Research on Quantitative Investment Strategy Based on Multi-factor Stock Selection Model

Fangfei Li(✉)

Business School, Hohai University, Nanjing 211100, China
13851286242@stu.sqxy.edu.cn

Abstract. Multi-factor stock selection is an important part of quantitative investment. The research on multi-factor stock selection is conducive to rational investment transactions. This paper selects the data of Shanghai 50 Index constituent stocks from April 2018 to March 2023 and relevant stock factors. PCA principal component analysis model, equal weight model and comprehensive scoring model are respectively used for stock selection. On this basis, it is found that compared with the other two stock selection models, the equal weight stock selection model has stronger stock selection ability. The IC value and IC_IR value of the factor are used to test the effectiveness of the factor in the back test. It is found that the two stock factors of PB and TR have stronger stock selection ability. Finally, this paper analyzes and compares the stock selection conditions under the three stock selection models, and draws relevant conclusions. The study presented in this paper has some theoretical importance as well as application to the creation of a quantitative investment strategy.

Keywords: Multi-Factor Stock Selection · Quantitative Investment · Principal Component Analysis · Equal-weight Model · Scoring Method

1 Introduction

Quantitative investment has been born out of investors' appetite for yield and aversion to risk. Quantitative investment is an investment strategy. It makes comprehensive use of statistics, mathematics and computer knowledge and tools to build quantitative models for investment decisions. The multi-factor stock selection model has become increasingly popular among institutions and people throughout the globe as one of the quantitative investment methods in recent years. It has made considerable contribution to the return of investment.

After long-term development, multi-factor quantitative stock selection strategy has formed a relatively mature research system and theoretical framework. In 1952, Markowitz established the portfolio management theory based on mean-variance model. He applied mathematical statistics methods and ideas to investment science, and formed the initial concept of quantitative investment [1]. This theory established the analysis paradigm of risk-return balance in finance and became the logical starting point of many quantitative investment models later. Based on Markowitz's mean variance theory and

S. Tehseen et al. (Eds.): ICEDBC 2023, AEBMR 258, pp. 187–193, 2024.
https://doi.org/10.2991/978-94-6463-246-0_22

its inherent limitations, Sharpe, Litner and Mossin explored the internal operating mechanism of the capital market, and combined with the optimal equilibrium conditions of asset pricing in the market. The target portfolio was represented by the capital asset pricing model (CAPM), which was developed by them, as a linear combination of risky and safe assets. CAPM also became the basic model for portfolio risk research later [2]. In 1976, Ross improved the CAPM model and established the arbitrage pricing theory (APT) from the perspective of statistical arbitrage. It elaborated the role of various economic factors affecting stock price changes and laid a foundation for subsequent multi-factor analysis [3].

To further explain the volatility of stock expected return, Fama and French introduced the "Fama-French three-factor model" based on the CAPM model and APT model [4]. According to this, Liew et al. studied the stock markets of 10 developed countries and found Fama-French, finding that the scale effect and value effect in the factor model can well predict the expected return rate of stocks [5].

Gruber and Carhart developed a four-factor model based on the Fama-French three-factor model by introducing bond return factor and momentum factor respectively, and found through empirical test that the model was more effective in portfolio risk identification and yield differential analysis [6, 7]. On top of the three factors introduced in 1993, Fama and French added the business profitability factor and investing style factor in 2015. They built a more complete five-factor model and confirmed the applicability of the five-factor model in the international market [8]. Guerard et al. constructed a 10-factor stock selection model including momentum factor and value factor, and verified that the model has outstanding stock selection ability and its selected stock portfolio can bring significant investment returns [9]. Balakrishnan tested the impact of scale, value and momentum factors on stock returns, and empirical results showed that they had a significant impact on returns, and the multi-factor stock selection strategy containing these factors could bring investors satisfactory excess returns [10].

