European Journal of Business, Economics & Management

Vol. 1 No. 2 2025

Article Open Access

Signal Fusion and Empirical Research on Volatility Factors in High-Frequency Trading of Crypto Assets

Minghao Chi [1,*]

[1] Global Markets Trading, Barclays Capital, New York, 10010, USA
* Correspondence: Minghao Chi, Global Markets Trading, Barclays Capital, New York, 10010, USA

Received: 18 May 2025
Revised: 27 May 2025
Accepted: 23 June 2025
Published: 26 June 2025

Abstract: Due to the wide application of high-frequency trading in the crypto asset market, and because volatility factors can accurately reflect the characteristics of price changes, they have long attracted attention from both research and practice. However, a single volatility factor is often disturbed by market noise and has a small adjustment range. In this paper, multiple high-frequency volatility factors are constructed, such as historical volatility, realized volatility, and jump volatility, and three fusion techniques are designed, namely linear weighting, statistical dimension reduction, and machine learning fusion methods. Through empirical tests using the transit-by-transaction data of BTC and ETH, the results show that the comprehensive signal strategy outperforms the single-factor strategy in terms of prediction effect, stability of positive returns, and risk control, demonstrating obvious trading advantages.

Keywords: crypto assets; high-frequency trading; volatility factor; signal fusion

1. Introduction

The crypto asset market, characterized by intense volatility, continuous trading, and decentralization, has become one of the important scenarios for the application and practice of high-frequency trading strategies. However, in a high-frequency environment, traditional factor models have problems such as rapid signal attenuation and insufficient real-time performance. Although the volatility factor has significant advantages in risk measurement, a single volatility measure cannot comprehensively consider the multi-dimensional dynamic characteristics of the market. For this purpose, this paper introduces multiple volatility factor construction mechanisms and, on this basis, proposes a fusion strategy to optimize signal stability. Through empirical research, the performance of factors under different fusion modes can be evaluated to prove their practicability and robustness in actual transactions.

2. The Influence of the Characteristics of the Crypto Asset Market on Factor Modeling

The crypto asset market is characterized by high volatility, asymmetric liquidity, and rapid changes in trading depth, and it is not fully applicable to traditional financial factor models. This is because the price of the crypto market fluctuates sharply and often experiences sudden jumps and drops, requiring the factor to have a high real-time response capability; The forms of transaction orders are variable and change over time, which can have adverse effects on factors based on static statistics [1]. 24-hour continuous trading and information disparities across different platforms increase the timeliness requirements of trading factors in both time and dimension. Therefore, in order to establish a volatility factor applicable to high-frequency environments, it is necessary to have dynamic adjustment of the time window, the ability to depict jumping fluctuations, and coupling with the market microstructure.

3. Data Sources and Construction of Volatility Factors

3.1. Data Description

This research is based on the high-frequency data of the Binance exchange and selects two major cryptocurrency trading pairs, BTC/USDT and ETH/USDT, as the research objects. The data covers transaction-by-transaction and order snapshots, with a sampling frequency of 1 second and a time span from June 2024 to May 2025. It encompasses various typical market conditions and is highly representative. Transit-by-transaction data includes the transaction price, quantity and timestamp per second, which is used to generate high-frequency return sequences [2]. The order data records the quotations for buying and selling and the number of pending orders, which is used to depict the changes in market liquidity and microstructure. In the data preprocessing stage, unify the time granularity and construct equally spaced data sequences, eliminate outliers and missing points, and normalize the key variables simultaneously. To enhance the explanatory power of the factors and to better identify transaction directions, this study captures the characteristics of active transaction behaviors and order flows. Ultimately, a standardized high-frequency data system including multi-dimensional features such as price, transaction volume and depth is formed, laying a data foundation for the subsequent construction of volatility factors and signal modeling [3].

