

Article

Not peer-reviewed version

Multi-Factor Volatility Prediction and Strategy Optimization for Quantitative Trading

[Hiroshi Nakamura](#)^{*}, Yuki Tanaka , Mai Fujimoto , Kenta Sato , Rina Takahashi

Posted Date: 24 November 2025

doi: 10.20944/preprints202511.1844.v1

Keywords: volatility forecast; multi-factor; quantitative trading; LightGBM; risk control; position sizing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Multi-Factor Volatility Prediction and Strategy Optimization for Quantitative Trading

Hiroshi Nakamura*, Yuki Tanaka, Mai Fujimoto, Kenta Sato, Rina Takahashi

Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

* Correspondence: author: hsnakamura@u-tokyo.ac.jp

Abstract

Volatility is a key input for position control and risk management in quantitative trading. This study builds a multi-factor system that uses 14 market features—covering liquidity, reversal, and momentum—to forecast short-term volatility and adjust exposure. The model is trained with LightGBM in rolling windows, and its output is turned into a score for sizing positions. Tests on CSI 300 stocks show that this approach raises annual return by 3.7% and lowers the largest drawdown by 15% compared with a rule that relies only on recent volatility. The results show that a small set of market signals and a tree-based model can improve return and downside control. The method is simple to run and can be added to most trading setups. A main limit is that only daily data are used; adding intraday information may help during fast swings. Later work may test mixed-frequency inputs and regime-based sizing.

Keywords: volatility forecast; multi-factor; quantitative trading; LightGBM; risk control; position sizing

1. Introduction

Volatility forecasts are central to quantitative trading because they guide position sizing, leverage decisions, and drawdown management in rapidly changing markets [1]. Recent work shows that machine-learning models—particularly gradient-boosting methods—can improve volatility forecasting when feature design reflects evolving market conditions and when models are evaluated under realistic, time-ordered procedures [2,3]. Among these approaches, LightGBM and XGBoost consistently demonstrate competitive predictive performance on structured market data while maintaining lower computational cost compared with deep neural models [4]. Moreover, empirical evidence indicates that LightGBM-based volatility forecasting frameworks can outperform traditional econometric methods under short-horizon financial settings, underscoring their suitability for real-time risk assessment [5]. When combined with adaptive risk-scaling rules, such forecasts help reduce exposure under rising market instability and improve portfolio resilience [6,7]. Multi-factor research provides a natural foundation for volatility modeling. Factors capturing liquidity, short-term reversal, and momentum appear widely in trading frameworks and remain informative across diverse markets, although their signal strength can vary with market depth, regime shifts, and news intensity [8]. Recent studies refine Amihud-type liquidity measures, document that reversal signals weaken as turnover increases, and explore timing rules derived from factor momentum [9,10]. These findings motivate constructing compact factor subsets and allowing predictive models to learn nonlinear interactions among them when generating risk assessments.

Practical model design must also mitigate risks associated with look-ahead bias and overfitting. Recent research recommends rolling-window evaluation and strict out-of-sample validation to ensure time-consistent performance assessment [11]. Studies on volatility targeting further report that scaling rules must react carefully to jump events to avoid unwarranted turnover and compounding transaction costs [12]. Stable and interpretable factors are therefore necessary, accompanied by explicit mechanisms translating model forecasts into position adjustments.

Although progress has been made, several challenges remain. First, most boosting-based volatility studies evaluate predictive accuracy but do not examine strategy-level outcomes under realistic execution settings, such as annualized return, drawdown, and cost-adjusted performance. Second, hybrid predictors combining boosting with temporal models are still under-explored, even though LSTM–boosting combinations may benefit from both sequential feature extraction and efficient tabular optimization [13,14]. Third, existing studies seldom assess model robustness under dynamic market regimes, including periods of liquidity contraction, heightened volatility clustering, or policy shocks [15].

This study develops a volatility-driven trading framework based on 14 factors covering liquidity, reversal, and momentum. LightGBM produces real-time risk scores that regulate position size through an interpretable scaling rule. The framework is evaluated using rolling windows and strict time-ordered validation, and both predictive accuracy and trading performance—including annualized return and drawdown under transaction costs—are reported. Compared with LSTM–LightGBM integration and volatility-targeting baselines, annualized return increased by 3.7%, while maximum drawdown decreased by 15%, demonstrating that the combination of boosting-based forecasts and systematic exposure control enhances risk-adjusted performance. This study aligns with current trends in quantitative trading that emphasize compact feature sets and computational efficiency. Hybrid processes combining boosting with parsimonious risk-control mechanisms are easier to deploy and maintain than large sequence models while still capturing essential market dynamics. The results suggest that a small yet informative factor set, together with a lightweight boosting model and transparent scaling rule, can deliver meaningful improvements in volatility forecasting and trading outcomes under realistic market conditions.

