Dean&Francis

ISSN 2959-6157

Theoretical Evolution, Practical Application, and Cutting-edge Challenges of Multi-factor Stock Selection Models

Yimin Yin

Massey Institute at NUFE Nanjing University of Finance & Economics, Nanjing, Jiangsu, 210023, China
*Corresponding author:2120231337@stu.nufe.edu.cn

Abstract:
With the development of financial computing, the importance of quantitative investing has steadily increased. Multi-factor stock selection models, as a crucial component of quantitative investing, play an irreplaceable and crucial role in investment practice. This article first reviews the theoretical evolution of multi-factor models, from the Capital Asset Pricing Model (CAPM) to the Fama-French Five Factors Model (FF5). Secondly, the theoretical models such as CAPM pointed out in this article need to screen out investable factors from academic factors and then be applied in the market to improve the fit between model theory and reality. Next, he demonstrated how multi-factor models can be localized in the Chinese market by introducing factors with high adaptability to the Chinese market, such as pricing factors. Finally, he discussed how, given the current pace of technological development, emerging machine learning, through its superior computational capabilities for high-dimensional, large-scale data, can help multi-factor models improve their predictive effectiveness for future stock market trends and their ability to interpret past data.

Keywords: Multi-factor, Random forest, CAPM model, Fama-French three-factor and five-factor models

1. Introduction

Based on the Efficient Market Hypothesis (EMH), Malkiel defined an efficient capital market as a market where market prices fully reflect all known information. He believed that market prices have fully reflected all available information, and it is difficult for investors to obtain excess returns on a sustained basis [1]. However, the EMH cannot fully explain the frequent asset pricing anomalies in real markets. Due to investor irrationality and market frictional arbitrage, the core contradiction in asset pricing theory revolves around the efficient market hypothesis and systematic pricing anomalies.

Based on the EMH, the Capital Asset Pricing Model (CAPM) was born as the earliest systematic pricing model. Later, in 1993, the Fama-French three-factor (FF3) model emerged, indicating that the previous

ISSN 2959-6157

CAPM model was too simplistic and failed to capture scale effects and value premiums in practical applications [2]. As theoretical models evolved, the concept of multi-factor models entering the market began to take shape, leading to the emergence of the Carhart Four-Factor Model and the Fama-French Five-Factor (FF5) Model. Experiments have shown that the Carhart model is not effective in predicting momentum, while the Fama-French five-factor model can optimize investment portfolios. The theoretical evolution of the model shows that not all factors are effective [3, 4]. Later, due to the blind mining of factors, the phenomenon of over-mining occurred, forming the so-called factor zoo [5]. This led to increased market efficiency, faster factor iteration, and increased investment difficulty. The emergence of the two extremes of single-factor models and factor zoos shows that there is a huge gap between academic theory and real market practice.

This article explores the use of a rational and effective multi-factor stock selection model to bridge the gap between theory and practice, mitigating the low correlation of single factors caused by the unique conditions of real-world markets. The goal is to reveal that markets do not always adhere to rational and objective pricing principles and to demonstrate that multi-factor models are a mainstream strategy in the asset management industry.

The main research issues include the transformation mechanism of academic models into tradable strategies, the special factors existing in the Chinese market, and the close connection between machine learning and stock selection models.

2. Evolution and Controversy of Academic Models

2.1 Capital Asset Pricing Model

The CAPM model, developed in 1964, is the earliest single-factor model, based on portfolio theory and capital market theory. The model assumes the existence and efficiency of market portfolios, the existence of frictionless markets, investors' homogeneous expectations, and investor rationality. However, real markets are not so idealistic. Irrational behavior, risk-based rates, investor perceptions, information biases, and behavioral biases all influence asset pricing. The CAPM model assumes that the only source of systematic risk for stocks is market risk, and that a stock's excess return is solely related to its risk relative to the market. It is precisely because market friction is not fully considered that valuation bias occurs and the explanatory power of the model is reduced. Therefore, after evaluating the CAPM model, Shokhruhmirzo et al. pointed out that although the model is theoretically groundbreaking, it is limited by strong assumptions and is out of touch with reality [6].

2.2 Fama-French Three-Factor Model

In 1993, the Fama-French three-factor model emerged, marking the transition of the multi-factor paradigm from theoretical foundation to market practice.

