Dean&Francis

ISSN 2959-6130

An Empirical Study of Machine Learning for Filtering Momentum Signals: Daily Frequency Strategy Analysis Based on the A-Share Market

Yanyouyou Wang*

Faculty of Arts & Science, University of Toronto, M5S 1A1,Toronto, Canada

*Corresponding author: yanyouyou.wang@mail.utoronto.ca

Abstract:

Momentum strategies are commonly used in quantitative investing, but they are susceptible to noise signals, false breakouts, and market reversals in the A-share market, leading to significant performance volatility. This paper selects 80 A-shares from January 2020 to May 2025 as samples to investigate whether machine learning filtering can improve the effectiveness of momentum signals. The study generates buy signals based on 20-day cumulative returns and rolling quantile thresholds, and constructs features such as momentum, volatility, trading volume, price position, and trend. A gradient boosting classifier is used to predict the return direction the day after the signal is issued, and the effectiveness of the strategy before and after filtering is compared through single-asset backtesting considering stop-loss and transaction costs. The results show that after machine learning filtering, the Sharpe ratio and win rate of most stocks are improved, and the improvement in Sharpe ratio is statistically significant. However, there is no stable positive correlation between AUC and strategy improvement, indicating that machine learning is more suitable as a signal evaluation tool.

Keywords: Momentum strategy, machine learning filtering, gradient boosting tree, A-share market

1. Introduction

Momentum strategies constitute one of the canonical research paradigms in quantitative asset pricing research, whose core theoretical logic centers on capturing cross-sectional and time-series persistent price trends of financial assets. Jegadeesh & Titman first established systematic empirical evidence for the existence of momentum effects in stock markets [1], while Moskowitz et al. extended the time-series momentum analytical framework to cross-border global capital markets. Distinct from overseas mature capital markets [2], the A-share market features a retail investor-dominated participant structure and

ISSN 2959-6130

asynchronous information transmission efficiency, which enables momentum strategies to generate measurable excess returns under specific market regimes. Existing empirical studies targeting China's equity market document that liquidity constraints, transaction cost friction and retail investor participation ratios exert substantial explanatory power over the return prediction accuracy of machine learning forecasting models [3]. Despite the above theoretical advantages, traditional momentum trading frameworks suffer from structural defects including unstable signal reliability, frequent false breakout signals and endogenous market regime switches, which lead to severe return volatility in practical trading deployment.

The continuous iteration of machine learning algorithms in recent years provides novel technical approaches to optimize the discrimination efficiency of quantitative trading signals. Gu et al. systematically benchmarked the out-of-sample forecasting performance of mainstream machine learning models within asset pricing regression frameworks [4], and verified that nonlinear tree-based models such as gradient boosting trees possess comparative advantages in identifying complex interactive relationships among predictive characteristic variables [5]. Subsequent comprehensive literature reviews further categorize tree ensemble models, regularization estimation methods and neural network architectures as three mainstream methodological branches of machine learning asset pricing research. Drawing on the above research conclusions, multiple scholars introduce machine learning classification models as signal screening tools, aiming to filter invalid noise signals generated by traditional factor strategies while retaining profitable trading opportunities. Nevertheless, extant relevant literature predominantly focuses on elevating the in-sample predictive accuracy of machine learning models and lacks systematic empirical examinations on whether signal-filtered strategies can generate statistically significant improvements in risk-adjusted returns under real transaction backtest constraints.

This paper selects 80 randomly sampled representative A-share listed firms with complete daily trading records from 2020 to 2025, and conducts empirical tests on whether gradient boosting tree screening frameworks can statistically improve the risk-return performance of baseline momentum strategies. Multi-dimensional technical characteristic variables are constructed based on daily trading panel data to predict the forward return direction of primitive momentum signals. A single-stock backtesting model embedded with fixed stop-loss rules is adopted to quantify performance discrepancies between unfiltered and filtered trading strategies. Comparative analysis covering Sharpe ratio, transaction win rate and profit-loss ratio, alongside a paired sample t-test, is deployed to resolve two core research questions: First, whether machine learning signal filtering generates statistically significant increments in risk-adjusted investment returns; Second, whether sustained positive correlation exists between model out-of-sample predictive capacity (AUC) and actual backtest performance. The empirical conclusions of this research deliver a quantitative reference basis for defining the functional positioning of machine learning algorithms in practical quantitative investment operations.

