

## Establishment and Optimization of Multi-factor Model

Pan Xu, Mengtian Wang, and Jun Liu\*

School of Management Science and Engineering, Nanjing University of Finance and Economics, No. 3  
Wenyuan Road, Nanjing, China

**Keywords:** Asset pricing model; Five-factor model; Multi-factor model

**Abstract:** Compared with the traditional investment method, quantitative investment has the advantages of fast data acquisition, rigorous processing method and free transaction logic. Quantitative investment in China is still in its infancy, so the research on quantitative model and quantitative investment strategy suitable for Chinese stock market has great development space and investment value. On the basis of capital asset pricing theory, quantitative trading develops from single factor to multi-factor model. Multi-factor stock selection strategy has become one of the most widely used strategies in the field of quantitative investment in China due to its advantages such as flexibility and strong ability to adapt to the market. With the development of big data and machine learning algorithms, scholars continue to combine artificial intelligence with multi-factor stock selection to build portfolios to obtain excess returns. Based on the capital asset pricing model and Fama-French five-factor model, this paper explores the future development direction of the multi-factor model. With the development of new theories, new methods, new data collection and new problems, new impact factors and new algorithms will be constantly discovered.

### 1. Introduction

Capital asset pricing is one of the three major areas of modern finance, and aims to explain the impact factors of stock returns from the perspective of the relationship between risk and return and quantification. The capital market of the United States started early and has developed many widely accepted and classical asset pricing theories. The classical CAPM model proves that the expected excess return of assets relative to risk-free assets is proportional to the risk-free assets of the assets through the single-factor model. With the development of time, in the later empirical tests, the interpretation of the CAPM model has gradually declined, and many financial anomalies cannot be explained. Since then, the validity and applicability of the CAPM model have been questioned by some scholars. Then, the intertemporal asset pricing model ICAPM and the arbitrage pricing model APT extended the stock impact factor from a single factor to multiple factors.

Under the multi-factor asset pricing, Fama and French (1993) [1] proposed the famous three-factor model based on the arbitrage pricing model, and established the multiple linear regression model of the excess stock return rate to market value factor, book-to-market equity and market factor. The Fama-French three-factor model, q-factor model and other multi-factor models have become asset pricing models to better explain the changes of stock returns in the United States in different periods, providing powerful tools for the study of capital market in the US. Fama and French (2015) [2] proposed the famous five-factor model on the basis of the original three-factor model, and added RMW and CMA on the basis of the original Fama-French three-factor model, so as to better explain the changes of stock yield in the United States.

Chinese stock market, established in 1990, is a typical emerging market. Compared with mature stock markets in Europe and the United States, there are still many imperfections. The capital asset pricing model is based on strict theoretical assumptions. Even in developed capital markets, it may not be able to fully meet these harsh assumptions. Because stock markets in different countries meet different levels of assumptions, the effectiveness of capital asset pricing models in different markets varies widely. In China, the A-share market is the main stock market. The A-share market has a long history of development, large investment and financing scale, and relatively mature development, but there is still a gap compared with mature European and American capital markets.

Although countless scholars and financial practitioners have conducted a large number of empirical studies on Chinese stock market, due to the short time of research on asset pricing in China, an effective theoretical system of asset pricing has not yet been formed. There are many differences between A-share stock market and foreign stock markets in terms of market operation mechanism and market environment. Existing foreign research theories and pricing models may not be applicable to A-share stock market. Therefore, to test the effectiveness of existing foreign asset pricing models in A-share market, and to explore the asset pricing models applicable to A-share market are research topics that have both theoretical and practical value.

With the development of multi-factor theory, people have begun to study more and more comprehensively use multiple factors to explain the reasons for stock changes. Since fundamental indicators are relatively easy to obtain, research on fundamental indicators has developed earlier. However, due to the continuous maturity and improvement of the capital market, the number of stocks is also increasing, the indicators of stocks are gradually becoming richer, and the indicators and the number of stocks that investors need to study are becoming more and more complex. In order to increase the speed of calculation and enhance the accuracy of investment, quantitative stock selection has gradually begun to combine big data and machine learning methods to study how to screen outstanding stocks and establish investment portfolios to obtain excess returns.

## **2. Literature Review and Classification**

Foreign research on quantitative investment has been carried out earlier. The more widely recognized article of origin is the mean-variance theory proposed by Markowitz (1952) [3] in his doctoral dissertation. For the first time, Markowitz analyzed the variance and average return of the investment portfolio in the article, and proposed to obtain the maximum return under the same risk. For the first time, quantitative thinking was used in the field of financial investment.

The famous economist William Sharp (1966) [4] published an article, proposed the famous CAPM model, and also laid the theoretical foundation for the multi-factor quantitative stock selection model. For the first time, CAPM divested the returns of individual stocks, and for the first time proposed the concepts of systemic risk and non-systematic risk.