This model builds on the CAPM by adding two risk factors, size and value, significantly improving its ability to explain cross-sectional stock return variations. Compared to the CAPM, the FF3 model supplements the Small minus big (SMB) and High minus low (HML) factors, providing more realistic factor interpretations, more accurate forecasts, and support for refined portfolio construction. Students from the Vietnam University of Economics and Law studied the application of the FF3 model in the Vietnamese market. The experiment showed that the average R2 of the FF3 model was 81.03%, higher than the 75.61% of the CAPM [7]. In the intercept term comparison, the FF3 model also gained an advantage with a smaller intercept term of 1.08 [7]. Through experimental data, it is found that the FF3 model is better than the CAPM in explaining the relationship between risk and return, especially for small-cap growth stocks (S/L), small-cap value stocks (S/H), and large-cap value stocks (B/L) [7]. Of course, due to the continuous decline in the historical premium of SMB and HML, the FF3 model still cannot perfectly fit the actual market.

2.3 Carhart Four Factors Model & Fama-French Five Factors Model

As factors continue to increase, the Carhart four-factor model and the FF5 model emerge. The Carhart model adds a price momentum factor to the FF3 model. In terms of predictive ability, the Mallows Cp, PRESS statistic, and error variance of the four-factor model are all lower than those of the FF3 model, indicating that the Carhart model has better predictive ability than the CAPM and FF3 models, and is more accurate in capturing short-term fluctuations in fund returns [8]. The FF5 model that emerged later added Robust Minus Weak (RMW) and Conservative Minus Aggressive (CMA) to the FF3 model, attempting to more comprehensively analyze profitability and investment growth, and to explain some anomalies that the FF3 model failed to explain. Paliienko et al. conducted an in-depth analysis of the FF5 model using data from 15 U.S. blue-chip stocks [9]. The results showed that the FF5 model had an average explanatory power of over 78% for six combinations: large-cap stocks-BtM combination, large-cap stocks-ROE combination, large-cap stocks-ROA combination, small-cap stocks-BtM combination, small-cap stocks-ROE combination, and small-cap stocks-ROA combination, but performed the weakest for the small-cap stocks-ROE combination [9]. In the large-cap portfolio, HML and RMW are not significant, and in the small-cap portfolio, HML, RMW, and CMA are all insignificant; all portfolios have significant positive alphas. This indicates that the FF5 model may have missed other risk factors and cannot fully explain the excess returns [9]. Therefore, when applying the FF5 model, it is necessary to dynamically adjust factor indicators, grouping thresholds, and risk assessment tools based on market effectiveness, portfolio composition, and risk distribution characteristics, and pay attention to the possible failure of HML and RMW in specific combinations.

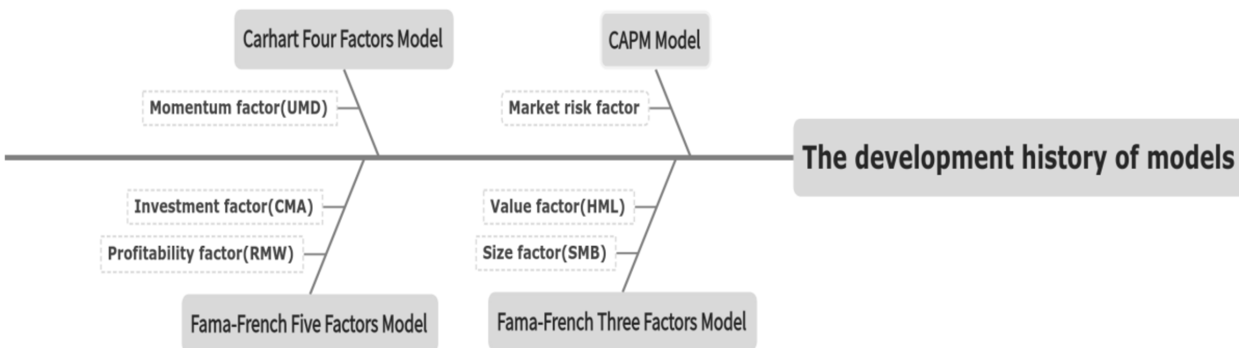

Fig. 1 The History of Theoretical Model Development and the Changes in Each New Model (Original)

Fig. 1 lists the factor models in chronological order from right to left, along with the factors that each model adds to the previous one. The evolutionary process illustrated in Fig. 1 demonstrates that none of the theoretical models described above fully aligns with the market and fully leverages the impact of each factor. Therefore, when applying theoretical models to the real market, careful attention must be paid to the selection of factors within the model.

