Received: 22 September 2021 | Revised: 21 July 2022 | Accepted: 22 July 2022

DOI: 10.1002/for.2899

RESEARCH ARTICLE

WILEY

Volatility forecasting for stock market incorporating macroeconomic variables based on GARCH-MIDAS and deep learning models

Yuping Song | **Xiaolong Tang** | **Hemin Wang** | **Zhiren Ma**

School of Finance and Business, Shanghai Normal University, Shanghai, China

Correspondence
Yuping Song, School of Finance and Business, Shanghai Normal University, Shanghai, Xuhui District, Guilin Road 100, 6 A, 402, China.
Email: songyuping@shnu.edu.cn

Funding information
Academic Innovation Team of Shanghai Normal University, Grant/Award Number: 310-AC7031-19-004228; Key Subject of Quantitative Economics of Shanghai Normal University, Grant/Award Number: 310-AC7031-19-004221; Ministry of Education, Humanities and Social Sciences project, Grant/Award Number: 18YJCZH153; Youth Academic Backbone Cultivation Project of Shanghai Normal University, Grant/Award Number: 310-AC7031-19-003021; General Research Fund of Shanghai Normal University, Grant/Award Number: SK201720; National Natural Science Foundation of China, Grant/Award Numbers: 11901397, 71973098

Abstract

Empirical experiments have shown that macroeconomic variables can affect the volatility of stock market. However, the frequencies of macroeconomic variables are low and different from the stock market volatility, and few literature considers the low-frequency macroeconomic variables as input indicators for deep learning models. In this paper, we forecast the stock market volatility incorporating low-frequency macroeconomic variables based on a hybrid model integrating the deep learning method with generalized autoregressive conditional heteroskedasticity and mixed data sampling (GARCH-MIDAS) model to process the mixing frequency data. This paper firstly takes macroeconomic variables as exogenous variables then uses the GARCH-MIDAS model to deal with the problem of different frequencies between the macroeconomic variables and stock market volatility and to forecast the short-term volatility and finally takes the predicted short-term volatility as the input indicator into machine learning and deep learning models to forecast the realized volatility of stock market. It is found that adding macroeconomic variables can significantly improve the forecasting ability in the comparison of the forecasting effects of the same model before and after adding the macroeconomic variables. Additionally, in the comparison of the forecasting effects among different models, it is also found that the forecasting effect of the deep learning model is the best, the machine learning model is worse, and the traditional econometric model is the worst.

KEYWORDS
deep learning model, GARCH-MIDAS model, macroeconomic variables, realized volatility forecasting

JEL CLASSIFICATION
G17, C45

1 | INTRODUCTION

The volatility of financial assets plays an extremely important role in investment decision (Alaali, 2020), option pricing (Recchioni et al., 2020), and stock market movement (Balli et al., 2015). Therefore, using appropriate models to forecast volatility possesses an important theoretical significance and practical value. Bollerslev

(1986) proposed generalized autoregressive conditional heteroskedasticity (GARCH) family models based on low-frequency data and used it to describe the volatility agglomeration phenomenon and forecast the volatility. However, the model did not take into account the intraday transaction information of financial assets. Andersen and Bollerslev (1998) calculated the realized volatility based on intraday high-frequency trading information data and used it to measure intraday volatility. Corsi (2009) proposed the heterogeneous autoregressive (HAR) model to forecast the volatility based on the heterogeneous market hypothesis and found that the HAR model had a better forecasting ability. However, the main theoretical basis of the above-mentioned traditional econometric models is the linear regression or quadratic function, which cannot describe the nonlinear characteristics between variables. In addition, other factors except for historical price information are not considered.

Deep learning models can describe the complex and nonlinear relationships of data based on multidimensional influencing factors. Some scholars have used deep learning models to forecast stock market volatility. Xiao et al. (2015) used the convolutional neural network (CNN) model to forecast the volatility of the Standard & Poor 500 (S&P 500) index, and Xiong et al. (2016) used the long short-term memory (LSTM) model to forecast the static volatility of the S&P 500 index. Chen (2018) forecasted the volatility of the Shanghai Composite Index based on deep learning models. However, the above-mentioned traditional econometric models and deep learning models did not take into account the impact of macroeconomic variables. Relevant literature such as Reitz (1988), Campbell and Cochrane (1999), and Barro (2006) had shown that macroeconomic variables would pass through a series of transmission mechanisms, such as marginal utility, cost of capital, discount rate, and other factors, to eventually lead to shocks in stock market. The different frequencies of macroeconomic variables and stock market volatility would lead to some difficulties when using traditional econometric models to deal with the problems between them. Silvestrini and Veredas (2008) pointed out that high-frequency data could be converted into low-frequency data, and Chow and Lin (1976) pointed out that low-frequency data could be converted into high-frequency data, but these methods would lose some information contained in the original data, which might lead to biases in the model design.

