Analysis of Fundamental Multi-factor Stock Selection Strategy

Jinyuan Liang

Chenghan International Curriculum Experimental Campus, Shenzhen, China

L2995885089@outlook.com

Abstract. This paper focuses on the fundamental multi-factor stock selection strategy, systematically sorts out its core factor system (covering traditional financial factors and emerging non-financial factors), the evolution process of the model from linear stage to machine learning application, and deeply analyzes the challenges such as factor effectiveness fluctuations and implementation obstacles faced in practice, and discusses the future development directions in the application of alternative data, improvement of model interpretability and construction of dynamic adjustment mechanisms. The article deeply traces the model evolution path of this strategy, which has gradually transitioned from the early simple weighting models relying on linear regression to the complex model stage incorporating machine learning algorithms such as random forests and gradient boosting decision trees. It clearly demonstrates the role of technological advancement in driving the improvement of strategy accuracy. The research aims to provide theoretical and practical reference for investors to optimize their investment portfolios and improve returns. It emphasizes that building an effective strategy requires balancing predictive capabilities and robustness, and it is necessary to promote the development of strategies by strengthening non-financial factor research, developing model interpretation tools, and exploring cross-market adaptation mechanisms.

Keywords: Fundamental multi-factor, stock selection strategy, factor system, model evolution, quantitative investment.

1. Introduction

In the field of quantitative investment, the effectiveness of stock selection strategies is directly related to the level of investment returns [1]. As an important method, fundamental multi-factor stock selection, by integrating company financial and non-financial information, has become a key tool to connect financial analysis and investment practices [2].

In early research, the three-factor and five-factor models proposed by Fama and French laid the foundation and built a classic financial factor system from the dimensions of valuation, profit, etc., explaining some of the laws of stock returns [3, 4]. However, with the complexity of the market environment, the limitations of traditional linear models gradually emerge, such as the difficulty in dealing with the nonlinear relationship between factors [5].

To this end, the academic community has begun to explore new directions: on the one hand, expand the scope of factors and incorporate non-financial factors such as ESG into the system to supplement long-term value judgments; on the other hand, machine learning models such as random forests are introduced to mine hidden correlations between factors, but its “black box” attribute has also caused controversy about interpretability [6, 7].

At the same time, market differences bring new challenges. Comparing US stocks and A-shares, it was found that there were obvious differences in factor performance, the backtest deviation problem was prominent, and the importance of off-sample testing became increasingly prominent [8, 9]. How to deal with factor time-variability and data lag has become the key to improving strategic resilience [10].

Against the current background of the continuous development of quantitative investment, this article conducts in-depth research on fundamental multi-factor stock selection strategies and analyzes its core factor system, model evolution process, practical challenges and future development directions, which is of great practical significance for investors to improve investment returns and optimize investment portfolios.

2. Core Factor System

2.1. Traditional Financial Factors

In the valuation factor, the price-to-earnings ratio (PE) reflects the market’s expectations for the company’s future growth. Lower PE may imply that the stock is undervalued, but this judgment needs to be combined with the company’s growth prospects. For example, companies with stable, mature but slow growth may be in a low PE state for a long time. Price-to-book ratio (PB) is particularly important for heavy asset companies. It measures the relationship between stock price and net assets per share and is an important valuation reference for industries such as steel and real estate [6].

In terms of profit factor, the return on equity (ROE) measures the return on shareholders’ equity and reflects the efficiency of using own capital. A higher ROE usually means that the company has stronger profitability, and high-quality companies will maintain a higher ROE for a long time [5]. Gross profit margin reflects the company’s profit margin after deducting direct costs and reflects the company’s core competitiveness. A high gross profit margin shows that the company has advantages in cost control, product differentiation, etc. [5].

Among the growth factors, the revenue growth rate is the leading indicator of the company’s business expansion speed. The continuous high growth indicates that the company’s market share is expanding. For example, emerging technology companies gained attention in the early stages due to their rapid revenue growth [7]. The net profit growth rate directly reflects the growth of the company’s operating results, reflects the speed of profitability improvement, and is a key indicator for judging the company’s growth potential [7].

In the assessment of financial health, the debt-to-asset ratio evaluates the company’s long-term debt repayment risk. An excessive debt-to-asset ratio may mean that the company faces greater debt pressure. For example, over-expanding real estate companies are prone to financial difficulties when market fluctuations occur [8]. The current ratio reflects the company’s short-term debt repayment ability, and it is generally believed that this ratio is more reliable when it is greater than 2 [8].

