Forecasting Sectoral Volatility with ESG Data Using GARCH MIDAS and Hybrid Machine Learning in China

Wenjing Cao

School of Economics, China Jiliang University, Hangzhou, 310018, China
constant@cjlu.edu.cn

Abstract. Environmental, Social, and Governance (ESG) ratings are increasingly recognized as potential drivers of stock volatility. However, the quarterly frequency of ESG ratings mismatches daily returns, posing a methodological challenge for forecasting. This study addresses this gap by systematically comparing nine volatility forecasting models across 31 Chinese A-share sectors from 2010 to 2025. The models include Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type models, Long Short-Term Memory (LSTM), eXtreme Gradient Boosting (XGBoost), and GARCH with Mixed Data Sampling (GARCH-MIDAS) and GARCH-MIDAS+LSTM hybrid models. The results reveal three main findings. GARCH-MIDAS+LSTM achieves the lowest forecast error (MSE=0.000814), with its MSE 52% lower than pure LSTM and 81% lower than GARCH(1,1). Mixed-frequency modeling via MIDAS substantially outperforms daily linear interpolation of quarterly ESG. Additionally, hybrid models do not universally beat traditional GARCH. In industries such as Power Equipment, GARCH(1,1) performs better, and adding GARCH conditional variance as LSTM input (GARCH-LSTM) degrades performance. This study confirms the advantage of the MIDAS structure for low-frequency ESG data. It also verifies the counterintuitive finding that deep learning does not always outperform simple econometric models, providing methodological guidance for volatility forecasting in sustainable finance.

Keywords: ESG rating, volatility forecasting, GARCH-MIDAS, machine learning.

1 Introduction

Environmental, Social and Governance (ESG) investing has become a major trend in the global financial market. In China, driven by the “dual-carbon” goals and stricter ESG disclosure regulations, the way ESG ratings affect asset prices has attracted growing attention. Volatility is a core metric for measuring financial risk. Whether it can be effectively predicted by ESG information has critical implications for risk management, asset pricing, and regulatory policy. ESG data are generally released quarterly, while stock price volatility is a daily or even higher-frequency phenomenon. This frequency mismatch makes a method problem: how to properly incorporate low-frequency ESG information into high-frequency volatility forecasting models.

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_153

Current studies have explored the relationship between ESG and volatility from various perspectives. Borkowski and Abdullazade compared Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type models with hybrid machine learning models (GARCH-LSTM, GARCH-RNN) based on U.S. sectoral data, and they found that the traditional MGARCH-DCC worked better than complex hybrid models, which is counterintuitive [1]. Ghani and Qin focused on climate-sensitive industries and used a regime-switching GARCH-MIDAS model to deal with low-frequency climate risk indicators, showing that the mixed-frequency methods can be useful in volatility forecasting [2]. In the Chinese market, Xu et al. found that ESG performance was negatively related to stock price volatility, and this effect became stronger in state-owned and declining firms. Even with these studies, two gaps remain [3]. First, many studies simply change ESG into daily frequency by interpolation, overlooking the potential benefits of mixed-frequency modeling. Second, there is still not enough systematic evidence about whether hybrid models outperform traditional GARCH at sectoral level in China.

To deal with these gaps, this study takes 31 Shenwan first-level industries in China as the sample and builds nine volatility forecasting models, with a focus on the GARCH-MIDAS mixed-frequency structure and the GARCH-MIDAS+LSTM hybrid model. The objectives are to determine whether mixed-frequency modeling outperforms daily interpolation, and whether hybrid models comprehensively beat traditional GARCH. The results are expected to provide empirical guidance for volatility forecasting in sustainable finance.

2 Methodology

2.1 Data

Data span from April 2010 to March 2025 across 31 Shenwan first-level industries in China's A-share market. This range covers multiple stages of volatility, such as economic upward, COVID-19 impact, recovery and adjustment.