The main contributions of this paper are as follows. Firstly, we take macroeconomic variables as exogenous variables and introduce them into the GARCH-MIDAS model. On the one hand, it solves the problem of different frequencies between macroeconomic variables and stock market volatility. On the other hand, one can use it to forecast the short-term volatility and take the predicted short-term volatility as the input indicator into machine learning and deep learning models to forecast the realized volatility. The hybrid model integrating the deep learning method with GARCH-MIDAS model in this paper solves the shortcomings that traditional econometric models cannot handle highly nonlinear problems and machine learning models cannot handle mixing problems, which provides new ideas for machine learning and deep learning models to forecast the realized volatility. Secondly, we use a variety of models, including traditional econometric models such as GARCH, machine learning models, and especially deep learning models to forecast the stock market volatility, and use a variety of criteria to evaluate the models' forecasting abilities. And it can be found that the forecasting accuracy of the deep learning models is the highest. Additionally, comparing the models before and after introducing the macroeconomic variables, it can be found that macroeconomic variables can improve the volatility forecasting effects of the models.

The remaining sections of this paper are as follows. Section 2 describes the principle of GARCH-MIDAS model and deep learning gated recurrent unit (GRU) model and comprehensively summarizes the evaluation criteria. Section 3 shows the empirical results on the out-of-sample volatility forecasting accuracy for seven models under five loss functions before and after adding the low-frequency macroeconomic variables. Section 4 concludes.

2 | COMPONENT MODELS AND EVALUATION CRITERIA

2.1 | GARCH-MIDAS model incorporating macroeconomic variables

Engle et al. (2008) had found that the short-term volatility component in the stock market would return to its mean value at a reasonable speed, which was very sensitive to sudden information in the market, while the long-term component was mainly affected by macroeconomic variables. If X_t represents macroeconomic variables, the stock return series can be expressed as follows:

$$r_{i,t} = \mu + \sqrt{\tau(X_t) g_{i,t}} \varepsilon_{i,t}, \quad (1)$$

where $r_{i,t}$ represents the return rate in the trading day i of the month (quarter or year) t for the stock market, μ represents the expected return, τ_t represents the long-term volatility component, and $g_{i,t}$ represents the short-term volatility component.

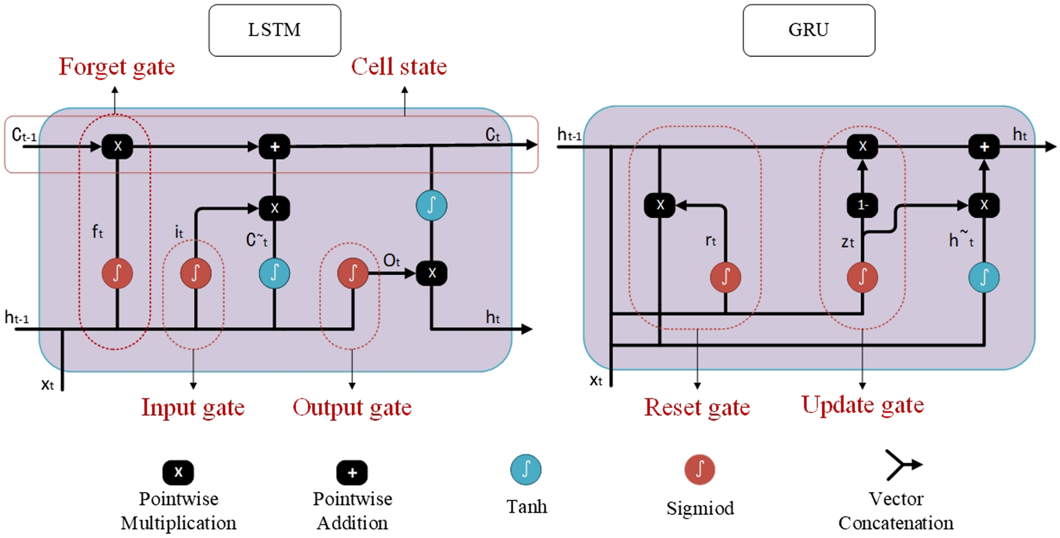

FIGURE 1 Structure diagram of long short-term memory (LSTM) and gated recurrent unit (GRU) models

