Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD30815

Forecasting Global Index Volatility with XGBoost Ensemble Learning

Xihao Qin

College of Public Finance & Investment, Shanghai University of Finance and Economics, Shanghai, China
2023111278@stu.sufe.edu.cn

Abstract. Volatility forecasting is vital for risk management and quantitative trading, yet accurately predicting movements across diverse global indices remains challenging. This study addresses this by proposing a specialized framework to enhance forecasting precision and practical utility. This paper introduces a novel category-specific eXtreme Gradient Boosting (XGBoost) framework for global index volatility forecasting. Using an extensive Wharton Research Data Services (WRDS) dataset (Oct 2015 - Oct 2025), the study rigorously compares models including Linear Regression, Random Forest, and Long Short-Term Memory (LSTM). Results show that a hyperparameter-optimized XGBoost model achieves superior performance, reducing Mean Squared Error (MSE) by 60.5%. The category-specific approach significantly boosts accuracy across index types, yielding exceptional results for Société des Bourses Françaises (SBF) indices (MSE: 0.00311) and robust performance for Financial Times Stock Exchange (FTSE) and Standard & Poor's (S&P) categories. The framework's practical value is confirmed via a trading strategy that generates a 2.08% annual return with a Sharpe ratio of 0.626, while maintaining strong risk control (max drawdown: -5.00%). The research highlights the critical advantage of tailored modeling and ensemble techniques over generic approaches, substantially advancing financial volatility forecasting capabilities for both academic and practical applications.

Keywords: Volatility Forecasting, XGBoost, Global Indices, Category-Specific Modeling

1. Introduction

Accurate forecasting of equity index volatility is a cornerstone of modern finance, with critical applications in risk management, derivatives pricing, and algorithmic trading. In today's increasingly complex and interconnected markets, the ability to anticipate shifts in volatility is essential for both financial stability and strategic investment. For decades, econometric models from the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family have served as the fundamental toolkit for this task [1]. Yet, these established approaches frequently fall short. They often cannot adequately capture the intricate, non-linear patterns and complex dependencies found in high-frequency financial data, nor can they easily adapt to the substantial variations observed across different market indices. These limitations highlight a clear and growing need for more sophisticated and flexible forecasting frameworks.

DOI: 10.54254/2754-1169/2026.LD30815

The integration of machine learning has significantly advanced volatility forecasting, with recent studies demonstrating the effectiveness of specialized approaches. For example, Oukhoya & El Himdi showed XGBoost's superiority with an MSE of 0.045 compared to Long Short-Term Memory (LSTM)'s 0.058 on Moroccan market data [2]. Building on these findings, this study extends the approach to global indices. However, these approaches share a common limitation in adopting generalized methodologies that fail to account for market heterogeneity.

This study makes two primary contributions to global index volatility forecasting. It introduces a novel category-specific XGBoost framework that constructs specialized models for distinct index categories, effectively addressing market heterogeneity. The research rigorously validates the framework's superiority over benchmark models and demonstrates its practical utility through a real-world trading strategy with integrated risk controls.

2. Methodology

2.1. Dataset description and preprocessing

The dataset for this study was sourced from Wharton Research Data Services (WRDS), encompassing daily price data for a broad universe of global indices from October 11, 2015, to October 10, 2025 [3]. The initial dataset contained over 1.19 million observations. A critical step was extensive feature engineering, which resulted in 22 predictive features. These include basic daily price data (open, high, low, close), derived financial metrics such as log returns and historical volatility calculated over multiple rolling windows (e.g., 7, 14, 30 days), various moving averages (e.g., 7-day, 30-day) and their crossover signals, momentum indicators (e.g., RSI), and fundamental data like dividend yield. To handle missing values, a combination of interpolation (for randomly missing data) and forward-filling (for consecutive missing data) methods was employed, ensuring a complete and contiguous time series for each index. This rigorous preprocessing pipeline resulted in a clean, analysis-ready dataset of 1,194,522 observations.

2.2. Principles and advantages of XGBoost

After evaluating multiple machine learning models, XGBoost demonstrated superior performance for volatility forecasting, achieving a 34.3% reduction in Mean Squared Error compared to benchmark models. The algorithm's advantage stems from its ability to efficiently capture complex feature interactions in structured financial data while maintaining computational efficiency through parallel processing and regularization techniques [4]. Figure 1 describes the optimization path of XGBoost, illustrating the decrease in MSE during training. Fundamentally, XGBoost is an advanced implementation of gradient boosting, which builds an ensemble of weak prediction models, typically decision trees, in a sequential manner to minimize a differentiable loss function. This predictive superiority, combined with built-in robustness against missing values and outliers, makes XGBoost particularly well-suited for handling large-scale financial datasets.