On the basis of CAPM, Rose(1976) [5] proposed the Arbitrage Pricing Theory (APT), which expanded the factors from one to a group based on the portfolio pricing model, and considered that the return of securities was linearly related to a group of factors. This idea is very close to the modern multi-factor stock selection model, which basically completed the theoretical foundation of the early stage of quantitative investment and established theoretical guidance for later researchers.

Fama and French (1993) [1] proposed a well-known three-factor model based on the arbitrage pricing model, and established a multiple linear regression model of stock excess return on market value factor, book market value ratio factor and market factor. At the end of the article, Fama stated that the model itself has limitations, and some excess returns in the market cannot be explained well by these three factors.

Fama and French (2015) [2] improved the original three-factor pricing model by adding two new factors, investment model and profitability, to build a five-factor pricing model.

### **2.1 Empirical research on the five-factor model in various markets**

Fama and French (2017) [6] did not stop at the study of the US market. They extended the application of the five-factor model to the international market and found that the applicability of the five-factor model to different countries' markets is different. This is consistent with the conclusions reached by some scholars in other countries. Foye (2018) [7] used samples from 18 countries in three emerging regions to study the applicability of the five-factor model in emerging markets, and concluded that the five-factor model performed well in eastern Europe and Latin America, but not significantly in Asian markets.

At this stage, most of the domestic researches on the Fama-French model focus on the empirical test of the applicability of the five-factor model in Chinese A-share market. However, in the empirical process, the selected sample interval and range, the factor construction method, and the

grouping method are different to varying degrees, so the research conclusions of such literatures are not the same.

Zhao et al. (2016) [8] empirically analyzed Chinese A-share market transactions and financial data, and found that the A-share market's market value effect and value effect price are obvious, and RMW and CMA are two factors that are not effective for stock portfolio returns. For the Chinese market, the three-factor model performs better than the five-factor model. Li et al. (2017) [9] believes that the five-factor model has a very strong interpretation ability. Compared with the CAPM model, the three-factor model, and the Carhart four-factor model, the five-factor model performs better, and the size and book market value ratio effect of the entire sample are significant. After adjusting through the three-factor model, its profit and investment is obvious, but there is no significant momentum or reversal effect. Guo et al. (2017) [10] provide out-of-sample tests of the five-factor model introduced by Fama and French for the Chinese stock market, and find strong size, value and profitability patterns in average returns, but weak investment pattern.

## **2.2 Explore new impact factors**

The multi-factor strategy is expanded on the basis of the Fama-French three-factor strategy. By continuously absorbing effective factors that conform to market rules, it achieves more accurate forecasting capabilities.

Later generations use three-factor or five-factor models to test in different countries on the basis of models, and at the same time continue to expand factors. Havery (2016) [11] constructed a multi-testing framework that accommodates 316 factors and validated the work of previous scholars. Green (2017) [12] selected 94 factors for regression calculation during the period from 1980 to 2014. The results show that only 12 factors can independently provide important information on the average monthly return of US stocks, and the remaining 82 factors cannot be independent to provide information.

Domestic scholars are also constantly looking for effective factors based on the characteristics of the Chinese stock market. Zhang (2014) [13] constructed 25 candidate factors based on comprehensive economic laws, market experience, and data availability, and used market and financial data from January 2000 to December 2009 to test the validity of the factors. Wang et al. (2018) [14] constructed a quantitative timing strategy and verified that the quantitative timing strategy is effective for the constituents of the Shanghai 180 Index. Li (2019) [15] used the scoring method to build an eight-factor stock selection model. Through back-testing from January 2016 to April 2019, the model has a relatively stable excess return compared to CSI 300.

## **2.3 Application of machine learning in multi-factor model**

Quantitative investment theory has been continuously improved and developed over the past decades, and the theory of multi-factor stock selection models has become increasingly mature. With the continuous development of computer algorithms, various machine learning models continue to emerge and develop, and scholars and financial industry practitioners from various countries also continue to combine artificial intelligence with multi-factor stock selection.

Kim et al. (2003) [16] used support vector machine model for stock price index prediction, and compared the performance of SVM and BP neural network in stock price prediction. Experimental results show that the support vector machine model has better prediction effect. Huang (2012) [17] established a GA-SVR model and tested its validity, and verified the effectiveness of the method by comparing it with benchmarks. Risul (2015) [18] uses support vector regression as a machine learning technique and uses different oligarch operators as input selection techniques or data preprocessing steps. Experiments were performed on three different stock indexes of the Bangladesh Dhaka Stock Exchange (DSE), the S & P 500 Index, and the IBM Index.

Cao et al. (2014) [19] used a random forest algorithm to establish a quantitative stock selection model. At the same time, they also compared the random forest stock selection model with a support vector machine model and found that the random forest algorithm performed better. Shi (2017) [20] combined principal component analysis and random forest algorithm and applied it to Barra's quantitative hedging model to design and implement a complete multi-factor stock selection

system. Wu (2018) [21] used 22 technical indicators such as KDJ, MACD, RSI, ROC, and Bollinger Bands to construct a quantified stock selection model based on the random forest algorithm, and conducted stock selection predictions every week. The results outperformed the SSE 500 Index.