Assuming that the short-term volatility component is a daily GARCH(1,1) process, then

$$g_{i,t} = (1-\alpha-\beta) + \frac{\alpha(r_{i-1,t}-\mu)^2}{\tau_t} + \beta g_{i-1,t}. \tag{2}$$

The long-term volatility component is expressed as the mixed data sampling (MIDAS) equation of a macroeconomic variable and realized volatility, which is

$$\begin{aligned}\tau_t = m + \theta_R \sum_{j=1}^{K} \varphi_j^R(\omega_{R1}, \omega_{R2}) RV_{t-j} \\ + \theta_X \sum_{j=1}^{K} \varphi_j^X(\omega_{X1}, \omega_{X2}) X_{t-j},\end{aligned} \tag{3}$$

where RV_{t-j} represents the realized volatility, $X_{t\text{-}j}$ represents the macroeconomic variables, θ_R and θ_X represents the influence degree of the realized volatility and macroeconomic variables on the long-term volatility component of the stock return series, respectively, K represents the number of lag periods, and $\varphi_j(\omega_1, \omega_2)$ represents the weight function.

Therefore, the low-frequency data of macroeconomic variables and the mid- and high-frequency data of the stock return series are organically combined through MIDAS filtering, which solves the problem of different frequencies between them.

2.2 | GRU neural network

LSTM neural network is a special recurrent neural network structure, including three gate structures: forget gate f_t, input gate i_t, and output gate o_t. By modifying the LSTM model, Cho et al. (2014) proposed the GRU model, in which there are only two gates: update gate z_t and reset gate r_t, as shown in Figure 1.

The function of the update gate z_t is to control the extent to which the state information from the previous moment is brought into the current hidden layer state h_t. The larger the value of the update gate is, the more state information from the previous moment is brought in. The function of the reset gate r_t is to control how much information of the previous state is written to the candidate set of the current hidden layer $\widetilde{h}_t$. The larger the value of the reset gate, the less information of the previous state is written.

The calculation formulas of the GRU model are as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \tag{4}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \tag{5}$$

$$\widetilde{h}_t = tanh\left(W_{\widetilde{h}} \cdot [r_t * h_{t-1}, x_t]\right), \tag{6}$$

$$h_t = (1-z_t) * h_{t-1} + z_t * \widetilde{h}_t, \tag{7}$$

where σ represents the activation function, W represents the weight matrix, [] represents the connection of two vectors, and $*$ represents the matrix multiplication.

2.3 | Evaluation criteria

In order to evaluate the forecasting effects between different models, this paper will use the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), and root mean square and percentage error (RMSPE) as evaluation criteria. The formulas of each criterion are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \tag{10}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \times 100\%, \tag{11}$$

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{y}_i - y_i}{y_i}\right)^2} \times 100\%, \tag{12}$$

where $\hat{y}_i$ represents the forecasting value, y_i represents the real value, and n represents the number of samples. The smaller each evaluation criterion is, the better the prediction effect of the model is.

3 | EMPIRICAL ANALYSIS

3.1 | Data source and variables

When constructing the macro factors, this paper selects the gross domestic product (GDP), consumer price index (CPI), purchasing managers' index (PMI), and the macroeconomic consensus index in the sample interval for principal component analysis where GDP is naturally logarithmized as *ln_GDP* and four principal component factors are obtained. Through the eigenvalues and the explanatory power contribution degree of principal component factors, this paper selects the first principal component to extract the macro factor, performs first-order difference on it, and finally obtains the macro factor data sequence defined as MACRO.

In the calculation of short-term volatility with macro factors, this paper adds the constructed macro factor variable MACRO into the GARCH-MIDAS model, generates the long-term volatility and short-term volatility based on historical data, and takes the short-term volatility series containing macro factors as one of the input variables into the machine learning and deep learning models.

TABLE 1 Description of variable

Name of indicators	Explanation
Volume	Daily transaction volume
Transactions	The volume of deals between buyers and sellers
Bias	(closing price of the day – 5-day average price)/5-day average price
DMA	Five-day moving average – 10-day moving average
CDP	(The highest price of the previous day + the lowest price of the previous day + 2 * closing price of the previous day)/4
AR	(Closing price – opening price)/(opening price – the lowest price) * 100%
BR	(The highest price – closing price)/(closing price – the lowest price) * 100%
Pct change	Range of rise and down
Overnight spread	Opening price – closing price of the previous day
GARCH-MIDAS	The short-run volatility with macroeconomic variables
RV	Realized volatility, $RV = \sum_{d=1}^{48} R_{t,d}^2$

Note: $R_{t,d} = 100\ (\ln P_{t,d} - \ln P_{t,d-1})$, where $\ln P_{t,d}$ is the 5-min high-frequency closing price of the stock index during the trading day of t.
Abbreviations: GARCH-MIDAS, generalized autoregressive conditional heteroskedasticity and mixed data sampling; RV, realized volatility.