On these bases, this paper selects stock factors like PE to conduct multi-factor quantitative analysis on Shanghai Stock Exchange 50 shares to verify the effectiveness and effect of multi-factor quantitative stock selection strategy.

2 Methods

2.1 Data Source and Basic Data Processing

The data of this study includes Shanghai 50 constituent stocks from March 2018 to March 2023. The data source is Wind Economic Database.

In terms of basic data processing, in order to better analyze the effect of each factor, the factors are mainly de-maximized and standardized, and the industry and market value are neutralized. First of all, check for missing data values and populate them. 3σ method is used to cut out the numbers that are too large or too small for each factor. This part adjusts the mean value and standard deviation of the original exposure value of the factor without the extremes and outliers to 0 and 1, which can retain more information about the distribution of factors. In order to put forward the influence of other factors when using a certain factor and make the stock selection results more dispersed, this part neutralizes the standardized factor value.

Table 1. The selected stock factors

Factor class	Factor shorthand	Factor name
Valuation factors	TTM	Price Earnings Ratio (Trailing twelve months)
	PE	Price Earnings Ratio
	PB	Price-to-book Ratio
	PS	Price-to-sales Ratio
	PCF	Price-to-cash Flow Ratio
Technical factors	TR	Turnover Rate
	ETR	Effective Turnover Rate

2.2 Factor Selection

In terms of factor selection, this paper refers to the economic theory and the previous experience of relevant scholars, and divides the fundamental factors of stocks into five categories. These five factors are: valuation factors, operating effectiveness factors, growth ability factors, profitability factors and technical factors. In this paper, two types of factors, valuation factors and technical factors are selected for research and analysis. The selection of specific factors is shown in Table 1.

2.3 Research Method

This research method mainly includes principal component analysis, equal weight method, comprehensive scoring method. Based on the selected stock factors, this paper respectively uses the above three methods to establish a stock selection model, back tests strategy returns and finally visualizes the revenue.

Principal Component Analysis Model. Too many variables will increase study complexity when employing statistical analysis to investigate a multi-variable topic. Therefore, it is very important to reduce the number of variables studied, that is, to reduce the dimension of data processing. The goal of PCA, or principal component analysis, is to identify the most crucial aspect of the data, replace the original data with this aspect, create as few new variables as possible—so that they are not correlated with one another—and use these new variables as closely as possible to the original data in order to preserve it. In this paper, the historical data of the previous 80 trading days are used to generate the weight of the corresponding stock factors, and PCA dimension reduction is carried out to integrate all factors into one factor. At the same time, the strategy backtest is carried out to forecast the earnings of the next trading day, calculate the average value, sort them, and select the top 5 stocks with the highest returns.

Equal Weight Model. Equal weight is to assign equal weight to each research variable so as to carry out data analysis. In this part, the weight of each factor is set to be equal, and on this basis, the factors are merged. Then, take the top 5 stocks with the highest returns when conducting a strategic return backtest.

Comprehensive Scoring Model. The objective of a comprehensive scoring approach is to quantify the items through scoring, which can be applied to the thorough evaluation of qualitative ranking issues. Its main objective is to give various grades of evaluation various scores and then base a thorough evaluation on those ratings. IC value, also known as information coefficient, represents the ability of factors to predict stock returns. The correlation between stock components from the current trading day and returns from the following trading day is referred to as the "IC value" in this study. The range of its value is [−1,1]. The larger the absolute value is, the better the prediction ability is. The IC value is used to score each factor, and the strategic return backtest is carried out. The top five stocks with the highest return are selected.

3 Results and Discussion

3.1 Analysis of Stock Selection Returns

To make it easier to compare the results of the three stock selection models, the line chart is used to present their returns and changing trends. Figure 1 shows the trend of returns over the past five years from April 2018 to March 2023 as a result of the three methods of stock selection.

