regularization. This study examined the impact of different inductive preferences on predictions.

To ensure standardization, this study introduces HAR-RV as a statistical baseline for comparison. The core idea of HAR-RV is to regress future volatility to historical volatility at different time scales to capture long-memory characteristics. Since the sample of this study is monthly frequency data, the study approximates it as a linear regression with the logRV of the current month and the average logRV of the past 12 months as the target variable, and uses OLS estimation in logarithmic space. When constructing the HAR-RV macro expansion model, like other models, the logarithm of the short-term/long-term ratio and three international macro factors, namely the change in the RMB against the US dollar, the change in the US 10-year Treasury bond yield, and the change in the federal funds rate (Corsi, 2009), are added. The models are compared under a unified dimension to ensure the fairness and persuasiveness of the experimental results.

To prevent future data leaks, this study employed Time Series Split cross-validation during modeling (scikit-learn developers, n.d.). This method systematically searched for hyperparameter combinations and selected the optimal hyperparameters based on the out-of-sample performance of the validation folds. The selected model was then evaluated on the test set. The test set evaluation metrics used were R^2, MAE, and RMSE. Results were presented side-by-side for both feature sets to compare the performance of the nonlinear model with the linear baseline and the incremental contribution of international macro factors.

1.2 Empirical Results and Discussion

The research results, shown in Table 1, show that incorporating international macroeconomic factors improves the forecasting accuracy of all models. This demonstrates the robust incremental value of information on foreign exchange, interest rates, and global financial conditions. The MLP model with international macroeconomic factors achieved the best results (R^2=0.412, RMSE=0.03857), followed by Xgboost. Ridge regression and HAR-RV, while showing more modest improvements, still align with the prevailing trend. In terms of incremental gains, the gains were more pronounced for nonlinear methods, consistent with economic intuition that models that can account for nonlinearities and interactions are more advantageous. To identify the sources of incremental gains, this paper evaluated variable contributions using permutation importance by monthly blocks and groups during the test period. The results show that permutation of the international macroeconomic group reduced R^2 by 0.138 and 0.171, respectively, in the machine learning models, with fluctuations in the RMB against the US dollar leading the way.

Table 1: Summary of prediction results of each model test set

Model	Feature Set	R^2	MAE	RMSE
HAR-RV	Base	0.270366	0.029249	0.042982
HAR-RV	all	0.282707	0.02904	0.042617
Ridge	Base	0.272467	0.031518	0.042920
Ridge	all	0.305556	0.031045	0.041932
XGBoost	Base	0.287808	0.030659	0.042465
XGBoost	all	0.376902	0.028616	0.039720
MLP	Base	0.256641	0.031814	0.043384
MLP	all	0.412388	0.029019	0.038572

From a model perspective, the MLP model with international macroeconomic factors in this study significantly outperformed other models in terms of out-of-sample R^2. This demonstrates its adaptability in the context of predicting monthly realized stock volatility after introducing international macroeconomic variables.

This is consistent with the findings of numerous studies using deep learning and machine learning for volatility and asset pricing forecasting (Gunnarsson et al., 2024, Bouri et al., 2021, Zhu et al., 2023). However, MLP also has certain limitations. For instance, it is slightly inadequate in terms of interpretability and training stability, which may be related to its high sensitivity to hyperparameters and initialization, as well as its greater tendency to overfit in small-sample scenarios.The XGBoost model with macroeconomic factors performed second best, but performed better in terms of mean absolute error, reflecting the tree model's ability to better capture piecewise nonlinearities and threshold effects. Ridge performed well in terms of stability and interpretability, indicating that Ridge can robustly absorb macroeconomic increments, but is limited by its linear model, resulting in limited marginal improvement and inadequate characterization of nonlinear interactions. As a statistical baseline, HAR-RV, without macroeconomic settings, primarily relies on historical volatility at different time scales to capture long-memory characteristics, resulting in interpretable and stable out-of-sample performance. Incorporating macroeconomic factors only slightly improves, reflecting the limited ability of a purely linear framework to absorb external financial conditions.

From the perspective of international macroeconomic factors and economic mechanisms, fluctuations in the RMB against the US dollar reflect changes in the US dollar cycle and cross-border liquidity. A weakening RMB typically reflects a tightening financing environment and rising risk aversion, which in turn increases price volatility. The US 10-year Treasury yield, on the other hand, represents fluctuations in the risk-free rate and term premium, while the federal funds rate captures short-term policy interest rates and liquidity conditions. For example, during rate hikes, rising financing costs and declining market-making capacity can lead to increased short-term volatility. All three characteristics provide contextual macroeconomic information, supplementing external shocks that are often lacking in individual stock-level variables, and therefore contribute significantly to forward-looking forecasts (Girardin and Joyeux, 2013, Liu et al., 2019).