The data of GDP, CPI, and PMI are obtained from the National Bureau of Statistics of China (http://www.stats.gov.cn/), the data of Macroeconomic Concordance Index are obtained from Sina Finance (https://finance.sina.com.cn/), and the data related to Shanghai Stock Index as trading volume, turnover rate, and overnight spreads all come from Wind. The specific variables used in this paper are shown in Table 1.

3.2 | Correlation analysis

Figure 2 shows the analysis of Pearson correlation coefficient between different variables.

From Figure 2, it can be found that the realized volatility (RV) has the highest correlation with the short-term volatility (GARCH-MIDAS) that includes macroeconomic variables, indicating that the realized volatility will be affected by macroeconomic variables. In addition, there is a high correlation between the realized volatility (RV), daily transaction volume (Volume), and the transaction volume between buyers and sellers (Transactions),

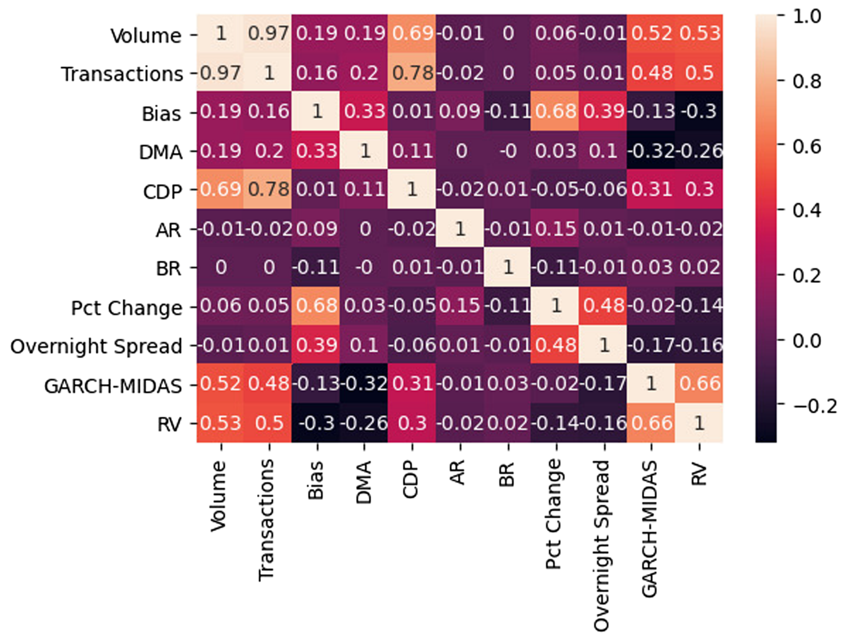

FIGURE 2 Correlation analysis of variables

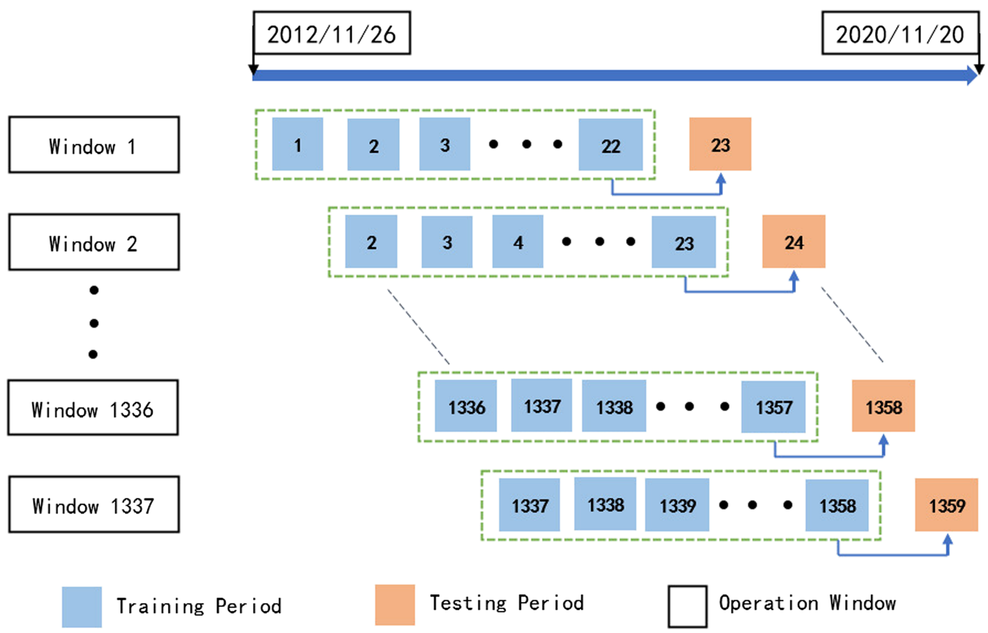

FIGURE 3 Schematic diagram of rolling time window method

which indicates that the realized volatility is affected by many factors not only the asset return series or historical volatility.

3.3 | Forecasting results

Due to the large sample size of the input data and the high randomness of the neural network model, in order to make the model capture the data features better and to better compare the forecasting results of different models, the data are split into the training set and the test set according to 7:3. And the hidden layer of both LSTM and GRU models is set to two layers, the number of memory units in the first layer is set to two times the number of input features, 22, and the *relu* is used as activation function. The optimizer is *RMSprop*, learning rate is 0.01, loss function is MAE, Batch Size is set to 32, and the epoch is 30. In addition, a dropout layer is added to prevent overfitting and set to 0.1.

This paper will use the rolling time window method to forecast the realized volatility of Shanghai Composite Index, taking 22 days of historical data to forecast 1-day volatility in the future as example shown in Figure 3.

Since the random seeds within the neural network model have random results on the model predictions, so to make the forecasting results more credible, we have performed LSTM and GRU models 100 times, respectively, and averaged them as the final forecasting results for comparison with other models. The results for both the LSTM and GRU models in Tables 2 and 5 are the average of 100 times.