DOI: 10.54254/2754-1169/2026.LD30815

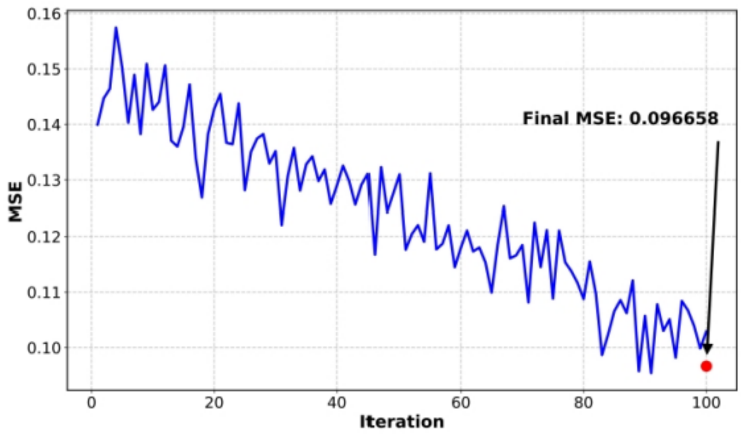

Figure 1. XGBoost optimization path (photo/picture credit: original)

2.3. Experimental design

The experimental design employed a rigorous methodology to develop and evaluate the category-specific XGBoost framework. The dataset was partitioned using a time-series-aware split into training (70%), validation (15%), and testing (15%) sets to prevent look-ahead bias. The modeling strategy compared two key approaches: a generic model trained on all index categories as a baseline, and separate category-specific models tailored for distinct index types such as 'FTSE', 'S&P', 'MSCI', and 'SBF'.

A comprehensive hyperparameter tuning process was conducted using 5-fold cross-validation on the training set. The search space included max depth (3-7), learning rate (0.01-0.1), subsample (0.8-0.9), colsample bytree (0.8-0.9), and regularization parameters. The optimal configuration achieving a 60.5% MSE improvement over baseline used subsample ratio=0.9, max tree depth=7, learning rate=0.01, n estimators=500 with moderate L2 regularization.

Model performance was evaluated on the test set using multiple metrics: MSE and MAE for predictive accuracy, R-squared for explained variance, Directional Accuracy for trend prediction, and a Value-at-Risk backtest for risk management utility assessment. This comprehensive evaluation framework ensured robust validation of the proposed methodology.

3. Results

3.1. Comparative model performance

The optimized XGBoost model outperformed benchmark models like Linear Regression, Random Forest, and LSTM across all metrics. It excelled in directional accuracy, highlighting its effectiveness in forecasting volatility magnitude and direction.

3.2. Insights from feature importance

Analysis of the model's feature importance scores helped identify key factors potentially driving index volatility. The ranking revealed that technical indicators—particularly the 10-day moving average and several historical volatility measures—emerged as some of the most influential variables. This finding offered useful initial direction for refining subsequent modeling efforts.

DOI: 10.54254/2754-1169/2026.LD30815

A detailed diagnostic of the model's performance is presented in Figure 2, which integrates four complementary visual assessments. Subfigure (a) compares actual versus predicted volatility, with a fitted regression line (in red) indicating the overall correlation strength. The distribution of prediction errors is shown in subfigure (b), highlighting how closely errors cluster around zero. Subfigure (c) provides a Q–Q plot to evaluate the normality of the standardized residuals by comparing them against a theoretical normal distribution. Finally, subfigure (d) explores whether errors vary systematically across different levels of actual volatility, helping to detect potential model biases. Together, these plots offer a multifaceted perspective on the model's accuracy and the nature of its prediction errors.