3.2. Volatility Factor Setting

This paper constructs multiple factors depicting the characteristics of market fluctuations based on high-frequency data, covering price fluctuations, return structure, and order flow dynamics. The historical volatility factor calculates the fluctuation level of recent returns through a sliding window to reflect market risks. The realized volatility measures the actual price fluctuations in the short term and is suitable for capturing instantaneous fluctuations. Jump volatility is used to identify sharp price changes and depict the impact of unexpected events on market dynamics. Furthermore, the order flow factor is introduced to reveal the changes in trading behavior through the differences in buying and selling volumes, reflecting liquidity pressure. All factors are constructed at one-second intervals and standardized to enhance the comparability among factors and the efficiency of subsequent signal fusion. These factors constitute the core basis of signal modeling in this paper (Figure 1).

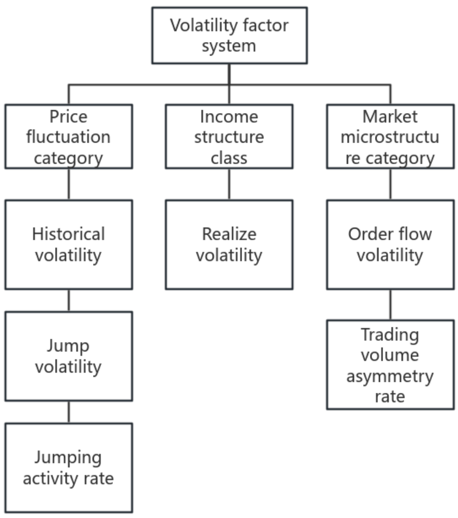

Figure 1. Framework Diagram of the Volatility Factor.

4. The Signal Fusion Method of Volatility Factor in High-Frequency Trading of Crypto Assets

4.1. Linear Weighting and Information Weighting

Linear weighting is the most fundamental method in signal fusion. It aggregates the standardized fluctuation factors using fixed weights to form a single composite signal. The most common method is equal-weight weighting, which assigns the same weight to all factors. This method is intuitive and efficient in execution. Information-based weighting is achieved by assigning differentiated weights to each factor through historical statistical indicators. Commonly used measurement indicators include the information coefficient (IC), signal direction accuracy rate, and prediction residual variance. These weights can be updated regularly to accurately reflect the historical performance of various factors, thereby enhancing the stability and forward-looking nature of the signal. The linear weighting method does not involve any parameter training, has a simple structure, is easy to understand, and is suitable for high-frequency trading scenarios with high requirements for effectiveness and controllability. Information weighting can further enhance the substantive role of factors and provide a comparable benchmark reference for the subsequent construction of more complex fusion models [4].

4.2. Statistical Dimension Reduction Model

In high-frequency trading, multi-factor models have problems such as high dimensionality, strong collinearity, and substantial redundant information. If they are not processed, it may lead to phenomena such as information distortion and model overfitting. Therefore, in this paper, the statistical dimension reduction method is adopted to compress a large number of original volatility factors into a few comprehensive signals with higher information density. Among them, the commonly used method is Principal Component Analysis (PCA). By retaining the maximum variance information, a smaller number of principal components are constructed, thereby eliminating redundant dimensions, merging highly correlated variables, and improving the stability and interpretability of the signal. Independent Component Analysis (ICA) is an effective statistical dimension reduction method, which identifies and extracts mutually independent signals from high-dimensional and multi-source factors that are not normally distributed. ICA pays more attention to statistical independence compared with PCA, and thus can identify potential structural features related to extreme fluctuations. Even in the case of factor distribution bias or the presence of long-tail features, it still has a good extraction ability. The signal after dimensionality reduction significantly reduces noise interference while retaining the core information, enhancing the robustness and generalization ability of the model.

4.3. Computer Learning Integration

Machine learning methods provide effective support for the integration and extraction of factor signals in the crypto asset market with complex high-frequency fluctuations and significant nonlinear characteristics. This paper employs two models-the ensemble learning model XGBoost and the time series model LSTM. XGBoost iteratively builds multiple regression trees to simulate the nonlinear relationship between factors and future returns. It has the ability of automatic feature selection and anti-multicollinearity, and is suitable for fusion tasks with high-dimensional inputs. Its input is a standardized set of multiple volatility factors, and the output is a signal of future short-term returns or price directions. LSTM is used to track the time series structure of factors, uses a gating mechanism to store valid historical information, and is capable of identifying trend changes as well as short-term impacts [5].