All data are obtained from reliable public sources. Daily industry index returns are computed as log-transformed daily percentage changes of Shenwan first-level indices from the Wind database. Market volatility is proxied by the Chinese implied volatility index (iVX) (code 000188), which reflects the implied volatility of SSE 50 ETF options and serves as a control variable for investor sentiment. The risk-free rate is the one-year government bond yield, used as an exogenous variable. Industry quarterly ESG scores are derived from Huazheng ESG quarterly ratings.

Through industry mapping and equal-weighted averaging of individual stocks within each industry, quarterly composite scores and sub-scores for Environment (E), Social (S), and Governance (G) are obtained. Daily ESG is obtained by assigning the quarterly ESG value to all trading days within that quarter and then forward-filling missing values. Daily returns, market volatility, risk-free rates, daily interpolated ESG, and raw quarterly ESG are aligned by industry and date, then merged. After cleaning, the final dataset contains approximately 70,000 valid observations.

The annualized realized volatility (RV) over a rolling 22-trading-day window is used as the forecast target:

$$RV_{22d,t} = \sqrt{252} \times \sqrt{\frac{1}{21}\sum_{i=0}^{21}(r_{t-i} - \bar{r})^2} \quad (1)$$

where r_{t-i} is return of industry index on day $t - i$, and $\bar{r}$ is average return of past 22 trading days. This indicator is updated daily and represents industry volatility over the past month.

2.2 Models

This study develops nine models categorized into three groups to compare the volatility forecasting performance of traditional GARCH-type models, pure machine learning models, and mixed-frequency hybrid models. All models take the realized volatility $RV_{22d,t}$ as the forecasting target, but differ in input features and structure.

Benchmark Econometric Models. GARCH(1,1) is a classic conditional heteroskedasticity model, using lagged squared residuals and lagged conditional variance to capture volatility clustering. Conditional standard deviations are generated and then annualized to obtain volatility forecasts.

Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) adds an asymmetric term to the variance equation of GARCH(1,1) to capture the different impacts of positive and negative news on volatility.

GARCH with Exogenous variables (GARCH-X) includes exogenous variables in the GARCH(1,1) variance equation. The risk-free rate and the daily-interpolated industry quarterly ESG score are added as exogenous variables, to examine whether low-frequency information can improve forecasts through a linear specification.

Machine Learning Models. Long Short-Term Memory (LSTM) is a long short-term memory network. Input features include lagged returns over the past 10 trading days, lagged realized volatility over the past 10 trading days and the current risk-free rate. The network consists of a single LSTM layer with 50 neurons, a Dropout layer (rate = 0.2) and a fully connected output layer. It is trained with the Adam optimizer and mean squared error (MSE) loss.

EXtreme Gradient Boosting (XGBoost) is an ensemble learning model based on gradient boosting decision trees. It uses the identical input features as the LSTM model. The hyperparameters are set as follows: 100 decision trees, maximum tree depth of 3, and a learning rate of 0.1.

LSTM-ESG adds the daily-interpolated industry ESG composite score as an additional input feature to the LSTM.

Hybrid Models. GARCH-LSTM fits a GARCH(1,1) model on the full sample to obtain daily annualized conditional variances. These variances are then used as extra features together with the original LSTM inputs (lagged returns, lagged RV, risk-free rate) to train an LSTM. This design tests whether volatility information extracted by GARCH can supplement deep learning models.

GARCH with Mixed Data Sampling (GARCH-MIDAS) is a mixed-frequency model that directly handles the frequency mismatch between quarterly ESG and daily returns. It decomposes realized volatility into a long-term component and a short-term component. The short-term component follows a standard GARCH(1,1) process, capturing volatility clustering. The long-term component is specified as an exponential function of a weighted sum of past quarterly ESG scores, where the weights follow a Beta lag polynomial. The model avoids interpolating daily ESG indicators, and directly learns how low-frequency variables affect long-term volatility through MIDAS polynomials.