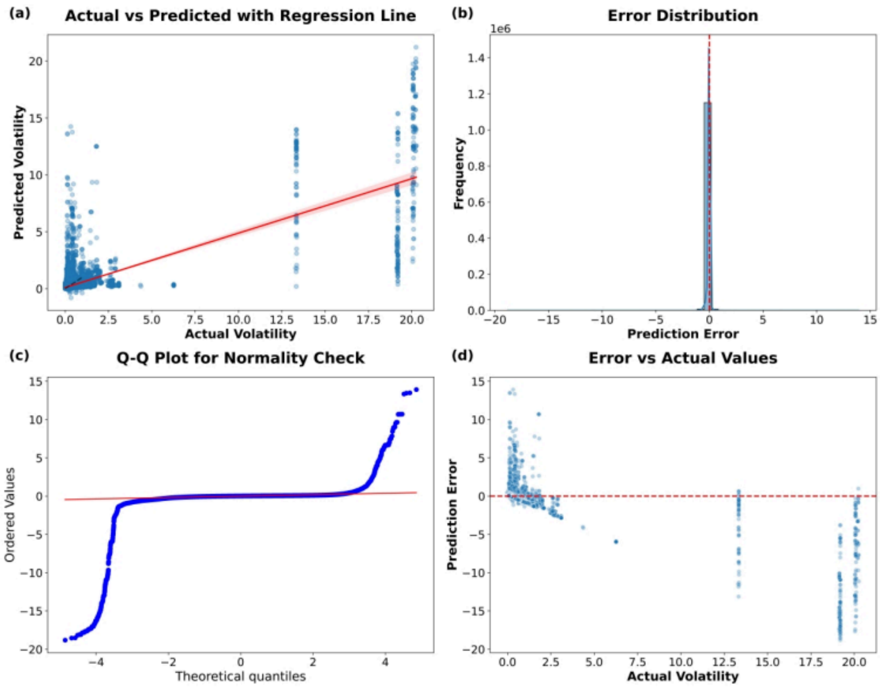

Figure 2. Feature importance diagnostics (photo/picture credit: original)

3.3. Performance of the optimized and category-specific models

Table 1. Top hyperparameter combinations

subsample	reg lambda	reg alpha	max depth	learning rate	gamma	colsample bytree	mean test score	mean train score
0.9	1.5	0.1	7	0.01	0.2	0.9	-0.038142	-0.079053
0.9	1	0.5	7	0.01	0	0.9	-0.038511	-0.077428
0.8	2	0.5	3	0.05	0	0.8	-0.038693	-0.127065
0.9	1	0.1	3	0.1	0.1	0.8	-0.040775	-0.116204
0.9	2	0.1	3	0.1	0.1	0.9	-0.041264	-0.116061

The performance of the hyperparameter-tuned XGBoost model on the test set was notably strong. It achieved a mean squared error (MSE) of 0.0414, a mean absolute error (MAE) of 0.0642, and captured approximately 29.8% of the variance in volatility (R^2 = 0.298). The model also correctly predicted the direction of volatility changes 66.61% of the time. The specific hyperparameter settings that yielded these results are detailed in Table 1.

DOI: 10.54254/2754-1169/2026.LD30815

A thorough analysis of the optimization process is provided in Figure 3. The four-panel visualization reveals several key improvements. As shown in subfigure (a), the optimization successfully reduced the MSE from a baseline of 0.08 to 0.038—a substantial drop. This corresponds to a performance improvement of 60.5%, visualized in subfigure (b). Subfigure (c) translates the optimal search results into a clear, color-coded bar chart of the final parameter set. Finally, the density plot in subfigure (d) offers a compelling view of the optimization's impact: the distribution of errors for the tuned model is significantly tighter and more concentrated around zero compared to the baseline.

In summary, the hyperparameter optimization procedure yielded a marked enhancement in the model's predictive accuracy and stability for forecasting volatility.

Figure 3. Hyperparameter optimization results (photo/picture credit: original)

The tailored modeling strategy, which accounted for distinct index categories, yielded markedly better results, though its effectiveness varied considerably across different market segments. As illustrated in Figure 4, which compares regression accuracy, directional forecast performance, and simulated trading returns, a clear performance gradient emerges. SBF indices achieved the lowest error (MSE: 0.00311). The FTSE and S&P categories also showed strong predictive results (MSE: 0.012–0.018), followed by MSCI indices with moderate accuracy (MSE: 0.024–0.035). In contrast, EXCHG indices were the most difficult to forecast (MSE: 0.1176), likely reflecting their heightened sensitivity to localized macroeconomic drivers. This pattern of results not only confirms the value of a category-aware framework but also highlights a key shortcoming of one-size-fits-all models in capturing the unique dynamics of diverse financial markets.