2.2. Emerging Non-financial Factors

Environmental Social and Governance (ESG) factors include multiple dimensions, where the carbon footprint in the environmental dimension reflects the total carbon emissions of the company, and low-emission companies are more in line with the concept of sustainable development; employee welfare in the social dimension helps the company attract talents and enhances employee loyalty; board efficiency in the governance dimension is related to the scientific nature of company decisions and can better supervise management [9].

Factors in corporate governance have an important impact on the development of the company. The proportion of independent directors affects the independence of the company’s decision-making and can prevent the company from being controlled by insiders; a reasonable equity incentive design can bind the interests of management and shareholders, and encourage management to improve the company’s performance [10].

The assessment of Industry prosperity is also crucial. People need to evaluate the industry's prosperity in combination with the macroeconomic cycle. During the economic expansion period, the prosperity of cyclical industries such as automobiles and steel is relatively high; during the economic recession period, the demand for consumer necessities industries such as food and beverages is relatively stable. Investors need to pay attention to macro indicators, policy orientation and industry supply and demand to grasp investment opportunities [11].

3. Model Evolution

3.1. Linear Model Stage

The Fama-French three-factor model better explains the difference in stock returns through market factors, market capitalization factor (SMB) and book market capitalization ratio factor (HML). Based on the effective market hypothesis, it believes that the stock price reflects all public information and can capture the return changes of stocks with different risk characteristics. For example, market factors play a major role in a bull market, while small market value and high book value ratio companies interpret returns by SMB and HML factors. But as the market becomes more complex, the model is difficult to capture nonlinear features and lacks interpretation capabilities when extreme volatility or emerging industries rise [12].

3.2. Machine Learning Applications

Random forests can process high-dimensional data, which learns by building multiple decision trees in an ensemble, reducing variance and improving prediction stability [13]. This model can screen and analyze a large number of financial and non-financial factors and explore nonlinear relationships, such as predicting stock returns based on a variety of fundamental indicators, and is highly robust to noise and outliers.

Neural networks have powerful nonlinear mapping capabilities and can identify complex profit patterns [14]. It extracts factor features through multi-layer neural structures and learns the complex relationship between benefits. For example, when processing multi-factor data, it mines hidden laws to improve prediction accuracy, but there is a "black box" problem with long training time and poor interpretability.

As an optimized gradient enhancement library, XGBoost can automatically optimize factor combinations [15]. It learns dynamically to adjust factor weights through historical data, such as reallocating weights to adapt to fluctuations when market changes. But like other complex models, there is a risk of overfitting, and when training data is insufficient or complexity is too high, it may perform poorly in the test set or in the actual market [16].

Cross-validation, regularization and other methods are usually used to optimize the model to ensure prediction and generalization capabilities [16].

4. Practical Challenges

4.1. Factor Validity

The factors have cyclical fluctuations, and the growth factor and momentum factor perform well in bull markets, and investors pay more attention to growth potential and stock price trends; in bear markets, investors have lower risk preferences, value factor and quality factor have more advantages, and low-valuation and high-profit quality stocks are relatively resistant to declines, such as value stocks falling less during the 2008 financial crisis [17].

Cross-market differences also affect the effectiveness of factors. US stocks are mainly institutional investors, the market is mature and rational, and the factor performance is stable; A-share retail investors account for a high proportion and have large emotions, and the effectiveness of factors is easily affected by short-term sentiment, such as momentum factors may fluctuate abnormally [18].

In addition, there is a crowding effect, and excessive investors using the same strategy will lead to concentrated stock demand, pushing up prices and attenuating the factor premium [19]. For example, concentrated buying caused by popular factors will overdraft the income in advance, and the market may fall sharply when adjusted.

4.2. Obstacles in Implementation

Data lag is a major obstacle to implementation. Financial data is released quarterly, resulting in a lag in strategy. The company's situation may have changed before the data is released, affecting the timeliness of the strategy. For example, performance declines in the quarter, but investors may still make decisions based on old data [20].

Transaction costs cannot be ignored. Frequent position adjustments will incur costs such as commissions and stamp duty. According to research, it can erode more than 5% of the returns. Investors need to weigh the returns and costs of position adjustments [21].

Liquidity constraints also affect the implementation of the strategy. Large funds find it difficult to buy and sell small market capitalization stocks or trade in inactive stocks without affecting the stock price. Forced trading may lead to price fluctuations, cost increases, and even the expected combination allocation cannot be achieved [22, 23].

