Research on Stock Selection Scheme Based on Quantitative Multi-factor Model

Haiwen Liu[1](✉), Yongxin Chen[2], Xiang Zhu[2], and Fan Wu[1]

[1] Cardiff Business School, Cardiff University, Cardiff, UK
{LiuH76,WuF16}@cardiff.ac.uk
[2] School of Management, Fudan University, Shanghai, China
ChenY210@cardiff.ac.uk, 20260691281@fudan.edu.cn

Abstract. With the continuous improvement of the theoretical foundations of mathematical and behavioral finance, quantitative investments have sufficient and necessary conditions to develop vigorously. In the society of increasing investment, quantitative investment has become an inevitable trend. In this paper, based on the multi-factor stock selection model, we analyze some financial indicators and trading data of listed companies in detail by quantitative methods, and test the model by using the data of A-share market in combination with the statistical test method. It achieves the purpose of providing investment reference and advice for stock investors. This study aims to help investors identify the most valuable stock portfolios.

Keywords: Multi-Factor · Quantitative · Stock Selection Model · In-Dividual Stock Portfolio

1 Introduction

With the rapid development of China's economy and the improvement of national living standards, the asset management industry is also booming and gradually becoming one of the main investment tools for institutions or individuals. In China today, the main force of asset management is slowly shifting towards fund companies. Since fund companies are established to invest and add value to assets, this is in line with the core philosophy of the asset management industry. Secondly, with the increasing regulatory pressure, fund companies with their strong talent pool and core strategies are under much less pressure in the face of regulation. Behind many mainstream private and public offerings, quantitative investment is quietly rising to become one of the important investment tools for fund companies. Among the stars of quantitative investment, multi-factor quantitative strategies are undoubtedly one of the brightest stars. Multi-factor model assumes that the market is inefficient or weakly efficient, and obtains excess returns through active portfolio management. The core idea of multi-factor stock selection is that market influences are multiple and dynamic, but there are always some factors that can play a stable role in a certain period of time. In quantitative practice, there are various multi-factor models constructed because different market participants or analysts have different understanding of market dynamics and factors [5].

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 586–594, 2023.
https://doi.org/10.2991/978-94-6463-010-7_60

2 Research Progress

Modern financial investment theory is mainly composed of portfolio theory, capital asset pricing model, arbitrage pricing theory, efficient market hypothesis, option pricing theory and behavioral finance theory. The development of these theories has greatly changed the traditional investment management practice, which relied mainly on fundamental analysis in the past, and made modern investment management increasingly develop in the direction of systematization, scientification and portfolioi-zation [1]. In 1952, Markowitz published his paper "Portfolio Selection" in The Journal of Finance (the top academic journal of finance), which initiated the modern theory of portfolio management. Markowitz pioneered the introduction of mean and variance to quantitatively characterize the return and risk of equity investments (considered the originator of quantitative trading strategies) and established the basic model for determining the optimal portfolio of assets. Subsequently, a growing number of economic and financial scholars have adopted quantitative models as well as peripheral market and investment trading problems.

In 1976, Stephen Ross proposed the Arbitrage Pricing Theory (APT) to address the shortcomings of the CAPM model in terms of untestability. Based on a multi-factor model of the yield formation process, the Arbitrage Pricing Theory argues that security yields are linearly related to a set of factors that represent some of the fundamental factors of security yields. Subsequently, researchers have discovered various types of factors affecting security returns from different perspectives, and the more classic ones include the Fama-French three-factor model. In 1997, Carhart argued that the study of stock returns should be based on Fama and French's three-factor model by adding momentum effects to construct a four-factor model. Although the four-factor model relates stock returns to the price itself, it has little to do with firm value [2]. In 2013, Asness, a student of Fama, quantified the "quality" of a company and proposed a five-factor model. Later, researchers found that the actual returns of low-volatility (low-beta) stock portfolios were higher than those of high-volatility (high-beta) stock portfolios, a phenomenon that could not be explained by the five-factor model, so Falarelli et al. introduced the volatility factor in 2013 to create a six-factor model.

3 Multi-factor Quantitative Stock Picking Model

3.1 Factor Selection

The first step of multi-factor stock selection model is to discover various types of factors related to stock returns. The selection of factors is mainly based on economic logic and market experience, based on classical market-wide common factors such as size, valuation, momentum, volatility, etc., and combined with various idiosyncratic factors to construct portfolios based on macro, industry, company fundamentals and market characteristics. The factors that affect stock price returns are diverse and subject to opinion. Referring to the summary of brokerage research reports, there are the following categories: overall market, valuation factor, growth factor, profitability factor, momentum reversal factor, delivery factor, size factor, stock price volatility factor, and analyst forecast factor [6].

3.2 Test of Factor Validity

The general test method mainly uses the sorting method to test the effectiveness of the candidate factors in stock selection. For example: it can be tested monthly. Specifically, for any candidate factor, the size of the factor of each normally traded stock in the market is calculated at the beginning of the first month of the model formation period, and the sample stocks are analyzed in ascending order. Sort them and divide them evenly into N combinations and hold them until the end of the month. At the beginning of the next month, rebuild N combinations in the same way and hold them until the end of the month, and repeat until the end of the model formation period. Another parameter is the number of candidate combinations. The optimal choice of specific parameters needs to be tested with historical data.