5. Empirical Research and Result Analysis

5.1. Experimental Setup

To enhance the effectiveness of the volatility factor fusion method in high-frequency trading of crypto assets, this paper constructs an experimental framework covering data processing, sample division, signal generation and strategy execution. The research object of this paper is BTC/USDT and ETH/USDT of Binance Exchange. The data frequency is 1 second, and the observation period is from June 2024 to May 2025. The samples are divided into a training set, a validation set, and a test set. After generating standardized signals, each model generates long and short positions, and the position size is determined by the signal intensity. With 10,000 USDT as the initial capital, a slippage equivalent to one second delay is assumed, and the bilateral transaction fee is 0.05%. The transaction proceeds at one-second intervals, and the net value and trading behavior are recorded in real time.

5.2. Evaluation Indicators

This paper makes judgments from four dimensions: return capacity, risk control, stability and signal prediction quality respectively, and verifies different volatility factor fusion strategies. All indicators are calculated based on the net asset value time series generated by strategy backtesting and signal outputs as follows:

The Annualized Return measures the overall profitability of the strategy and is defined as:

$$R_{annual} = \left(\frac{V_T}{V_0}\right)^{\frac{252\times24\times60\times60}{T}} - 1 \tag{1}$$

In the actual measurement, the XGBoost fusion strategy achieved an annualized rate of return $R_{annual} \approx 63.15\%$ that was superior to that of PCA fusion at 45.97% and linearly weighted at 38.24%.

Maximum Drawdown measures the most severe loss that a strategy may face in history and is defined as:

$$MaxDrawdown = \max_{t\in[0,T]}\left(\frac{\max_{s\le t} V_s - V_t}{\max_{s\le t} V_s}\right) \tag{2}$$

The maximum drawdown of the XGBoost model was 8.2%, significantly smaller than 12.9% of the linear weighted strategy, indicating that it has stronger risk resilience.

In order to measure the risk-adjusted return performance of the strategy, this paper calculates the Sharpe Ratio and the Calmar Ratio. The Sharpe ratio is defined as:

$$Sharpe = \frac{E[R_t - R_f]}{\sigma(R_t)} \tag{3}$$

Among them, R_t is the daily rate of return of the strategy, R_f is the risk-free interest rate (set as 0 in this paper), and $\sigma(R_t)$ is the standard deviation of the rate of return. The Calmar ratio is the ratio of annualized return to maximum drawdown:

$$Calmar = \frac{R_{annual}}{MaxDrawdown} \tag{4}$$

The Sharpe ratio of the XGBoost strategy is approximately 2.41, and the Calmar ratio reaches 7.70, both of which are higher than those of PCA fusion (1.75, 4.38) and linear weighting (1.43, 2.96), indicating that it has a better risk-return profile.

In terms of signal prediction, this paper introduces the information ratio to measure the correlation degree between the signal and future returns, which is defined as:

$$IR = \frac{E[r_t \cdot s_t]}{\sigma(r_t \cdot s_t)} \tag{5}$$

The higher the information ratio is, the more consistent the signal's direction is with the actual market returns. In the empirical study, the IR of the XGBoost model is 0.79, which is higher than that of PCA (0.58) and linear weighting (0.42).

In addition, the Accuracy of signal direction prediction is also calculated, namely:

$$Accuracy = \frac{1}{T}\sum_{t=1}^{T} I\big(sign(s_t) = sign(r_t)\big) \quad (6)$$

The direction prediction accuracy rate of the XGBoost model reaches 60.5%, significantly higher than that of other models, further proving that its fused signal has good predictive power and tradability.