From the perspective of overall returns, before 2020, that is, the first two years of return backtest, the three stock picking techniques' returns do not significantly differ from one another. However, since 2020, the equal-weight stock selection model is significantly better than PCA stock selection model and comprehensive stock selection model, and this advantage continues to expand in the following three years. From 2018 to the beginning of 2021, namely the first three years of return backtest, PCA principal component analysis stock selection model and comprehensive stock selection model have great similarity in return volume and change trend. However, since the middle of 2021, the returns brought by PCA stock selection model began to be gradually higher than those brought by comprehensive scoring model, and this advantage basically showed an expanding trend.

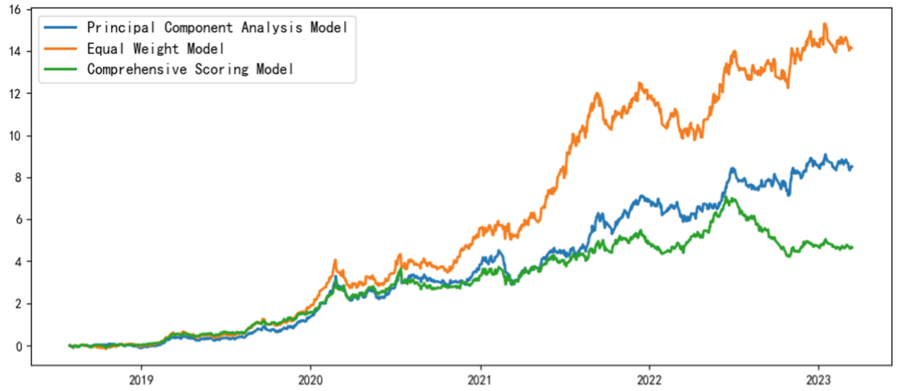

Fig. 1. The line chart of returns of three stock selection models

From the perspective of the volatility of returns, the three stock selection models all present a certain degree of volatility and the volatility trend is roughly the same in the strategy backtest time range. In the first two years of the return backtest, the volatility of the three stock selection models is not obvious, basically showing a rising trend, and the rising amplitude increases in the early 2020. In 2020, it basically shows irregular fluctuations, while in 2021, it basically shows an upward trend. The rising range of the equal-weight stock picking model is significantly larger than the other two models. During this period, the returns of the three stock picking models all show slight fluctuations. In early 2022, the returns of the three stock picking models all showed a slight decline, and then began to rise, reaching a peak in mid-2022. From mid-late 2022 to early 2023, PCA principal component analysis stock selection model and equal weight stock selection model both show an upward trend, while the comprehensive stock selection model shows a downward trend and tends to be flat.

3.2 Analysis of Factor Validity

The effectiveness of factor stock selection is the basis of the success of multi-factor model. Based on the above analysis of stock selection returns, this part analyzes the effectiveness of the selected factors with the IC value and IC_IR index of the selected factors.

It is thought that a factor has a favorable stock selection effect if the long-term average IC of the factor has an absolute value greater than 0.02. Stocks with a high exposure value to the factor should be long, according to the IC positive indicator, which shows that the factor has a positive connection with the stock's future returns. Conversely, negative IC indicates stocks with low multifactor exposure.

In the return test interval, the IC value of each factor of each trading day is calculated, and the average IC value and IC_IR value are obtained as shown in Table 2.

The information ratio derived from the IC value is known as the IC_IR index. Calculation is done by dividing the IC mean by the IC standard deviation. The stock selection effect of this factor is better the greater the IC_IR's absolute value. The IC_IR index's stock selection criteria are in line with the IC value. Stocks having a high exposure

Table 2. IC value and IC_IR value of each factor

Factors	IC Mean	IC_IR
PE_TTM	−0.014149	−0.060897
PE	−0.009846	−0.045319
PB	−0.027065	−0.105445
PS	−0.006961	−0.030149
PCF	0.003385	0.022275
TR	−0.022535	−0.091979
ETR	−0.006054	−0.027051

Table 3. The top eight picks of the three methods

Model	Top eight most picked stocks
PCA	603986, 603288, 600436, 600809, 603501, 603799, 600519, 601066
Equal Weight	603986, 600436,603501, 603288, 601066, 603799, 600276, 603259
Scoring Method	603986, 603288, 603501, 600436, 601066, 600276, 603259, 603799

value to the overallocation factor will be chosen if IC_IR is positive; stocks with a low exposure value to the overallocation factor will be chosen if IC_IR is negative.