3. Academic Factors and The Transformation of Tradable Strategies in Market Practice

An academic factor refers to a characteristic or risk exposure that has been empirically shown in economic or financial academic research to significantly explain the cross-sectional return differences of stocks.

Investable factors are defined, systematic risk factors that can be effectively traded and allocated in real markets through specific investment rules and financial instruments. Investable factors are selected from academic factors to identify those that can be leveraged in real markets to generate returns. Table 1 outlines the differences between academic factors and investable factors.

Table 1. Comparison between Academic Factors and Investable Factors

Feature Dimension	Academic Factors	Investable Factors
Construction Method	Annual financial data Decile Long-Short Hedge	Daily or quarterly financial data Factor weighting
Feature	Idealization Purity Emphasizing the explanation of market phenomena	Reality Robustness Cost awareness
Cost Consideration	Transaction costs (such as commissions and stamp duties), impact costs, and taxes are usually ignored.	Fully consider all transaction costs, capital costs, management fees, and custody fees.
Portfolio Turnover	The turnover rate is very high, and little regard is paid to practical operational feasibility.	Strictly control the turnover rate.

As shown in Table 1, academic factors are the theoretical foundation, providing the logical basis for investable factors. Investable factors, on the other hand, are the foundation for strategy implementation, focusing on generating

ISSN 2959-6157

returns.

When screening investable factors, factor cyclicality is an important consideration. Bender et al. noted in their report that from 2001 to 2007, factors like momentum and value performed well, but quality performed poorly. After 2007, the quality factor began to take the lead, while the momentum and value factors became relatively weak [10]. Therefore, the investable factors are not fixed at different times. The selection of investable factors requires incorporating and combining cyclical factors. Therefore, we can see that the cyclical nature of factors directly affects the effectiveness and practicality of strategies. Bender et al. pointed out in their experiments that it is possible to diversify risks and deal with the problem of cyclical changes in factors by constructing a portfolio containing multiple low-correlated factors [10].

In summary, in order to implement the theory, it is necessary to screen out investable factors from academic factors, and gradually transform academic theories into implementable strategies by optimizing stock selection, controlling costs, and managing risks, so as to realize the productization of the model.

4. The Conflict and Integration of Theory and Practice in the Chinese Market

The Chinese stock market is one of the most important financial markets in the world. Due to its unique development path, regulatory environment, investor structure, and macroeconomic background, the Chinese market has many distinct characteristics (as shown in Fig. 2). Investors in the Chinese market are primarily retail investors, with high turnover and short-term speculation. Furthermore, the Chinese market has its own unique regulatory system, with a heavy presence of state-owned enterprises.

Following the example of foreign countries using quantitative investment, and combining the characteristics listed in Fig. 2, some Chinese scholars have also tried to introduce this method into the Chinese market. For example, Luan Qinghai used the CAPM and FF3 models as the research basis and selected the daily data of all components of the Shanghai and Shenzhen 300 Index in 2022 and 2023 for empirical analysis [11].

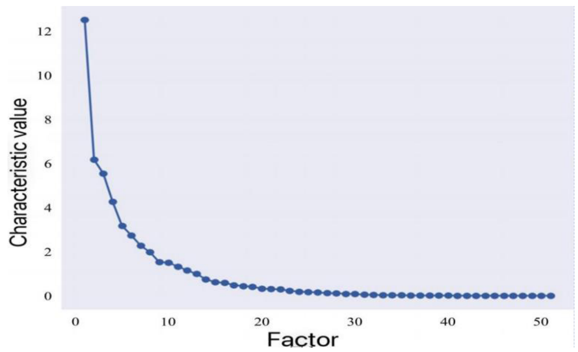

Fig. 2 Scree plot (This picture is from Luan Qinghai's research paper) [11]

This indicates that the explanatory power of the factors added subsequently is decreasing. Therefore, to improve the explanatory power of the model, it is necessary to select appropriate factors. Due to the complexity of the Chinese market, it is far from enough to describe it using the number of factors covered by the CAPM and FF3 models. It is also necessary to explore some factors unique to the Chinese market to improve the adaptability of the model.