5.3. Comparative Analysis of Signal Prediction Capabilities

This paper mainly examines the predictive ability of different volatility factors and their fusion signals for future market returns. Specifically, historical volatility (HV), realized volatility (RV), and jump volatility (JV) are evaluated as single factors respectively; Meanwhile, the statistical fusion signal (PCA) obtained through Principal Component analysis (PCA) and the machine learning fusion signal constructed based on XGBoost are taken as the test pairs. The direction of the payoff in the next five seconds is used as the criterion for judging the validity of the signal, and then the Pearson correlation coefficient between the information coefficient and the future payoff is calculated. The stability of the signal is measured by the standard deviation (Table 1).

Table 1. Comparison of the Predictive Capabilities of Different Volatility Signals and Fusion Models under the BTC/USDT Trading Pair.

Signal type	Direction accuracy rate (%)	information coefficient (IC)	IC standard deviation
Historical volatility HV	51.2	0.021	0.035
Achieve volatility RV	52.9	0.034	0.028
Jump Volatility JV	53.1	0.037	0.030
Statistical Fusion (PCA)	56.8	0.058	0.021
ML Fusion (XGBoost)	60.5	0.079	0.018

Studies show that the predictive ability of a single volatility factor is relatively weak, the direction accuracy rate is basically equivalent to the random guessing level (i.e., around 50%), the information coefficient fluctuates greatly, and it is easily affected by the microstructure of the market. Compared with historical volatility, realized and jump volatility factors, which incorporate high-frequency structural information, show slightly improved predictive performance, indicating that the introduction of high-frequency information has certain value. The statistical fusion method (PCA) can be utilized to further enhance the predictive performance. The accuracy rate is increased to 56.8%, the information coefficient is significantly improved, and the signal performance is more stable. The XGBoost fusion signal achieved the best performance, with an accuracy rate of over 60% and an information coefficient of 0.079, indicating that it has obvious advantages in identifying complex nonlinear relationships among factors.

5.4. Comparative Analysis of Fusion Strategies

After completing the evaluation of the signal prediction ability in the previous section, this paper further constructs multiple high-frequency trading strategies based on the fused signals and builds a backtesting system that closely simulates real-world trading conditions to measure the performance of the strategies in terms of overall returns, risk control and stability. Specific strategies include linear weighting strategies (such as equal-weight and information-weighted), statistical dimensionality reduction strategies (such as PCA), and machine learning strategies (such as XGBoost). All strategies generate trading signals at a frequency of seconds and allocate funds proportionally based on the signal strength. The backtest settings include a bilateral transaction fee rate of 0.05%, a simulated slippage equivalent to a 1-second execution delay, a trading subject of BTC/USDT, and a test period from March 2025 to May 2025 (Table 2).

Table 2. Comparison of Backtest Performance of Different Fusion Strategies under the BTC/USDT Trading Pair.

Strategy type	Annualized rate of return (%)	Maximum drawdown (%)	Sharpe ratio	Karma ratio	Average daily turnover rate (%)
Linear weighting (equal-weight)	38.2	12.9	1.43	2.96	28.5
Information weighting	45.7	11.3	1.74	4.05	26.1
PCA fusion	52.6	10.5	1.98	5.01	23.7
XGBoost fusion	63.1	8.2	2.41	7.70	21.9

Studies show that the nonlinear fusion strategy is generally superior to the traditional linear weighting method in terms of return and risk control. The XGBoost strategy performed the best, with an annualized return rate of 63.1% and a maximum drawdown of only 8.2%. It ranked first in indicators such as the Sharpe and Calmar ratios, demonstrating excellent stability and predictive power. The PCA strategy mitigates factor redundancy by applying dimensionality reduction techniques. The Sharpe ratio is close to 2, and the maximum drawdown is controlled within 11%, demonstrating strong robustness. In contrast, the linear weighting method is vulnerable to market noise interference and exhibits unstable performance. In addition, the XGBoost strategy has a lower average daily turnover rate and higher execution efficiency.