2. Method

2.1 Data Acquisition and Preprocessing

This study randomly selected 80 A-share stocks (79 of which were valid), covering multiple industries including energy, finance, consumer goods, pharmaceuticals, manufacturing, new energy, real estate, transportation, and media. Energy Resources Sector: China Shenhua Energy (601088), Sinopec (600028), and Zijin Mining (601899). Consumer Goods Sector: Wuliangye (000858), Yili Group (600887), and Shuanghui Development (000895). Financial Sector: China Merchants Bank (600036) and Industrial and Commercial Bank of China (601398). New Energy and Manufacturing Sector: CATL (300750) and BYD (002594). The samples generally come from several different industries for these stocks. The sample period will be from January 2020 to May 2025, and over five years of daily trading data will be available.

Data will be obtained from the Baostock securities data interface, and the following items are available daily: open price, high price, low price, close price, trading volume and turnover. All the price data have been normalised for a stock split to ensure comparability. Logarithmic returns are used for the pre-processing stage, and outliers exceeding three standard deviations are removed to prevent abnormal trading data from affecting the results of strategy backtesting. Sample data for each stock meets the backtesting requirements, and enough daily-frequency observations need to be saved. A minimum of 50 valid original signal samples are needed for the model training stage.

2.2 Traditional Momentum Signal Generation

The method of the above study is a momentum approach. The first idea of the momentum approach is that stocks that have risen in recent years are likely to keep increasing. Cumulatively, the returns of all stocks over the past 20 trading days have been calculated.

$$past\_return_t = \frac{Close_t}{Close_{t-2}} - 1. \tag{1}$$

Where where past_returnt denotes the 20-day cumulative return ending on trading day t, Closet and Closet-2 are the adjusted closing prices on trading days t and t−2, respectively, and t represents the trading day index.

The first purchase signal is generated when the cumulative return over the past 20 trading days exceeds the 80th percentile threshold of the stock's historical return distribution in the past 252 trading days. The above parameter settings are based on the typical values in the literature for classic momentum strategies and aim to track medium-term price changes [1,2]. To prevent extended bias, no trading signals will be generated for the first 20 trading days.

2.3 Feature Engineering and Tag Construction

To build a machine-learning filter model in this paper, several features that can help predict the reliability of the original signal were selected. The above feature sets are momentum features, volatility features, volume features, price position features and trend features. Five-, ten-, and twenty-day momentum indicators are calculated by summing the price changes over the corresponding periods to show the trend at different times. Volatility features are used to describe market fluctuations and changes by calculating the rolling standard deviation of 5-day and 20-day returns and their ratio. Volume features are abnormal changes in trading activity, which can be expressed as the ratio of the current day's volume to the average volume over the past 5 days. Price position features determine whether the price is in an overbought or oversold area by calculating how far the current price is from the 20-day low relative to the range of the 20-day high and low (i.e., (current price - 20-day low) / (20-day high - 20-day low)). Trend features are binary indicators of whether the price has been rising over the past five days, and they can be used as proxies for the direction of a short-term trend. The other parts of the model are 60-day momentum, an adaptive threshold, turnover rate and the ratio of 5-day/20-day moving averages.

Label is a binary variable indicating whether the return on the next trading day will be positive (label = 1 if it will increase in the future, 0 otherwise). Only when the original signal day is valid (non-zero) will a label be added; otherwise, the model will not learn a prediction.

2.4 Machine Learning Model Training

The first is a gradient boosting classifier for signal filtering. Gradient Boosting Trees are suitable for addressing non-linear relationships and interactions among features automatically, as well as other problems [5, 6]. Second, it is also prone to overfitting. Model training used a 7:3 time series split, and the first 70% of the data were used as the training set, while the latter 30% served as the test set. Instead of using future information for feature standardization, the mean and standard deviation were calculated on the training set and then applied to the test set.

Time-series cross-validation was employed, with a maximum of 3-fold cross-validation, and ROC-AUC was used as the optimization objective. The search space included: n_estimators □ {80, 120}, max_depth □ {2, 3}, and learning_rate = [0.05]. The optimal model's AUC was calculated on the test set as an evaluation metric for its predictive ability.