GARCH-MIDAS+LSTM adds the long-term component and short-term component estimated from GARCH-MIDAS as extra features into the LSTM. This is a "first decompose, then learn" deep hybrid model.

2.3 Experimental Design

All models are divided by time order. For each industry, the first 80% of trading days are used for training, and the rest 20% for testing, ensuring no look-ahead bias. The evaluation uses mean squared error (MSE) and mean absolute error (MAE).

The Diebold-Mariano (DM) test is used to compare the predictive accuracy of each model with the benchmark GARCH(1,1). The test statistic is calculated from the loss differential between two sets of forecast errors, using squared errors as the loss function. If it is statistically significant, it shows whether one model outperforms the other in statistical terms.

To examine industry heterogeneity, all models are run separately for each of the 31 industries, and then the results are put together to calculate the cross-industry average and standard deviation.

3 Results

3.1 Overall Performance across All Industries

Table 1 presents the mean and standard deviation of MSE and MAE for the nine models across all 31 industries.

Table 1. All-industry average prediction error (MSE, MAE) of each model.

Model	MSE		MAE	
	mean	std	mean	std
GARCH	0.004215	0.007048	0.038365	0.015154
EGARCH	0.004432	0.005797	0.040290	0.012917
GARCH-X	0.004215	0.007048	0.038365	0.015154
LSTM	0.001688	0.004184	0.018441	0.014919
XGBoost	0.002422	0.006859	0.017274	0.016102
LSTM-ESG	0.002032	0.004345	0.021485	0.016063
GARCH-LSTM	0.002009	0.004620	0.021438	0.016385
GARCH-MIDAS	0.003039	0.002744	0.036748	0.011552
GARCH-MIDAS+LSTM	**0.000814**	**0.002543**	**0.008657**	**0.010971**

First, GARCH-MIDAS+LSTM performs best. Its average MSE is 0.000814, 52% lower than pure LSTM and 81% lower than the benchmark GARCH(1,1) (DM $p < 0.01$).

Second, mixed-frequency models yield substantially lower errors than models using daily interpolated data. GARCH-MIDAS+LSTM (quarterly ESG) significantly outperforms LSTM-ESG (daily interpolation), and standalone GARCH-MIDAS performs similarly to LSTM-ESG.

Third, GARCH-LSTM does not improve over pure LSTM (MSE 0.002009 vs. 0.001688). This suggests that simple feature stacking may introduce redundancy or noise.

Fourth, EGARCH does not outperform the symmetric GARCH (MSE 0.004432 vs. 0.004215), indicating that volatility asymmetry is not a prevalent feature across industries in China's stock market.

3.2 Diebold-Mariano Significance Tests

Taking GARCH(1,1) as the benchmark, DM test statistics are computed for each model in each industry. Table 2 presents the DM test results for selected industries.

Table 2. DM test statistics (vs. GARCH) for selected industries.

Industry	Model	t statistic	p value	Significant
Banking	LSTM	8.643	<0.001	Yes
	XGBoost	8.610	<0.001	Yes
Utilities	LSTM	10.871	<0.001	Yes
	XGBoost	10.959	<0.001	Yes
Power Equipment	LSTM	-0.932	0.294	No
	XGBoost	-3.244	0.002	Yes (worse)
Social Services	LSTM	1.064	0.288	No
	XGBoost	1.022	0.307	No
Retail Trade	XGBoost	-0.231	0.366	No
	LSTM ESG	-0.261	0.411	No

The results show that the forecast errors of LSTM, XGBoost, LSTM-ESG and GARCH-LSTM are significantly lower than GARCH ($p < 0.01$) in most industries. Exceptions include Power Equipment, Social Services and Retail Trade. GARCH-MIDAS+LSTM significantly outperforms GARCH MIDAS in all industries except Beauty & Personal Care ($p = 0.879$). This verifies that LSTM can further enhance the MIDAS components.