This paper will use the variables with and without macro factors and multiple models to make one-step forecasting. The effects of macro factors and the forecasting capabilities of each model are analyzed through different evaluation criteria. The results are shown in Table 2.

Table 2 demonstrates the corresponding forecasting evaluation criteria values of the traditional econometric model, machine learning model, and deep learning model. First of all, from the perspective of the forecasting accuracy of each model, the deep learning model is higher than that of the machine learning model, and the performance of the machine learning model is better than the traditional econometric model. On the analysis of the forecasting results adding the GARCH-MIDAS indicator, it is found that the GRU model with the best performance in the deep learning model has a reduction of 27.21% in the MSE compared with the support vector machine (SVM) model with the best performance in the machine learning model, and a reduction of 71.65% compared with the HAR-RV model, the MAE index reduced by 12.98% and 51.24%, and the RMSPE index reduced by 16.14% and 38.89%, respectively. From the perspective of adding the GARCH-MIDAS indicator or not, the forecasting accuracy of the GRU model adding the GARCH-MIDAS indicator is higher than that of the GRU model that does not consider the factor. The forecasting

TABLE 2 Comparison of one-step forecasting results

		GARCH	HAR-RV	SVM	Random forest	XGBoost	LSTM	GRU
MSE	+G-M	-	2.9264	1.0570	1.7532	1.5249	0.8671	**0.8296**
	−G-M	3.8908	3.6954	1.4289	1.3015	1.2991	1.3345	**1.0098**
RMSE	+G-M	-	1.7107	1.0184	1.3241	1.2349	0.9312	**0.9041**
	−G-M	1.9725	1.9223	1.1731	1.1408	1.1398	1.1552	**1.0005**
MAE	+G-M	-	1.2089	0.6773	0.8249	0.6673	0.6111	**0.5894**
	−G-M	0.8205	1.3842	0.7616	0.7823	0.6745	0.6594	**0.6406**
SMAPE	+G-M	-	0.2589	0.168	0.1724	0.1675	0.1699	**0.1553**
	−G-M	-	0.2699	0.1722	0.1822	0.1637	0.1967	**0.1587**
RMSPE	+G-M	-	2.6545	1.9696	1.9344	1.6765	1.6989	**1.6223**
	−G-M	-	3.114	2.0839	1.9194	**1.6896**	2.2457	1.7485

Note: "+GM" means adding GARCH-MIDAS indicator, "−GM" means not adding GARCH-MIDAS indicator; the numbers in bold in the table are the optimal values corresponding to each evaluation criterion. In order to verify the stability of the deep learning model, the evaluation criteria corresponding to the LSTM and GRU models are the average of 30 forecasting results.