2.5 Signal Filtering and Backtesting Framework

After training, the optimal model is used to predict the future rise of all original signals, obtaining the probability of each signal increasing daily. Only signals with a predicted probability of at least 0.6 (signal_filtered = 1) are retained, and the rest are discarded. This threshold choice strikes a balance between signal quality and signal quantity.

This backtesting framework uses a single-asset, tick-by-tick trading simulation with a stop-loss mechanism. The initial capital is 100,000 RMB. Specifically, when a buy signal is triggered, the strategy uses all available funds to buy at the closing price of the day. If a position is already held, a new buy signal will be used as one of the conditions for closing the position, rather than opening a new position on the same day. During the holding period, if the price falls more than 5% from the buy point, the stop-loss mechanism is triggered, and the position is sold at the closing price of the day. If the stop-loss is not triggered, the position is held until the next signal day or the end of the backtesting period, after which it is closed. Regarding transaction costs, the model considers both slippage and commissions. Slippage is 0.02%, commission is 0.01%, and the total cost of a one-sided transaction is 0.03%. Finally, a net asset value curve is calculated based on daily holdings and returns, with the corresponding transaction costs deducted during the calculation.

3. Results and Discussion

3.1 Analysis of Model Predictive Ability

The results of the backtest show that the gradient boosting tree model has limited performance in predicting the rise and fall of the next day. The AUC values of the test set for each stock range from 0.3479 to 0.6390. This result shows that the ability to predict short-term returns based solely on traditional technical indicators such as price, volume and volatility is limited, which is consistent with the expectation of the efficient market hypothesis that there are a lot of unpredictable noise components in daily price fluctuations. Even in studies that show relative advantages in stock prediction tasks such as gradient boosting, the prediction target is usually the direction of medium- to long-term returns [7], and short-term (1-day) prediction is much more difficult.

ISSN 2959-6130

It is worth noting that there is no clear positive correlation between AUC value and subsequent strategy performance. As a classifier performance metric, AUC measures the model's ability to distinguish between positive and negative samples. However, the actual performance of a trading strategy depends not only on prediction accuracy but also on factors such as the degree of probability calibration, signal threshold settings, and trading costs. A model with strong ranking ability but a large probability calibration bias may produce suboptimal trading decisions at certain thresholds.

3.2 Strategy Performance Comparison

Based on the average performance of these 80 randomly selected stocks, the original momentum strategy had a mean Sharpe ratio of -0.0753 and a median of -0.0445. After ML filtering, the mean Sharpe ratio increased to 0.1186, the median increased to 0.0391, and the average improvement was 0.1939.

In terms of improvement, the Sharpe ratio improved for 52 out of 79 stocks, accounting for 65.8%. The most significant improvement was seen in China State Construction (601668), with a ΔSharpe of 1.3587 (see Figure 1). Conversely, the most significant deterioration was observed in China Merchants Bank (600036), with a ΔSharpe of -1.2242.

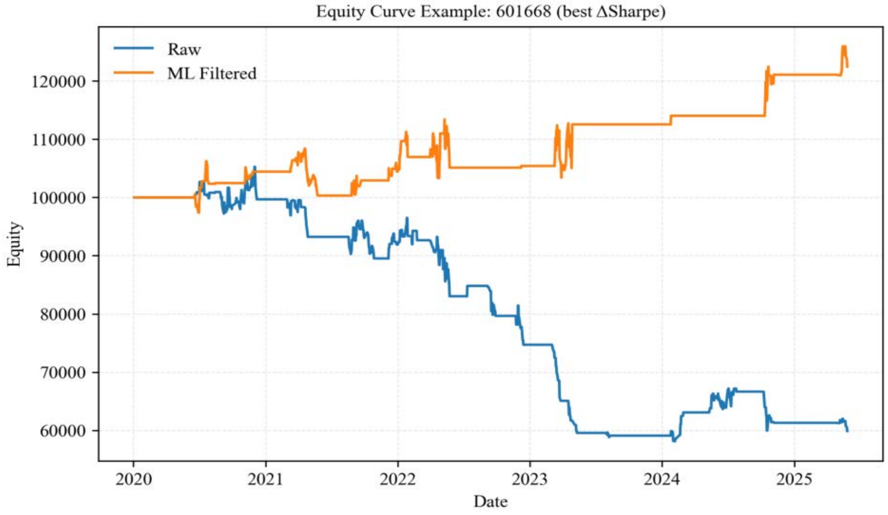

Fig. 1. Equity Curve Example 601668 (Picture credit: Original)