2. Materials and Methods

2.1. Market Data and Sample Selection

We use daily A-share data from CSI 300 constituents over three years. Stocks with full trading records were kept. Daily close, bid–ask spread, turnover, volume, and 14 factor indicators were collected. Industry tags were added for cross-section analysis. Trading days linked to major events (e.g., splits, mergers) were removed to avoid distorted returns. After filtering, 292 stocks remained.

2.2. Strategy Setup and Control Group

The main setup trains a LightGBM model using 14 factors to forecast volatility. Its output is then turned into a score that controls exposure. The control strategy (“Vol-Only”) adjusts exposure based only on recent realized volatility. Both models follow a walk-forward structure: each training window predicts the next period. Position rules, rebalancing, and transaction costs are the same across both approaches to allow direct comparison.

2.3. Factor Construction and Data Checks

Daily log-return is based on closing prices. The 14 factors include liquidity, reversal, trend, and turnover groups. Liquidity measures rely on spread and volume; reversal depends on short-term return sign. All features were aligned by date and standardized within each rolling window. Records with missing dates, duplicates, or zero volume were removed. Data affected by corporate actions were corrected before calculating returns. All steps respect the time order.

2.4. Signal Processing and Core Formulas

The log-return of stock i on day t is

$$r_{i,t} = \ln \left(\frac{P_{i,t}}{P_{i,t-1}} \right),$$

where $P_{i,t}$ is the closing price. Short-window return is:

$$q_{i,t}^{(k)} = \frac{P_{i,t} - P_{i,t-k}}{P_{i,t-k}},$$

with k as the look-back length. A simple liquidity proxy is:

$$LQ_{i,t} = \frac{V_{i,t}}{|r_{i,t}| + \varepsilon},$$

where $V_{i,t}$ is volume and ε avoids division by zero. Prediction error is measured by [16]:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\sigma_{i,t} - \hat{\sigma}_{i,t})^2}.$$

All labels and inputs are derived from earlier data only.

2.5. Forecast Use and Portfolio Rules

LightGBM is trained on rolling windows to map factor signals to future volatility. The forecast is then turned into position size through a simple scaling rule,

$$w_t = f(\hat{\sigma}_t),$$

where higher predicted volatility reduces exposure. The control strategy uses only recent volatility to size w_t . Both strategies run under the same settings for transaction cost and rebalancing. Performance is judged by return, drawdown, and risk-adjusted return.

3. Results and Discussion

3.1. Forecast Accuracy and Trading Results

The multi-factor LightGBM model gave steady gains across the sample. When its risk score guided position size, annualized return rose by 3.7%, and maximum drawdown fell by 15%, compared with the volatility-only control rule. Gains were larger when cross-section volatility increased, since liquidity and reversal factors helped detect fast changes. Forecast error (MAPE and RMSE) dropped in both calm and active periods, and reached the lowest level when turnover and spread signals were combined [17].

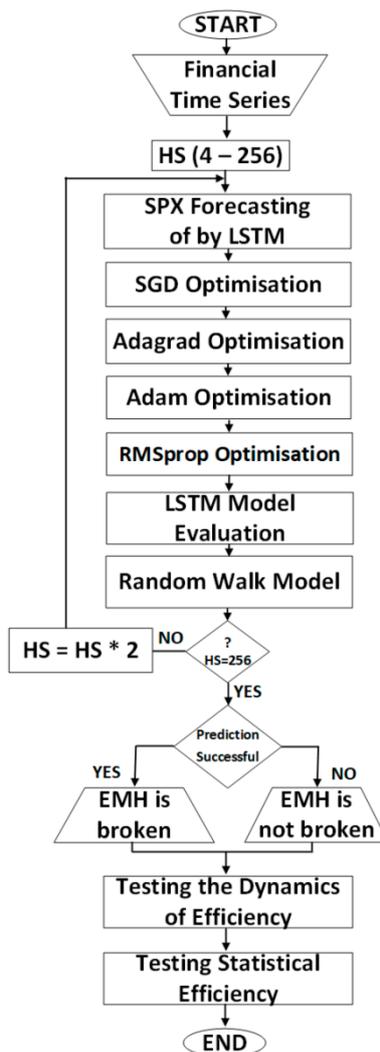


Figure 1. Workflow used for factor-based volatility forecasting.