Abbreviations: GARCH, generalized autoregressive conditional heteroskedasticity; GRU, gated recurrent unit; AE, mean absolute error; LSTM, long short-term memory; MSE, mean square error; RMSE, root mean square error; SMAPE, symmetric mean absolute percentage error; SVM, support vector machine.

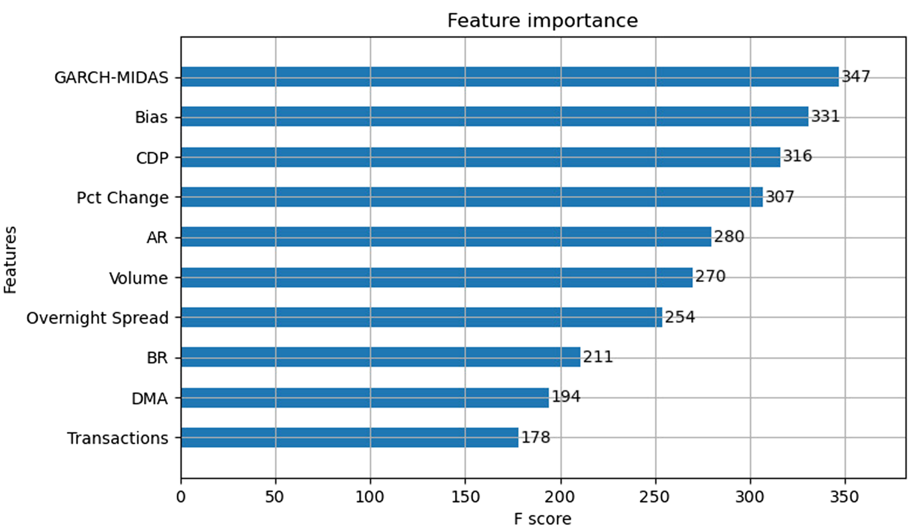

FIGURE 4 Ranking of feature importance

accuracy of the MSE, MAE, and RMSPE criteria are reduced by 17.85%, 8%, and 7.22%. In addition, the forecasting accuracy of other models that consider the GARCH-MIDAS indicator has also been improved to different degrees.

In short, the forecasting performance of deep learning models is the best, followed by machine learning models, and the forecasting performance of econometric models is the poorest. This is mainly because deep learning and machine learning models can better fit the nonlinearity between independent variables and dependent variables, while the essence of the econometric model is the linear regression, which cannot fully fit the highly nonlinear relationship between the influencing factor and the volatility. In addition, the GARCH-MIDAS indicator can significantly improve the accuracy of the model's forecasting results, indicating that this indicator is one of the important factors affecting the volatility of the Shanghai Stock Exchange Index. This paper uses the extreme gradient boosting (XGBoost) algorithm to rank the importance of independent variables, and the results are shown in Figure 4.

As shown in Figure 4, the GARCH-MIDAS indicator ranks first in the feature importance, indicating that this factor has a strong ability to explain the volatility of the Shanghai Composite Index and has an important contribution to the volatility forecasting of the Shanghai Composite Index.

In order to explain the differences between different models, this paper uses two-sample Kolmogorov–Smirnov (KS) test and Diebold–Mariano (DM) test to test the residual series of the models in pairs. The KS test assumes that the residual series of the two models have the same distribution and calculates and compares the significance level p and the size $\alpha = 0.05$. If the value p is less than the value α, the null hypothesis is rejected, which means that the residual series can be considered to have different distributions, indicating that there are differences between the models.

Table 3 shows the p-values corresponding to the KS test between the models considered in this paper. It can be seen from the table that the p-values between different models are all approximately equal to 0 and less than 0.05, rejecting the null hypothesis, indicating that the residual series between different models have different distributions, and there are large differences between the models.

In addition, in order to compare the forecasting ability between the two models, we will conduct DM test on each model. The test results are shown in Table 4.

TABLE 3 Comparison of KS test results of different models

	HAR-RV	SVM	Random forest	XGBoost	LSTM	GRU
HAR-RV	1.0000	-	-	-	-	-
SVM	0.0000	1.0000	-	-	-	-
Random forest	0.0000	0.0000	1.0000	-	-	-
XGBoost	0.0000	0.0000	0.0000	1.0000	-	-
LSTM	0.0000	0.0000	0.0000	0.0000	1.0000	-
GRU	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Note: The test data are the residual error of the one-step forecasting value of each model with the GARCH-MIDAS indicator added.
Abbreviations: GARCH, generalized autoregressive conditional heteroskedasticity; GRU, gated recurrent unit; KS, Kolmogorov–Smirnov; LSTM, long short-term memory; SVM, support vector machine.