Traditional asset pricing models are primarily based on the efficient market hypothesis and the rational investor assumption. However, in reality, investors often exhibit behavioral biases. Individual investors make up the majority of China's A-share market. Therefore, when developing models, investor preferences must also be considered. Based on prospect theory, Zhao Shengmin and Liu Xiaotian introduced investor preferences into the multi-factor model in the form of pricing factors (BMG), established a behavioral asset pricing model, and empirically selected data from China's A-share market from 1995 to 2015. Comparing all the regression models in the experiment, the correlation coefficients of the BMG factors were significantly different from zero [12]. The regression coefficients were statistically significant at the 5% or 10% level [12]. After the introduction of BMG, the adjusted R2 of the regression of each portfolio generally increased, and the unexplained excess returns of most portfolios approached zero [12]. Analysis of the data showed that the BMG factor can significantly improve the ability to explain the cross-sectional differences in the returns of China's A-share market [12].

In summary, choosing a multi-factor model to describe market conditions is the right approach in increasingly complex stock markets. To address the unique characteristics of the Chinese market, unique factors like BMG should be appropriately added to enable the model to transition from theory to practice. However, we must also be mindful of avoiding a factor zoo and avoid the risk of overfitting by blindly exploring factors.

5. Cross-Innovation between Theory and Practice

Traditional financial econometric models are typically linear and parametric. Their strong reliance on theoretical assumptions, economic intuition, and pre-defined factors makes it difficult for these models to capture the complexities of financial markets, such as nonlinearity, high latitude, and interactive effects. However, machine learning (ML), as an emerging technological tool, can largely make up for the weaknesses of traditional models and fundamentally reshape the theory and practice of asset pricing and portfolio management.

ML models are capable of processing high-dimensional data and can simultaneously process a large number of potential factors, including company fundamentals, macroeconomic indicators, and alternative data. Furthermore, ML models do not require pre-assumed linear relationships and can automatically learn and capture complex nonlinear relationships. Unlike the core goal of traditional models, ML aims to improve out-of-sample forecast accuracy, which is highly consistent with the goal of investment decision-making. Therefore, it is wise to introduce machine learning into multi-factor stock selection models to assist research.

In the study of machine learning and portfolio allocation, Pinelis & Ruppert focused on the method of using random forests to optimize the portfolio allocation between market index and risk-free assets. Monthly data of the U.S. stock market from 1927 to 2019 were selected to compare the performance of the benchmark strategy (Base), linear model (OLS), elastic network (Elastic Net), and random forest models in excess return prediction, volatility prediction, Sharpe ratio, drawdown control, Henriksson-Merton, and Treynor-Mazuy tests. It was found that random forest achieved the only positive out-of-sample R2 (0.52%) in excess return prediction, with a directional accuracy of 64.52% [13]. In volatility prediction, its directional accuracy was the highest (80.91%) [13]. The strategy using random forest for return-risk timing has the highest Sharpe ratio (0.73) and an annualized Alpha of 3.37% [13]. In terms of drawdown control, its maximum drawdown does not exceed 30% [13].

The above excellent experimental results not only demonstrate the strong data processing capabilities of machine learning but also prove that it can significantly improve the performance of portfolio allocation and provide new methods and support for active asset management.

6. Limitations

Looking at the development of multi-factor stock selection models, current research is still not perfect enough.

First, there's the issue of factor validity, requiring attention to the impact of market conditions, economic cycles, policy adjustments, and investor structure on factors. Second, previous models relied too heavily on historical data, resulting in inaccurate predictions for the future. The model's decision-making is also not flexible enough when responding to unexpected market conditions. Furthermore, the model has high requirements for both the quality and quantity of data. Since ordinary investors have difficulty accessing large amounts of data, they can only make predictions and investments based on data of limited dimensions. The quantity and quality of such data cannot meet the standards of the model, thus affecting the accuracy and reliability of the model. Finally, we should be sensible when mining factors. Too many factors and overly complex parameters will lead to overfitting of the model, which is not conducive to model operation.