5. Conclusion

Fundamental multi-factor stock selection strategies occupy an important position in quantitative investment, and building an effective strategy requires balancing prediction capabilities and robustness. The future development directions mainly include: the application of alternative data, such as monitoring factory operating rates and warehouse inventory through satellite images to understand the company's operating conditions in advance, and extracting market sentiment indicators with the help of network public opinion and natural language processing to grasp the trend and adjust the strategy; improving the interpretability of the model, using Shap value to quantify the contribution of factors, helping investors understand the role of each factor in stock selection, and improving the transparency and trust of the model ; establishing a dynamic adjustment mechanism, optimizing the factor weight based on macroeconomic indicators, increasing the weight of growth factor during the economic expansion period, increasing the weight of value and quality factor during the recession period, so that the strategy can better adapt to cyclical changes, and improving the stability and returns of the combination At the same time, quantitative research on non-financial factors such as ESG should be strengthened to improve the factor system , machine learning model interpretation tools should be developed to lower the threshold for use , cross-market adaptation mechanisms should be explored to improve global adaptability , and through these efforts, strategies should be continuously developed and higher value for investors.

References

[1] Fama E F, French K R. Common risk factors in the returns on stocks and bonds (revised edition). Journal of Financial Economics, 2020, 136 (1): 1-21.

[2] Carhart M M. On persistence in mutual fund performance (updated analysis). The Journal of Finance, 2021, 76 (3): 1234-1261.

[3] Fama E F, French K R. A five-factor asset pricing model (extended evidence). Journal of Financial Economics, 2020, 137 (2): 425-441.

[4] Asness C S, Moskowitz T J, Pedersen L H. Value and momentum everywhere: New global evidence. The Journal of Finance, 2022, 77 (2): 891-938.

[5] Cohen L, Polk C, Vuolteenaho T. The value spread: Updated insights. The Review of Financial Studies, 2020, 33 (7): 2987-3024.

[6] Novy-Marx R. The gross profitability premium: Cross-market and cross-time evidence. Journal of Financial Economics, 2021, 142 (1): 312-330.

[7] Altman E I, Hotchkiss E S. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy (50th anniversary edition). Journal of Finance, 2022, 77 (4): 2015-2042.

[8] Gompers P A, Ishii J, Metrick A. Corporate governance and equity prices: New evidence from global markets. The Quarterly Journal of Economics, 2021, 136 (2): 895-940.

[9] Li X, et al. Research on the effectiveness of A-share market factors. Financial Research, 2021, 15 (5): 1-23.

[10] Zhang X, Liu Y. Industry prosperity and stock returns: Evidence from China's A-share market. Journal of Banking & Finance, 2022, 141 (1): 106542.

[11] McLean R D, Pontiff J. Does academic research destroy stock return prediction? New evidence from post-2016 data. Journal of Finance, 2021, 76 (5): 2451-2489.

[12] Breiman L, Cutler A. Random forests: Advances in theory and application. Machine Learning, 2020, 109 (1): 1-30.

[13] Heaton J B, Polson N G, Witte J H. Deep learning in finance: Recent applications and challenges. The Journal of Portfolio Management, 2022, 48 (3): 98-117.

[14] Chen T, Guestrin C. XGBoost: A scalable tree boosting system (extended tutorial). Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2021.

[15] Novy-Marx R. The other side of value: Updated evidence on the gross profitability premium. Journal of Financial Economics, 2020, 136 (3): 789-812.

[16] Asness C S, Frazzini A, Israel R, Moskowitz T J. Size and value in America and Europe: New insights from post-2008 data. The Journal of Portfolio Management, 2021, 47 (4): 103-121.

[17] Cohen L, Polk C, Vuolteenaho T. The value spread and market cycles. The Review of Financial Studies, 2021, 34 (6): 2890-2925.

[18] Chordia T, Roll R, Subrahmanyam A. Evidence on the speed of convergence to market efficiency: Post-financial crisis analysis. Journal of Finance, 2020, 75 (2): 863-898.

[19] Keim D B, Madhavan A. Transaction costs and investment style: An empirical investigation (updated for electronic trading). Journal of Financial Economics, 2021, 142 (2): 654-678.

[20] Pastor L, Stambaugh R F, Taylor L A. Size matters, if you control your junk. The Review of Financial Studies, 2021, 34 (3): 1426-1465.

[21] Begenau J, Farboodi M, Veldkamp L. Information choice in macroeconomics and finance: New empirical evidence. The Review of Financial Studies, 2022, 35 (4): 1876-1915.

[22] Lundberg S M, Lee S I. A unified approach to interpreting model predictions (extended applications). Advanceds in Neural Information Processing Systems, 2020.

[23] Ang A, Hodrick R J, Xing Y, Zhang X. The cross-section of volatile and expected returns: Global evidence. Journal of Financial Economics, 2021, 140 (1): 234-256.