3.3 Industry Heterogeneity

Although GARCH-MIDAS+LSTM performs best on average, forecasting difficulty varies considerably across different industries. Table 3 lists the MSE of three typical models for representative industries.

Table 3. MSE Comparison of three types of models in typical industries.

Industry	GARCH MSE	LSTM MSE	GARCH-MIDAS+LSTM MSE
Banking	0.00093	0.00011	0.00389×10-3
Utilities	0.00095	0.00013	0.00549×10-3
Environmental Protection	0.00973	0.00286	0.00196
Beauty & Personal Care	0.03994	0.02297	0.01397
Power Equipment	0.00579	0.00687	0.00276

For highly predictable industries such as Banking and Utilities, the MSE of GARCH-MIDAS+LSTM is below 0.00003, and the predicted values closely match the true realized volatility (RV). These industries share the characteristics of stable ESG ratings and strong correlations with return volatility.

In contrast, for industries that are difficult to predict, such as Environmental Protection and Beauty & Personal Care, even the best model yields an MSE above 0.001, which is higher than the benchmark GARCH errors of other industries. This may stem

from higher idiosyncratic volatility in these industries or weaker links between ESG ratings and stock price volatility.

For counterintuitive industries such as Power Equipment, the traditional GARCH(1,1) actually achieves a lower MSE than LSTM and XGBoost. It suggests that when volatility dynamics are simple (e.g., weak volatility clustering), linear parametric models are sufficient, while machine learning models tend to overfit noise.

4 Discussion

4.1 Superiority of the Mixed-Frequency Method over Daily Interpolation

This study finds that the MSE of GARCH-MIDAS+LSTM is much lower than LSTM-ESG (0.000814 vs. 0.002032), which uses linearly interpolated quarterly ESG. This result is consistent with Ghani and Qin and Liu et al. [2, 4]. A frequency gap naturally exists between low-frequency ESG indicators and high-frequency returns. Simple daily interpolation may create artificial smoothing effects and weaken the extra information contained in low-frequency variables. In contrast, GARCH-MIDAS uses a Beta weighting function to give higher weights to recent quarterly ESG adaptively, while retaining long-term trends. In this way, ESG information is brought into volatility forecasting more effectively.

4.2 Explanations for the Underperformance of Hybrid Models

Contrary to expectations, GARCH-LSTM performs worse than pure LSTM, and in industries such as Power Equipment, the benchmark GARCH(1,1) outperforms all machine learning models.

There are three possible explanations. First, feature redundancy may be responsible for the performance degradation. LSTM is capable of capturing volatility dynamics from lagged return sequences. Adding GARCH conditional variance introduces highly correlated variables, increasing the risk of overfitting [5].

Second, the simplicity of volatility dynamics in certain industries makes linear models sufficient. In industries with weak volatility clustering (e.g. Power Equipment), linear GARCH(1,1) is optimal. Machine learning models attempt to fit nonlinear patterns that do not exist, thereby increasing forecast errors.

Third, data frequency constraints limit the advantage of deep learning. ESG data is released on a quarterly basis, while deep learning models like LSTM require dense time steps to exert their strength in sequence learning. When low-frequency indicators dominate the model, the parametric MIDAS structure delivers higher efficiency.

This counterintuitive finding is consistent with the conclusions based on the U.S. market reported by Borkowski & Abdullazade, indicating that the fact that hybrid models do not always outperform traditional econometric models is likely a cross-market phenomenon [1].

4.3 Limitations and Future Directions

This study has some limitations. First, the ESG data are only available at quarterly frequency, limiting the ability of hybrid models to capture high-frequency volatility information. Second, this research only adopts ESG ratings from Huazheng Index; disagreements across rating agencies are not taken into account. Existing research shows that divergences in ESG ratings can substantially increase stock price volatility [6]. Third, the hybrid models are based solely on LSTM, without exploring more advanced structures such as Transformer and temporal convolutional networks.