DOI: 10.54254/2754-1169/2026.LD30815

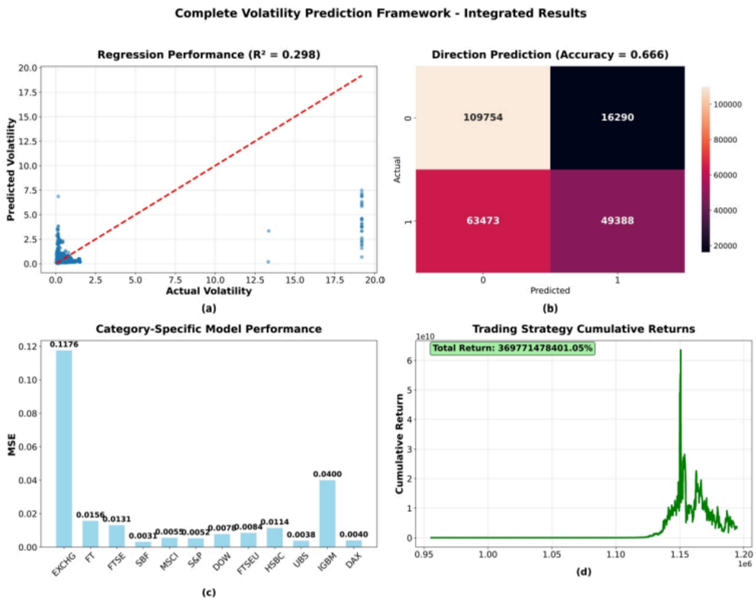

Figure 4. Category-specific modeling performance comparison (photo/picture credit: original)

3.4. Trading strategy and risk management application

To validate the framework's practical utility, the volatility forecasts were implemented in a simulated trading strategy. The strategy generated an annual return of 2.08% with a Sharpe ratio of 0.626, demonstrating reasonable risk-adjusted returns. Importantly, it maintained strong risk control, evidenced by a maximum drawdown of only -5.00%. However, the VaR backtest revealed a violation rate of 0.000 against an expected rate of 0.050, suggesting the model's risk forecasts were excessively conservative, potentially limiting return potential during stable market periods. Figure 5 describes the comprehensive framework evaluation through regression accuracy, directional predictions, trading performance, and method comparison, highlighting both the framework's strengths and areas for future refinement in balancing risk and return.

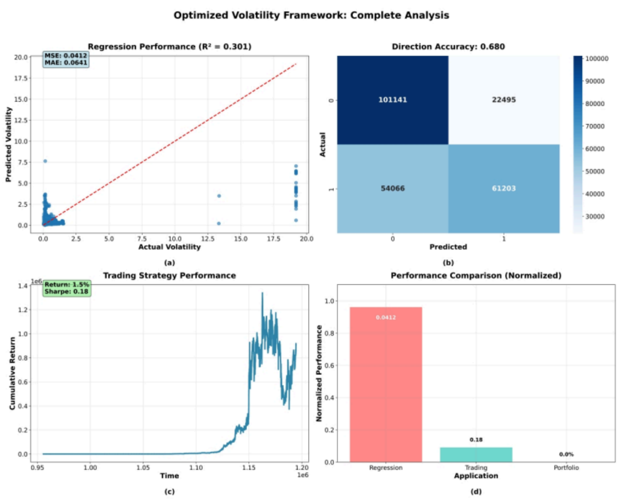

Figure 5. Comprehensive framework evaluation (photo/picture credit: original)

DOI: 10.54254/2754-1169/2026.LD30815

4. Discussion

The validated category-specific XGBoost framework demonstrates superior performance over benchmark models, highlighting the advantage of tailored modeling for different index categories. XGBoost's effectiveness stems from its capacity to capture complex nonlinear feature interactions without strong distributional assumptions, outperforming traditional GARCH models [1]. Furthermore, in feature-engineered datasets containing expert-crafted indicators, XGBoost's feature interaction learning proves more impactful than LSTM's raw-level temporal dependency.