In terms of win rate, the average win rate of the original strategy was 49.30%, which increased to 61.57% after ML filtering. If the win rate increases but the Sharpe ratio improves only slightly, it usually means that the filtering model has eliminated some low-quality signals, but at the same time it may also sacrifice some high-yield trading opportunities. This is consistent with the finding that the returns of momentum stocks may have a bimodal distribution, which leads to the asymmetry of risk of traditional momentum signals [8]. If the win rate and Sharpe ratio decrease at the same time, it means that the probability threshold may be too aggressive, resulting in too few samples retained or signal selection bias.

The results of the paired t-test showed that, for the current sample, the Sharpe improvement brought about by ML filtering was statistically significant ($t = -3.8831$, $p = 0.0002$).

3.3 Signal Filtering Intensity Analysis

This experiment uses rolling quantiles as the adaptive momentum threshold. The in-sample average adaptive threshold is 0.0912, and the median threshold is 0.0769 (see Figure 2). This setting allows stocks with different volatility, industries, and market capitalization/liquidity to use signal triggering criteria that are more reasonable relative to their historical state. In terms of signal filtering strength, the average signal retention rate after ML filtering is 32.13%.

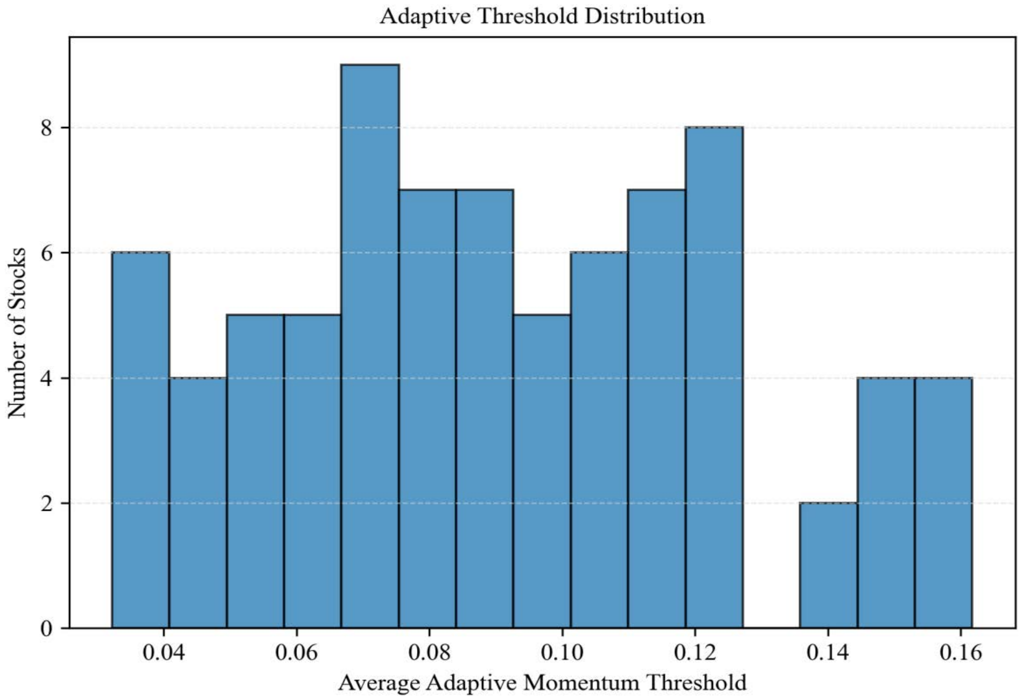

Fig. 2. Adaptive Threshold Distribution (Picture credit: Original)

3.4 Limitations Analysis

Limitations of data frequency. Whether the conclusions of the daily frequency research can be extended to high-frequency environments still needs further verification. In fact, existing literature points out that noise and signal are highly intertwined in high-frequency data, and the difficulty of factor extraction and signal filtering is much higher than that of daily frequency data [9, 10].