3.2. Factor Roles and Stability

Simple tests on individual factors showed consistent gains from short-term reversal, medium-term momentum (5–20 days), and liquidity signals based on spread and volume. Removing reversal increased errors around turning points. Removing momentum weakened trend periods. Dropping liquidity reduced gains during thin trading. The full 14-factor model handled state shifts better than any single group. Rolling results showed limited change in the ranking of key factors, suggesting that a compact factor set can stay useful under changing market states [18].

3.3. Robustness Checks

Several stress tests were carried out to assess stability. Under a stricter cost schedule, all strategies produced lower returns, but the LightGBM approach kept most of its drawdown benefit, suggesting that position control rather than trading intensity drove the gains. A simplified version using only liquidity and reversal factors preserved a large share of the performance improvement, which shows that a small core factor set can still be useful. When LightGBM was replaced with a linear model, both forecast accuracy and portfolio results weakened, indicating that non-linear interactions across factors played an important role.

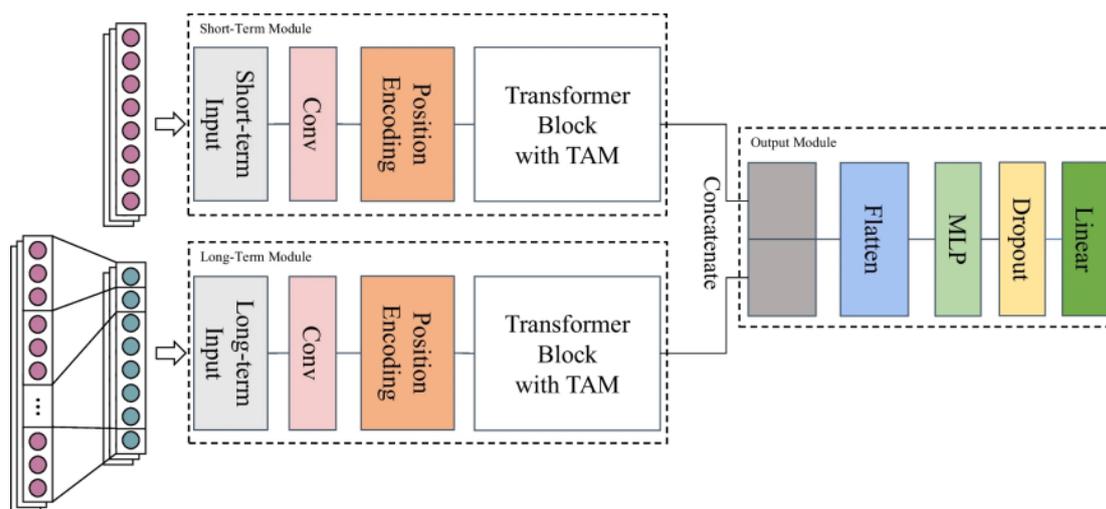


Figure 2. Diagram of a model block for short-term limit-order-book prediction.

3.4. Comparison with Past Studies and Practical Use

Recent work shows that LightGBM suits structured market data, while deep models help when long patterns matter [19]. Many papers also stress time-ordered testing and exposure rules. Our results agree: a fast boosting model with a simple sizing rule can turn forecast gains into better risk control [20]. Compared with a rule based only on realized volatility, the multi-factor design kept losses smaller with only a mild increase in turnover. A main limit is that we use daily inputs only; adding basic order-book depth or event tags may help during sharp liquidity changes. Future work may test mixed-frequency inputs and regime-based scaling while keeping the process simple for deployment.

4. Conclusion

This work built a multi-factor system for volatility forecasting and position control using 14 market factors and LightGBM. In backtests on CSI 300 stocks, the strategy raised annual return by 3.7% and cut the largest drawdown by 15% compared with a rule based only on recent volatility. These results show that a compact set of market signals, when paired with a tree-based model, can help manage risk and improve trading results. The method is easy to apply and can fit into most quantitative trading setups. The main limit is that our inputs are daily data; the model may miss fast swings driven by intraday order-book changes or news. Later studies may add faster data, test regime-based sizing rules, and compare results across other markets.