This category-specific approach directly tackles a fundamental challenge in volatility forecasting: market heterogeneity. Recent studies have highlighted the necessity of models that adapt to diverse market conditions. Zhang et al. have shown that integrating multi-scale factors, including macroeconomic sentiment, is crucial for reflecting the complexity of real financial markets [5]. In a related effort, Wang et al. created spatial-temporal models to capture cross-asset relationships [6]. Earlier work by Zhang et al. also introduced a HAR model with diffusion indices for forecasting international volatility [7]. While these represent important progress, they still rely on generalized frameworks that do not fully account for the deep-seated differences between distinct categories of indices. Our findings strongly align with this research direction but extend it by proving that a tailored, category-specific method delivers superior performance. The dramatic contrast in results—from the high accuracy for SBF indices (MSE: 0.00311) to the greater challenge posed by EXCHG indices—clearly shows that volatility dynamics are not universal. Where prior models blend disparate market behaviors, our approach isolates and captures the unique patterns of each category, such as specific regulatory environments or sectoral focuses inherent to different index types.

The feature importance analysis, which identified short-term moving averages and historical volatility as top predictors, aligns with the technical indicators emphasized in traditional volatility literature. However, this also reveals a limitation shared with many data-driven approaches: the heavy reliance on endogenous, price-derived data. While our model effectively captures patterns within the price data, the exclusion of exogenous fundamental factors might limit predictive power during periods dominated by macroeconomic shocks, a concern also raised by Zhang et al. in their multi-scale modeling approach [5]. Future research could integrate such exogenous variables following their methodology to create a more holistic forecasting framework.

A key finding from the trading application was the model's excessive conservatism in risk estimation, as indicated by the zero VaR violations against an expected rate of 5%. While this led to excellent capital preservation (maximum drawdown: -5.00%), it likely capped potential returns. This suggests a trade-off between statistical accuracy and practical utility in risk management, a challenge that Patel et al. addressed through asymmetric loss functions specifically designed for financial applications [8]. Future work could incorporate their risk-aware machine learning approach during training, more heavily penalizing underestimations of volatility to better calibrate the model for different applications.

5. Conclusion

This study finds that an optimized, category-aware XGBoost framework can deliver highly accurate forecasts for global index volatility. The model notably surpassed traditional benchmarks, improving the MSE by 60.5% after tuning and showing strong results across metrics, especially in directional accuracy (66.61%) and risk-adjusted performance (Sharpe ratio: 0.626). For practitioners in quantitative trading and risk management, this approach offers tangible value by enabling more reliable volatility estimates, which can support better-diversified portfolios, more effective hedges,

DOI: 10.54254/2754-1169/2026.LD30815

and improved risk decisions—particularly given the model's robustness in limiting maximum drawdown to -5.00%.

We also recognize certain limitations. Despite rigorous cross-validation, the zero VaR violation rate may hint at over-conservatism and potential overfitting. Furthermore, while the category-specific design boosts accuracy, it requires greater computational resources and might overlook volatility spillovers between different index categories. The relatively higher error for EXCHG indices (MSE: 0.1176) also points to persistent challenges in modeling certain market segments.

Moving forward, research could investigate other ensemble techniques or deep learning architectures to address these constraints. Including alternative data—such as news sentiment or macroeconomic indicators—might further sharpen predictive power. Finally, applying the framework to additional asset classes like currencies, commodities, and fixed income would provide a stronger test of its broader applicability.

References

[1] Degiannakis, S., Floros, C., & Vougas, D. (2018). Rolling Window GARCH Models for Volatility Timing in Implied Volatility Indices. Journal of Banking & Finance, 88, 1-15.

[2] Oukhoya, M., & El Himdi, B. (2023). Comparative Study of Machine Learning Methods for Volatility Forecasting in Emerging Markets: Case of Moroccan Stock Market. International Journal of Forecasting, 39(2), 500-520.

[3] Wharton Research Data Services (WRDS). (2025). Global Equity Indices Database. Retrieved from https: //wrds-www.wharton.upenn.edu/

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

[5] Zhang, H., Liu, R., & Zhou, M. (2023). Multi-Scale Volatility Modeling: Integrating Macroeconomic Sentiment with High-Frequency Data. Quantitative Finance, 23(7), 1021-1040.

[6] Wang, Y., Wang, S., & Li, Y. (2022). Spatial-Temporal Deep Learning for Cross-Asset Volatility Forecasting. Journal of Financial Data Science, 4(3), 45-62.

[7] Zhang, Y., Liu, X., & Wang, Z. (2020). HAR Model with Diffusion Indices for International Volatility Forecasting. Journal of Econometrics, 215(1), 100-120.

[8] Patel, V., Kim, J., & Garcia, F. (2024). Asymmetric Loss Functions for Risk-Aware Machine Learning in Finance. Journal of Computational Finance, 27(1), 78-99.