TABLE 4 Comparison of DM test results of different models

	HAR-RV	SVM	Random forest	XGBoost	LSTM	GRU
HAR-RV	-	2.472***	1.639***	2.919***	2.556***	2.565***
SVM	−2.472***	-	1.334***	2.455**	1.041***	1.102***
Random forest	−1.639***	−1.334***	-	1.792**	1.767***	1.812***
XGBoost	−2.919***	−2.455**	−1.792**	-	2.631***	2.667**
LSTM	−2.556***	−1.041***	−1.767***	−2.631***	-	1.095***
GRU	−2.565***	−1.102***	−1.812***	−2.667**	−1.095***	-

Note: The test data are the residual error of the one-step forecasting value of each model with the GARCH-MIDAS indicator added.
Abbreviations: DM, Diebold–Mariano; GARCH, generalized autoregressive conditional heteroskedasticity; GRU, gated recurrent unit; LSTM, long short-term memory; SVM, support vector machine.
$^{***}p < 0.001$. $^{**}p < 0.01$. $^{*}p < 0.05$.

TABLE 5 Comparison of five-step forecasting results

		GARCH	HAR-RV	SVM	Random forest	XGBoost	LSTM	GRU
MSE	+G-M	-	1.9669	1.3087	1.7213	1.3043	1.1222	**1.0854**
	−G-M	4.8579	2.671	1.8923	1.4848	1.2531	**1.1981**	1.4114
RMSE	+G-M	-	1.4025	1.1377	1.3119	1.1421	1.2101	**1.0375**
	−G-M	2.2041	1.6343	1.368	1.2185	1.2194	1.2479	**1.1818**
MAE	+G-M	-	0.9503	0.7848	0.8853	0.7399	0.7901	**0.7117**
	−G-M	0.9328	1.1123	0.8866	0.8695	0.7409	0.7963	**0.7668**
SMAPE	+G-M	-	0.2255	0.1828	0.1847	0.1887	0.1937	**0.1757**
	−G-M	-	0.2511	0.1844	0.1987	0.1845	0.2101	**0.1779**
RMSPE	+G-M	-	2.7079	2.4532	2.1571	**1.8458**	2.0161	2.2545
	−G-M	-	3.2439	2.7438	2.3237	**2.0487**	2.1236	2.5348

Note: "+GM" means adding GARCH-MIDAS indicator, "−GM" means not adding GARCH-MIDAS indicator; the numbers in bold in the table are the optimal values corresponding to each evaluation criterion. In order to verify the stability of the deep learning model, the evaluation criteria corresponding to the LSTM and GRU models are the average of 30 forecasting results.

Abbreviations: GARCH, generalized autoregressive conditional heteroskedasticity; GRU, gated recurrent unit; LSTM, long short-term memory; MAE, mean absolute error; MSE, mean square error; RMSE, root mean square error; SMAPE, symmetric mean absolute percentage error; SVM, support vector machine.

Table 4 above shows the DM test results between the models considered in this paper. The null hypothesis of the DM test is that the performance of model A and model B is the same. When the statistical results are meaningful, the hypothesis is rejected. If the DM test statistical value is positive, the model in the column performs better than the model in the row, and so on, and the greater the difference between the models, the greater the absolute value of the statistical value. It can be seen from Table 4 that the DM test statistics of GRU model, HAR-RV model, and XGBoost model are relatively large, indicating that the forecasting ability of GRU model is significantly better than HAR-RV model and XGBoost model. From the rightmost column in Table 4, the DM test statistics of the GRU model and other models are positive, indicating that the forecasting ability of the GRU model is better than all other models, which is consistent with the results shown in Table 2.

In summary, through the comparison and analysis of the evaluation criteria of the forecasting results with the KS test and the DM test, it can be found that the forecasting ability of the GRU model is the best among all models, and the addition of the GARCH-MIDAS factor further improves the forecasting accuracy.

The one-step forecasting result reflects the short-term forecasting ability of the model, and the multistep forecasting results can better reflect the stability of the model. Therefore, after the one-step forecasting, we also make a five-step forecasting shown in Table 5.

It can be seen from Table 5 that in the five-step forecasting results, the forecasting performance of the GRU model is still better than the machine learning model and the traditional econometric model on the whole, indicating that the deep learning model has better stability while ensuring the forecasting performance.