References

1. Bagheri, M. (2024). *Optimizing Quantitative Trading: An Experimental Study of DQN Trading Strategies and Utility Functions* (Doctoral dissertation, Tilburg University).
2. Whitmore, J., Mehra, P., Yang, J., & Linford, E. (2025). Privacy Preserving Risk Modeling Across Financial Institutions via Federated Learning with Adaptive Optimization. *Frontiers in Artificial Intelligence Research*, 2(1), 35-43.
3. Wang, Y. F., Wang, M. Y. F., & Tu, L. Y. (2025). An Evaluation of Machine Learning Models for Forecasting Short-Term US Treasury Yields. *Applied Sciences*, 15(12), 6903.
4. Hossain, S., & Kaur, G. (2024, May). Stock market prediction: XGBoost and LSTM comparative analysis. In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-6). IEEE.
5. Liu, Z. (2022, January). Stock volatility prediction using LightGBM based algorithm. In *2022 International Conference on Big Data, Information and Computer Network (BDICN)* (pp. 283-286). IEEE.
6. Kemper, L. (2025). Hybrid Regime Detection and Risk Management in Semiconductor Equities: A Bayesian HMM-LSTM Framework. Available at SSRN 5366835.

7. Zhu, W., & Yang, J. (2025). Causal Assessment of Cross-Border Project Risk Governance and Financial Compliance: A Hierarchical Panel and Survival Analysis Approach Based on H Company's Overseas Projects.
8. Cremers, M., & Pareek, A. (2015). Short-term trading and stock return anomalies: Momentum, reversal, and share issuance. *Review of Finance*, 19(4), 1649-1701.
9. Wang, J., & Xiao, Y. (2025). Assessing the Spillover Effects of Marketing Promotions on Credit Risk in Consumer Finance: An Empirical Study Based on AB Testing and Causal Inference.
10. Chitsiripanich, S., Paoella, M. S., Polak, P., & Walker, P. S. (2024). Smoothing Out Momentum and Reversal. *Swiss Finance Institute Research Paper*, (24-47).
11. Li, T., Liu, S., Hong, E., & Xia, J. (2025). Human Resource Optimization in the Hospitality Industry Big Data Forecasting and Cross-Cultural Engagement.
12. Al Janabi, M. A. (2024). Crises to Opportunities: Derivatives Trading, Liquidity Management, and Risk Mitigation Strategies in Emerging Markets. In *Liquidity Dynamics and Risk Modeling: Navigating Trading and Investment Portfolios Frontiers with Machine Learning Algorithms* (pp. 169-256). Cham: Springer Nature Switzerland.
13. Hu, Q., Li, X., Li, Z., & Zhang, Y. (2025). Generative AI of Pinecone Vector Retrieval and Retrieval-Augmented Generation Architecture: Financial Data-Driven Intelligent Customer Recommendation System.
14. Stuart-Smith, R., Studebaker, R., Yuan, M., Houser, N., & Liao, J. (2022). Viscera/L: Speculations on an Embodied, Additive and Subtractive Manufactured Architecture. *Traits of Postdigital Neobaroque: Pre-Proceedings (PDNB)*, edited by Marjan Colletti and Laura Winterberg. Innsbruck: Universitat Innsbruck.
15. Muzaffar, Z., & Malik, I. R. (2024). Market liquidity and volatility: Does economic policy uncertainty matter? Evidence from Asian emerging economies. *Plos one*, 19(6), e0301597.
16. Yang, J., Li, Y., Harper, D., Clarke, I., & Li, J. (2025). Macro Financial Prediction of Cross Border Real Estate Returns Using XGBoost LSTM Models. *Journal of Artificial Intelligence and Information*, 2, 113-118.
17. Mähleke, N., & Lundtofte, F. S. (2025). Regime-Based Nasdaq Futures Trading.
18. Scott, D. (2024). Using Conditional Factor Performance to Analyse a Market-Beating Portfolio Strategy (Master's thesis).
19. Alfaisal, R., Abousamra, R., Mansoori, A., Tahat, K., Tahat, D. N., Habes, M., & Salloum, S. A. (2025). Analysis of Machine Learning Models for Market Action Prediction: A Case Study with K-Means Clustering and Light Gradient Boosting Machine. In *Generative AI in Creative Industries* (pp. 317-331). Cham: Springer Nature Switzerland.
20. Sarioguz, O., & Miser, E. (2024). Integrating AI in financial risk management: Evaluating the effects of machine learning algorithms on predictive accuracy and regulatory compliance. no. November.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.