4. Conclusion

This paper takes 80 representative A-share listed enterprises as research samples to conduct empirical tests on the risk-return improvement effect of machine learning signal filtering within momentum trading strategies. Empirical backtest results show that gradient boosting tree screening frameworks can effectively discriminate primitive momentum signals and have increased the Sharpe ratio for 65.8% of the sampled stocks, with an average improvement in the magnitude of approximately 0.1939.

Based on the empirical results, the application of quantitative investment operations needs to be carried out in a particular way; although machine learning models can be used as a supporting discrimination tool to evaluate the reliability of quantitative trading signals, their incremental impact on strategy profitability in isolation is likely to be overestimated in practice. The next two directions of research are to add high-frequency intraday trading data at the minute level to verify the reliability of the previous study's results in a high-frequency trading environment, and to examine how well sequence feature extraction can be performed by deep learning models, such as LSTM. Based on previous studies of asset pricing, deep neural networks have been employed to extract high-dimensional conditional information from price series and enhance the out-of-sample forecasting stability of quantitative strategies.

Future research can further expand this study in several ways. First, the research scope can be expanded from daily data to minute-level or even higher frequency trading data to verify the applicability of machine learning signal filtering methods in high-frequency trading environments and further analyze the impact of market microstructure noise on model performance. Second, deep learning models such as LSTM and Transformer can be introduced to fully explore the temporal dependencies in price series and combine them with attention mechanisms to improve signal recognition capabilities. Furthermore, future research can integrate multi-source information such as macroeconomic indicators and industry factors to construct a multimodal prediction framework, thereby further enhancing the machine learning model's ability to identify the effectiveness of momentum signals and providing more reliable technical support for intelligent quantitative investment strategies.

This study still has certain limitations. First, the sample only includes 79 A-share listed companies with complete trading data from 2020 to 2025, resulting in a relatively

ISSN 2959-6130

limited sample size and time span. The generalizability of the research conclusions in other market environments and longer time periods needs further verification. Second, this paper only uses Gradient Boosting as the machine learning classification model and has not yet conducted a systematic comparison with Random Forest, XGBoost, LightGBM, and deep learning models. Therefore, the performance differences between different algorithms still warrant further investigation. Furthermore, this study mainly constructs features based on technical indicators and does not fully consider external factors such as macroeconomic conditions, company fundamentals, and market sentiment. Their incremental contribution to the model's predictive ability still needs further investigation in subsequent studies. Finally, although the backtesting considered transaction costs and stop-loss mechanisms, it has not fully simulated the liquidity constraints, market impact costs, and slippage changes in actual trading. Therefore, the performance of the strategy in the real market still needs to be verified through experiments that more closely resemble real-world trading environments.

References

[1] Jegadeesh N, & Titman S. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance, 1993, 48(1): 65-91.

[2] Moskowitz T J, Ooi Y H, & Pedersen L H. Time series momentum. Journal of Financial Economics, 2012, 104(2): 228-250.

[3] Leippold M, Wang Q, & Zhou W. Machine learning in the Chinese stock market. Journal of Financial Economics, 2022, 145(2): 64-82.

[4] Gu S, Kelly B, & Xiu D. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 2020, 33(5): 2223-2273.

[5] Bagnara M. Asset Pricing and Machine Learning: A critical review. Journal of Economic Surveys, 2024, 38(1): 27-56.

[6] Azevedo V, & Hoegner P. Enhancing stock market anomalies with machine learning. Review of Quantitative Finance and Accounting*, 2023. (SSRN Working Paper No. 3752741)

[7] Padhi D K, Padhy N, Bhoi A K, Shafi J, & Ijaz M F. A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. Mathematics, 2021, 9(21): 2646.

[8] Han C. Bimodal Characteristic Returns and Predictability Enhancement via Machine Learning. Management Science, 2022, 68(10): 7701-7741.

[9] Liu Y, Xu C, Hou M, Liu W, Bian J, Liu Q, & Liu T. Digger-Guider: High-Frequency Factor Extraction for Stock Trend Prediction. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(12).

[10] Konstantinou C, Ibikunle G, Rzayev K, & Cao Y. Cutting Through the Noise: A Deep Learning-Enhanced Particle Filter for High-Frequency Market Forecasting. Working paper, University of Edinburgh, 2024.