## 2.4 Chapter summary

By combing relevant literature, we can find that at present foreign scholars focus on various anomalies in the financial market on the one hand, and verify whether the factor model has an explanatory ability to the capital market rate of return, on the other hand, they are mining the theoretical basis of the factor model in order to fill the gap of weak factor model theory.

Domestically, most researches focus on exploring the applicability of the Fama-French five-factor model in Chinese securities market, but the conclusions are not the same. With the continuous development of the capital market, more and more financial anomalies cannot be explained by Fama-French factor model, such as momentum effects. In view of the characteristics of the A-share market, some scholars began to make factor corrections on the basis of the original model to form an extended multi-factor model to explain the stock return of the A-share market. At the same time, more and more scholars use models such as machine learning and deep learning to mine sample information. Various machine learning algorithms are also more widely used in multi-factor stock selection.

## 3. Conclusion

Capital asset pricing theory prices assets from the relationship between risk and return. The premise of capital asset pricing theory is that the market is efficient. It mainly expounds the relationship between expected return rate of assets and risk under the state of equilibrium market. The theory of asset pricing model has been developed from single-factor model to multi-factor pricing model. Fama-French five-factor model has relatively fully described the shape of influencing the change of return rate. At present, it has pushed the study of asset pricing to the front and the limit of the cognitive ability of the financial circle. However, this is not the final destination of capital asset pricing model research. With the proposal of new theories, the discovery of new methods, the collection of new data and the emergence of new problems, new influencing factors will be constantly discovered. At the same time, with the combination of theoretical empirical research and practical investment, people can always find new market anomalies one after another. When these abnormal variables are added into the model, they can better explain the systematic change pattern in the rate of return and challenge the original model.

## Acknowledgements

This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19\_1431).

## References

- [1] Fama, E.F., French, K.R. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 1992, 47: 427-465.
- [2] Fama, E.F., French, K.R. A Five-factor Asset Pricing Model. *Journal of Financial Economics*, 2015, 116: 1-22.
- [3] Markowitz, H.M. Portfolio selection. *Journal of Finance*, 1952, 7: 77-91.
- [4] Sharpe, W.F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 1964, 19: 425-442.
- [5] Ross, S.A. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 1976,

13: 341-360.

[6] Fama, E.F., French, K.R. International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 2016, 123: 441-463.

[7] Foye, J. A Comprehensive Test of the Fama-French Five-Factor Model in Emerging Market. *Social Science Electronic Publishing*, 2017, 37: 199-222.

[8] Zhao, S., Yan, H., Zhang, K. Is the Fama-French five-factor model better than the three-factor model - empirical evidence from Chinese a-share market. *Nankai economic research*, 2016, 2: 41-59.

[9] Li, Z., Yang, G., Feng, Y., Jing, L. Empirical test of Fama-French five-factor model in Chinese stock market. *Finance research*, 2017, 6: 191-206.

[10] Guo, B., Zhang, W., Zhang, Y.J. and Zhang, H. The Five-Factor Asset Pricing Model Tests for the Chinese Stock Market. *Pacific-Basin Finance Journal*, 2017, 43: 84-106.

[11] Harvey, C.R., Yan, L., He, Z. The Cross-Section of Expected Returns. *Electronic Journal*, 2015, 29: 5-68.

[12] Green, J., Hand, J.R.M., Zhang, F. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *Social Science Electronic Publishing*, 2017, 30: 4389-4436.

[13] Zhang, L. Quantitative stock selection based on multi-factor model. *Hebei university of economics and business*, 2014.

[14] Wang, C., Liu, G., Wang, Q. Multi-factor quantitative stock selection model and timing strategy. *Journal of Dongbei University of finance and economics*, 2018, 05: 81-87.

[15] Li, Q. Establishment and optimization of multi-factor quantitative stock selection model. *Shandong university*, 2019.

[16] Kim, K. Financial time series forecasting using support vector machines. *Neurocomputing*, 2003, 55: 307-319.

[17] Huang, C.F., Hsieh, T.N., Chang, B.R. A comparative study of stock scoring using regression and genetic-based linear models. 2011 IEEE International Conference on Granular Computing, GrC-2011, Kaohsiung, Taiwan, 2011.

[18] Rasel, R.I., Sultana, N., Meesad, P. An efficient modelling approach for forecasting financial time series data using support vector regression and windowing operators. *Inderscience Publishers*, 2015, 02: 134-138.

[19] Cao, Z., Ji, H., Xie, B. Random forest algorithm to select high-quality stocks. *Journal of capital university of economics and business*, 2014, 16: 21-27.

[20] Shi, W. Application of PCA and random forest in BARRA quantitative hedging model. *Xi An university of science and technology*, 2017.

[21] Wu, W. Application of random forest in quantitative stock selection of technical indicators. *University of electronic science and technology of China*, 2018.