To ensure the robustness and persuasiveness of our results, we repeated the experiment with various parameter settings. The overall pattern remained consistent: regardless of parameter adjustments, the model with international macroeconomic factors consistently outperformed the model without them. This was confirmed across all three models, reinforcing the view that international macroeconomic factors do provide additional explanatory power.

It is important to note that the overall R^2 level of this study's results is low. This is primarily due to the financial market context, data characteristics, and limitations of the research setting, and does not necessarily indicate a lack of model predictive power. Realized volatility is significantly impacted by shocks such as unexpected events and stock-specific factors, resulting in a high proportion of noise. Furthermore, the monthly data used in this study have a limited span, resulting in a small number of observations and limited statistical power. Therefore, this article focuses on relative comparisons between different models and feature settings, rather than absolute numerical differences.

1.3 Conclusion

This study investigates the prediction of realized stock volatility by comparing factor specifications and models. Empirical results demonstrate that, under a unified training and testing framework, incorporating monthly fluctuations in the RMB/USD exchange rate, the US 10-year Treasury yield, and the federal funds rate improves overall out-of-sample performance. Nonlinear methods generally outperform linear baselines, with machine learning models showing a clear advantage, with shallow neural networks performing the best. Model rankings remain consistent across different specifications.

Further analysis found that the increase mainly came from macroeconomic variables, especially exchange rates and US interest rates. These external financial conditions provide common time-varying information across individual stocks, which can significantly improve the forecast effect while retaining the core volatility factors.

This study has certain limitations: the sample period and the selection of macroeconomic features are limited, and the realized volatility series is affected by factors such as jumps, structural change points, and microstructural noise, which limit its predictability. Consequently, the performance of common fitting indices is low. Future research could test robustness with longer sample periods and higher frequencies, expand the feature set, explore methods such as combined forecasting and time-varying parameters, introduce more cutting-edge machine learning methods, and further refine the international factor mechanism to enhance forecast accuracy and economic explanatory power. Given these shortcomings, future research could extend the sample period and test robustness with higher-frequency data. Furthermore, it could apply more machine learning models, incorporate more macroeconomic factors into the multivariate framework, and strengthen uncertainty assessment and interpretability.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P., (2003). Modeling and forecasting realized volatility. *Econometrica,* vol. 71**,** no. 2, pp. 579-625.

Bouri, E., Gkillas, K., Gupta, R. and Pierdzioch, C., (2021). Forecasting realized volatility of bitcoin: The role of the trade war. *Computational Economics,* vol. 57**,** no. 1, pp. 29-53.

Corsi, F., (2009). A simple approximate long-memory model of realized volatility. *Journal of financial econometrics,* vol. 7**,** no. 2, pp. 174-196.

Department of Cooperation of the Ministry of Commerce and State Administration of Foreign Exchange, (2025). *Brief Statistics of China's Outward Direct Investment in All Industries in 2024* [Online]. Available: http://www.mofcom.gov.cn/tjsj/gwjjhztj/art/2025/art_044cb9f995184b4e88970994c22a7881.html [Accessed December 12, 2025].

Girardin, E. and Joyeux, R., (2013). Macro fundamentals as a source of stock market volatility in China: A GARCH-MIDAS approach. *Economic Modelling,* vol. 34, pp. 59-68.

Gunnarsson, E. S., Isern, H. R., Kaloudis, A., Risstad, M., Vigdel, B. and Westgaard, S., (2024). Prediction of realized volatility and implied volatility indices using AI and machine learning: A review. *International review of financial analysis,* vol. 93, p. 103221.

Liu, J., Ma, F. and Zhang, Y., (2019). Forecasting the Chinese stock volatility across global stock markets. *Physica A: Statistical Mechanics and Its Applications,* vol. 525, pp. 466-477.

Scikit-learn developers, (n.d.). *TimeSeriesSplit - scikit-learn 1.8.0 documentation* [Online]. In scikit-learn. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html [Accessed December 12, 2025].

Tushare, (2025). *Tushare Pro platform and interface documentation* [Online]. Available: https://tushare.pro/document/2 [Accessed December 16, 2025].

Zhu, H., Bai, L., He, L. and Liu, Z., (2023). Forecasting realized volatility with machine learning: Panel data perspective. *Journal of Empirical Finance,* vol. 73, pp. 251-271.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.