5.5. Sensitivity and Stability Tests

To further verify the effectiveness and stability of the volatility factor fusion method, this paper conducts scalability tests from two dimensions. On the one hand, it examines the performance of the strategy under various market conditions, including normal volatility, trending markets, and high-turbulence periods. On the other hand, an analysis is conducted around the impact of changes in key parameters on the strategy results, such as changes in settings such as slippage rate, transaction fees, and prediction window length. This section mainly compares the performance of the XGBoost and PCA fusion strategies, under the above conditions, and evaluates them in terms of revenue performance, robustness, and execution efficiency (Table 3).

Table 3. Performance of Strategy Robustness under Different Market Conditions and Parameter Settings.

Test scenario	Annualized return rate of XGBoost Fusion (%)	Maximum drawdown (%)	Annualized rate of return of PCA fusion (%)	Maximum drawdown (%)
Regular Volatile Market (March 2025)	58.6	8.9	46.1	10.7
Obvious unilateral increase (Early April 2021)	65.3	7.4	51.2	9.5
Sudden sharp market decline (Late April 2021)	60.1	9.6	43.7	12.2
The slippage has been increased to 0.15%	55.4	10.2	40.6	13.1
The handling fee is doubled (0.1%)	50.7	9.8	39.3	12.5
The prediction window has been changed from 5 seconds to 10 seconds	66.8	8.5	48.4	10.2

The experimental results show that the XGBoost fusion strategy demonstrates the advantages of stable returns and low fluctuations in various market environments, indicating strong adaptability and robustness. Even under sharp market fluctuations or adverse trends, the XGBoost strategy maintains an annualized return above 60%, reflecting its strong resilience, reflecting its strong ability to withstand pressure. In contrast, the overall performance of the PCA fusion method declines when facing sudden price changes or rising transaction costs, further highlighting the advantages of machine learning methods in capturing the structural characteristics of the market. Although both strategies are affected by external disturbances, XGBoost is less impacted, indicating that it is more stable in the model construction and execution stages. When the prediction window was extended from 5 to 10 seconds, the XGBoost strategy achieved higher returns, showing strong compatibility with lagging features, while the PCA strategy exhibited reduced adaptability under this condition.

6. Conclusion

This study focuses on high-frequency trading of crypto assets, constructs a multi-factor system incorporating historical, realized, jump, and order book-based volatility measures, and proposes three types of signal fusion methods: linear weighting, statistical dimensionality reduction and machine learning. Empirical results show that the fusion strategy outperforms the single-factor model in terms of revenue, drawdown control and signal stability, with XGBoost performing the best. Further sensitivity tests verified the adaptability and robustness of the model. The findings demonstrate that integrating multi-source volatility information significantly enhances predictive accuracy and execution efficiency, providing methodological support for high-frequency quantitative trading strategies.

References

1. M. M. Karim, L. Zhao, H. Nguyen, C. Li, A. Cornelius, and S. Patra et al., "Exploring asymmetries in cryptocurrency intraday returns and implied volatility: New evidence for high-frequency traders," *Int. Rev. Financ. Anal.*, vol. 96, p. 103617, 2024, doi: 10.1016/j.irfa.2024.103617.
2. L. Zhao, V. H. Nguyen, and C. Li, "The volatility-liquidity dynamics of single-stock ETFs," *Finance Res. Lett.*, vol. 69, p. 106163, 2024, doi: 10.1016/j.frl.2024.106163.
3. A. Cornelius, A. Ogbonna, M. Yusuf, E. Adebayo, J. Eze, and I. Lawal et al., "Trading volume, intraday and overnight volatility of stock returns in the Nigerian Stock Exchange," *Asian J. Probab. Stat.*, vol. 26, no. 9, pp. 110–122, 2024, doi: 10.9734/ajpas/2024/v26i9649.
4. S. Patra, "An empirical analysis of the volume-volatility nexus in crude oil markets under structural breaks: Implications for forecasting," *Int. Rev. Econ. Finance*, vol. 94, p. 103434, 2024, doi: 10.1016/j.iref.2024.103434.
5. T. Chen, J. Deng, and J. Nie, "Implied volatility slopes and jumps in bitcoin options market," *Oper. Res. Lett.*, vol. 55, p. 107135, 2024, doi: 10.1016/j.orl.2024.107135.