4 | CONCLUSION

Based on whether to consider the GARCH-MIDAS indicator, we have performed different models to forecast, and through a variety of evaluation criteria, KS test and DM test, it can be found that the forecasting accuracy of the deep learning GRU model is the highest, followed by machine learning, and the forecasting accuracy of the traditional econometric model is the worst whether short-term local volatility or long-term volatility. The macroeconomic indicator can significantly improve the forecasting accuracy of the model. In addition, correlation analysis and feature importance ranking indicate that the macroeconomic factor plays an important role in the volatility forecasting of the stock market.

ACKNOWLEDGMENTS

This research is funded by the National Natural Science Foundation of China (11901397 and 71973098), Ministry of Education, Humanities and Social Sciences project (18YJCZH153), Youth Academic Backbone Cultivation Project of Shanghai Normal University (310-AC7031-19-003021), General Research Fund of Shanghai Normal University (SK201720), Key Subject of Quantitative Economics of Shanghai Normal University (310-AC7031-19-004221), and Academic Innovation Team of Shanghai Normal University (310-AC7031-19-004228).

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Yuping Song has designed the framework of this paper and substantively revised it. Xiaolong Tang has performed the corresponding empirical volatility forecasting. Hemin Wang and Zhiren Ma have drafted the work and assisted code debugging and preprocessed data.

DATA AVAILABILITY STATEMENT

The dataset for the empirical analysis can be derived from Wind that is a service company in mainland China providing financial data and information as Bloomberg.

ORCID

Yuping Song https://orcid.org/0000-0002-7506-1719

REFERENCES

Alaali, F. (2020). The effect of oil and stock price volatility on firm level investment: The case of UK firms. *Energy Economics*, *87*, 104731. https://doi.org/10.1016/j.eneco.2020.104731

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905. https://doi.org/10.2307/2527343

Balli, F., Hajhoj, H. R., Basher, S. A., & Ghassan, H. B. (2015). An analysis of returns and volatility spillovers and their determinants in emerging Asian and middle eastern countries. *International Review of Economics & Finance*, *39*, 311–325. https://doi.org/10.1016/j.iref.2015.04.013

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, *121*, 823–866. https://doi.org/10.1162/qjec.121.3.823

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption----based explanation of aggregate stock market Behavior. *Journal of Political Economy*, *107*(2), 205–251. https://doi.org/10.1086/250059

Chen, W. (2018). Forecasting volatility of Shanghai composite index with deep learning. *Statistics & Information Forum*, *33*(5), 99–106.

Cho, K., Merrienboer, B., Gulcehre, C., Hdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv*, *1406*, 1078.

Chow, G. C., & Lin, A.-L. (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association*, *71*(355), 719–721. https://doi.org/10.1080/01621459.1976.10481554

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Engle, R. F., Gonzalo, J., & Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, *21*(3), 1187–1222. https://doi.org/10.1093/rfs/hhn004

Recchioni, M. C., Lori, G., Tedeschi, G., & Ouellette, M. S. (2020). The complete gaussian kernel in the multi-factor Heston model: Option pricing and implied volatility applications. *European Journal of Operational Research*, *293*(1), 336–360. https://doi.org/10.1016/j.ejor.2020.11.050

Reitz, T. A. (1988). The equity risk premium: A solution. *Journal of Monetary Economics*, *22*(1), 117–131. https://doi.org/10.1016/0304-3932(88)90172-9

Silvestrini, A., & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, *22*(3), 458–497. https://doi.org/10.1111/j.1467-6419.2007.00538.x

Xiao, D., Yue, Z., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. IJCAI'15: Proceedings of the 24th International Conference on Artificial Intelligence, 2327–2333.

Xiong, R., Nichols, E. P., & Shen, Y. (2016). Deep learning stock volatility with google domestic trends. arXiv:1512.04916v3.

AUTHOR BIOGRAPHIES

Yuping Song is an Associate Professor at the School of Finance and Business, Shanghai Normal University, China. He is an expert on analysis of financial large data.

Xiaolong Tang is a Master of Finance at the School of Finance and Business, Shanghai Normal University, China. His research interests are financial statistics and modeling.

Hemin Wang is a Master of Applied Economics at the School of Finance and Business, Shanghai Normal University, China. Her research interests are financial statistics and modeling.

Zhiren Ma is a Master of Applied Economics at the School of Finance and Business, Shanghai Normal University, China. His research interests are financial statistics and modeling.

How to cite this article: Song, Y., Tang, X., Wang, H., & Ma, Z. (2023). Volatility forecasting for stock market incorporating macroeconomic variables based on GARCH-MIDAS and deep learning models. *Journal of Forecasting*, *42*(1), 51–59. https://doi.org/10.1002/for.2899