Future research can improve along the following directions. First, higher-frequency (e.g., monthly) ESG data can be adopted. Second, an ESG rating divergence index can be introduced as a moderating variable. Third, innovative hybrid architectures including KAN-GARCH-MIDAS (Kolmogorov-Arnold Network, KAN) can be explored [4]. Fourth, empirical analysis can be extended to cross-market settings to test the robustness of research conclusions [7-10].

5 Conclusions

This study compares nine volatility forecasting models across 31 Chinese A-share industries (2010-2025), including GARCH-type models, pure machine learning models (LSTM, XGBoost), and mixed-frequency models with quarterly ESG (GARCH-MIDAS, GARCH-MIDAS+LSTM).

First, GARCH-MIDAS+LSTM achieves the lowest forecast error (MSE 52% below LSTM and 81% below GARCH(1,1)). This confirms that MIDAS with LSTM improves forecasting accuracy.

Second, mixed-frequency modeling outperforms daily interpolation. GARCH-MIDAS beats LSTM-ESG, and GARCH-MIDAS+LSTM improves even further. This indicates that retaining the original frequency of low-frequency data is more effective than simple interpolation.

Third, hybrid models do not universally outperform traditional GARCH. In industries like Power Equipment, GARCH(1,1) produces smaller forecast errors than all machine learning models, and GARCH-LSTM performs worse than pure LSTM. This suggests that linear models are sufficient when volatility dynamics are simple, and feature redundancy may harm deep learning.

Limitations include quarterly ESG from one source and hybrid models only based on LSTM. Future research may use monthly ESG, divergence index, or advanced architectures like KAN-GARCH-MIDAS. This study confirms the advantage of mixed-frequency methods for low-frequency ESG and verifies the counterintuitive finding that hybrid machine learning does not always outperform traditional GARCH. In practice, GARCH-MIDAS provides a cost-effective volatility forecast for institutions with only quarterly ESG data.

References

1. Borkowski, P., Abdullazade, Z.: ESG indicators and sectoral volatility: GARCH vs. hybrid machine learning. Finance Research Letters 85(PC), 108103 (2025)
2. Ghani, M., Qin, Q.: Forecasting climate-sensitive industries' volatility: A regime-switching GARCH-MIDAS approach with multiple climate risk indicators. International Review of Financial Analysis 105, 104412 (2025)
3. Xu, Z., Liu, D., Li, Y., Guo, F.: ESG and Stock Price Volatility Risk: Evidence from Chinese A-share Market. North American Journal of Economics and Finance 75(PA), 102277 (2025)
4. Liu, T., Choo, W., Xinping, H., Li, L.: Combining deep learning with econometric models: volatility forecasting using the KAN-GARCH-MIDAS framework. Journal of Applied Economics 28(1), 2555479 (2025)
5. Han, B., Liu, A., Chen, J., Knottenbelt, W.: Can machine learning models better volatility forecasting? A combined method. The European Journal of Finance, 1–22 (2025)
6. Ding, H., Yang, W.: ESG rating divergence, investor confidence and stock price volatility. Communication of Finance and Accounting (11), 61–64 (2025)
7. Awijen, H., Jabeur, B. S., Pillot, J.: Interpretable machine learning models for ESG stock prices under transition and physical climate risk. Annals of Operations Research 357(1), 1-31 (2024)
8. Rosinus, M., Lansky, J.: Predictive Power of ESG Factors for DAX ESG 50 Index Forecasting Using Multivariate LSTM. International Journal of Financial Studies 13(3), 167 (2025)
9. Onomakpo, H.O.: The Illusion of Predictability: Deconstructing the ESG-Alpha Link with Robust Econometrics and Validated Machine Learning. International Journal of Management Research and Economics 5(2), 35-54 (2025)
10. Palanisamy, K., Parthasarathy, K.: Stock returns and volatility: Evidence from Nifty 50 index. Asian Journal of Research in Social Sciences and Humanities 6(5), 434–442 (2016)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