In terms of innovation, as far as machine learning is concerned, the degree of its integration with models cannot yet meet all the needs of the market. Bartram et al. conducted experiments to test the application of ML in key aspects of active investment management, such as signal generation, portfolio construction, and execution, and evaluated its actual effects. The results show that the 13 exchange-traded funds analyzed in the experiment had small but rapidly growing assets under management, high management fees, and generally failed to generate significant alpha. Their risks primarily stemmed from their exposure to style factors, particularly their bias towards small-cap and less liquid stocks. The above results show that financial data is noisy, has a short time series, and has an unstable distribution, which can limit the learning effect of ML models. In addition, it is easy to fit noise in high-dimensional data. At the same time, the computation and implementation costs of ML models are also very high [13]. Therefore, the integration of machine learning and multi-factor models still has broad development prospects.

The limitations are particularly evident for models applied to the Chinese investment market. Due to the different maturity levels of China's capital markets compared to overseas markets, simply copying factor combinations would affect the effectiveness and sustainability of these models. Therefore, in terms of theoretical localization, it is necessary to systematically build a library of driving factors specific to the Chinese market. Further attention should be paid to the survival patterns of these factors in real trading, and the model should be continuously refined based on market feedback. Ultimately, this will ensure a balance between scientific validity and profitability.

ISSN 2959-6157

7. Conclusion

This article reviews the evolution of multifactor models from academic theory to practical application, focusing on the challenges and development prospects of these models in the Chinese market. Currently, in order to make multifactor models more adaptable in the Chinese market, we need to explore more relevant factors. When building a factor library specifically for the Chinese market, it is also important to consider the rationality of the number of factors. The application of machine learning combined with models also demonstrates the broad development prospects of multi-factor models. Through layer-by-layer analysis of theoretical evolution, practical application, localization conflicts, and cutting-edge innovative development, the article draws the following conclusions: the multi-factor model successfully builds a bridge between financial theory and investment practice, helping to achieve the transformation from theory to strategy implementation.

Looking ahead, the development of multi-factor models will increasingly rely on the feedback loop between theory and practice. Multi-factor models are not static formulas, but rather analytical systems that evolve dynamically with the market and practice. The core of its effectiveness stems from the continuous interaction and verification between theoretical research and market practice. Only by establishing an efficient feedback mechanism between theory and practice can the model be continuously optimized and upgraded in increasingly complex market conditions, ultimately achieving a synergistic unity of scientific validity and profitability.

References

[1] Tian A. G. The efficient market hypothesis: review of specialized literature and empirical research. Procedia Economics and Finance, 2015, 32: 442–449.

[2] Fama E. F., French K. R. The capital asset pricing model: theory and evidence. Journal of Economic, 2004, 18(3): 25–46.

[3] Paliienko O., Naumenkova S., Mishchenko S. An empirical investigation of the Fama–French five-factor model. Investment Management and Financial Innovations, 2020, 17(1).

[4] Tazi O., Aguenaou S., Abrache J. A comparative study of the Fama–French three-factor and the Carhart four-factor models: empirical evidence from Morocco. International Journal of Economics and Financial Issues, 2021, 12(1): 58–66.

[5] Harvey C. R., Liu Y. A census of the factor zoo. SSRN Electronic Journal, 2019.

[6] Boynazarov S., Mustafakulova S., Mustafakulov I., Djurayeva Z. A critical evaluation of asset pricing models: CAPM, APT, and the Fama–French multifactor model. American Journal of Social Science, 2025, 3(4).

[7] Nguyen A. P., Tran V. H. Applying Fama and French three factors model and capital asset pricing model in the stock exchange of Vietnam. International Research Journal of Finance and Economics, 2012, 95.

[8] Bello Z. Y. A statistical comparison of the CAPM to the Fama–French three-factor model and the Carhart's model. Global Journal of Finance and Banking Issues, 2008, 2(2).

[9] Bender J., Briand R., Melas D., Subramanian R. A. Foundations of factor investing. Foundations of Factor Investing, 2013.

[10] Zhao H. Research on a multi-factor stock selection model based on the CSI 300 Index. Operations Research and Fuzziology, 2023, 13(6): 6204–6210.

[11] Zhao S., Liu X. Introducing a multi-factor model based on investor preferences: an analysis from the perspective of prospect theory. China Economic Studies, 2019, 2.

[12] Pinelis M., Ruppert D. Machine learning portfolio allocation. The Journal of Finance and Data Science, 2022, 8: 35–54.

[13] Bartram S. M., Branke J., De Rossi G., Motahari M. Machine learning for active portfolio management. WRAP Journal, 2021.