According to the results, PB and TR are significant for the same type of variables, IC absolute values are bigger than 0.02 and IC IR absolute values are higher, which indicates that the PB and TR factors of stocks have stronger stock selection ability compared with other factors selected. On this basis, the IC values of these two factors are negative, indicating that they have a long-term and stable negative relationship with future earnings. In contrast, factors PE, PS and ETR have weak stock selection ability and can be eliminated.

3.3 Analysis of Stock Selection

Based on the three methods of stock selection, the selection times of each stock are analyzed and discussed. The top eight stock tickers in each method are shown in Table 3.

It can be seen from the table that the top stocks in the stock selection times of the three stock selection models have certain similarities. Stocks with codes of 603986, 603288, 600436, 603501, 603799 and 601066 were selected at the top of the list under the three stock selection models, indicating that these stocks have good profitability.

Under the three stock selection methods, GigaDevice Semiconductor Inc. (stock code: 603986) is selected the most frequently, accounting for 9.33%, 11.68% and 8.37% of the total stock selection times respectively. Foshan Haitian Flavoring and Food Company Ltd. (stock code: 603288) is selected only second to GigaDevice, accounting for 9.24%, 8.41% and 8.33% respectively under three stock selection meth Zhangzhou Pientzehuang Pharmaceutical Co., Ltd. (stock code: 600436) ranks the third in overall selection times, accounting for 9.01%, 9.64% and 7.10% respectively.

4 Conclusion

This paper analyzes and compares the stock selection effects of three stock selection models, namely principal component analysis model, equal weight stock selection model and comprehensive type stock selection model, through the stock factors of Shanghai Stock Exchange 50 components from April 2018 to March 2023, and the return test. Through the visualization analysis of the returns of the three stock selection models, it can be concluded that the equal weight stock selection model has obvious strong stock selection ability, followed by PCA model. From the medium- and long-term perspective, even if the return fluctuates in the return test interval, the three stock selection models have

profit effect. In addition, factor effectiveness analysis is carried out, and it is found that factors PB and TR have strong stock selection ability. It is found that stock codes 603986, 603288 and 600436 are at the forefront of stock selection, showing good profitability. However, only seven factors such as TTM and PE were selected in this study, so the Stock-picking results have limitations.

References

1. Markowitz, H.: Portfolio selection. The Journal of Finance 7(1), 77-91 (1952).
2. Sharpe, W. F.: Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425-442 (1964).
3. Ross, S.: The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13(3), 341-360 (1976).
4. Fama, E. F., K. R.: French Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3-56 (1993).
5. Liew, J., Vassalou, M.: Can book-to-market, size and momentum be risk factors that predict economic growth. Social Science Electronic Publishing 57(2), 221-245 (1999).
6. Edwin, J., et al.: The persistence of risk-adjusted mutual fund performance. The Journal of Business 69(2), 133 (1996).
7. Carhart, M. M.: On persistence in mutual fund performance. The Journal of Finance 52 (1), 57-82 (1997).
8. Fama, E. F., French, K. R.: A five-factor asset pricing model. Journal of Financial Economics 116(1), 1-22 (2015).
9. Guerard, J. J., et al.: Earnings forecasting in a global stock selection model and efficient portfolio construction and management. International Journal of Forecasting 31(2), 550-560 (2015).
10. Balakrishnan, A.: Size, Value, and Momentum Effects in Stock Returns: Evidence from India. Vision 20(1), 1-8 (2016).