3.3 Elimination of Redundant Factors

Different stock selection factors may have a high consistency in stock composition and returns due to the same intrinsic drivers, so some of them need to be eliminated as redundant factors, and only the one with the best returns and highest differentiation among similar factors should be retained. For example, there is a clear correlation between volume and liquidity indicators. The higher the liquidity, the higher the volume, so in the stock selection model, only one of these two factors is selected. The method of redundancy factor elimination: Assuming that K valid factors need to be selected with a total sample period of M months, the specific steps of redundancy factor elimination are shown below.

- First, the N portfolios under different factors are scored, and the score is related to the return of the portfolio in the whole model formation period, the larger the return, the higher the score;
- Calculate the correlation matrix between the scores of different factors for individual stocks on a monthly basis;
- After calculating the monthly factor score correlation matrix, calculate the average of the correlation matrix for the entire sample period;
- Set a score correlation threshold, and keep only those factors corresponding to elements in the score correlation average matrix that are greater than this threshold that are less correlated with other factors and more valid, while other factors are eliminated as redundant factors.

3.4 Judgment Method of Multi-factor Stock Selection

The judgment method of multi-factor stock selection is divided into regression method (OLS) and scoring method. The regression method is to use the stock's historical return to regress the screened multi-factor, estimate the regression equation coefficients, and then bring the latest factor into the regression equation to estimate the stock's future return as the basis for stock selection [3]. The scoring method involves scoring stocks based on the magnitude of each factor, then weighting them according to certain weights to obtain an overall score, and screening stocks based on the overall score. Finally, stocks are ranked according to the composite average score derived from the model, and then the top-ranked stocks are selected as needed.

4 Empirical Analysis Study

4.1 Establishment of the Model and Its Parameters

4.1.1 Model Building

The construction of a multi-factor quantitative stock selection model begins with scoring each effective factor on individual stocks. This process we generally take a ranking approach. The scores of all stocks i at the beginning of position building are ranked according to the size of different factors k, from 1 to n. To facilitate the calculation and to make the empirical results easy to observe, the scores are divided by n, so that the scores of individual stocks are distributed in the interval [1/n, 1]. The score of stock selection indicator k of stock i is recorded as $Z_{i,k}$. In this paper, the Z scores of individual stock selection indicators are aggregated into a multi-factor composite Z score using the equal weighting method. The multi-factor equal-weighted composite Z score is:

$$Z_i = (\frac{1}{K}) \sum_{k=1}^{k=K} Z_{i,k}$$

The frequency of portfolio adjustment is determined by ranking all stocks according to the total score Z_i and selecting a number of stocks with the highest scores into the portfolio.

4.1.2 Portfolio Stock Selection

Evans, Archer (1968) proposed a regression measure of risk diversification using the standard deviation of a portfolio of stocks as a proxy for diversification and the number of stocks in the portfolio as a proxy for portfolio size. The specific methods are as follows:

(1) The standard deviation of the stock portfolio is used as the dependent variable to regress the inverse of the portfolio size k. The regression coefficients $\alpha$and β are estimated, and the regression equation is:

$$S_p^k = \alpha + \beta * k^{-1} + \boldsymbol{\varepsilon_k}$$

where: S_p^k is the standard deviation of the stock portfolio, k is the number of stocks included in the portfolio, a represents non-diversifiable risk, and $\beta * k^{-1}$ represents diversifiable risk.

(2) For different portfolio sizes k, the smaller the value of residual diversifiable risk as a percentage of $\beta * k^{-1}/\alpha$, the smaller the unsystematic risk of such a portfolio.

For different stock selection factors, the size of the optimal portfolio may be different. The factors that will be used in this paper are the validity-tested and non-redundant factors in the previous chapter: earnings yield (EP) book-to-market ratio (BM), cash yield (CR) ROA change, PEG, and turnover rate change. The selected stock pool is the sample stocks in the CSI 300 index (CSI 300 index is typically representative), and the selected empirical period is January 2010–December 2013 out-of-sample testing period totaling 48 periods [4]. We try to find out the optimal portfolio size from the perspective of risk diversification based on Evans, Archer's theory, which is suitable for the portfolio composed of the effective factors involved in this paper.

For the screened effective factors: earnings yield (EP), book-to-market ratio (BM), cash return (CR), ROA change, PEG, and turnover rate change, all stocks in the pool are scored and summed at the beginning of each month according to the method described in the previous section, and then ranked in descending order of total score. The portfolio index is used to measure the change of the portfolio's return. The portfolio is constructed using the top K stocks: K = 1 portfolio size 1, K = 2 portfolio size 2,… The maximum portfolio size is 300; a total of 300 portfolios are constructed. The standard deviation of the monthly logarithmic return series of each portfolio index is found, and then the standard deviation of the portfolio and the inverse of the portfolio size are regressed to obtain the estimation equation of the standard deviation of the portfolio and the portfolio size.

$$S_p^k = 0.0255 + 0.0137 * k^{-1} + \boldsymbol{\varepsilon}_k$$

where K = 1.2,… 300, 0.0255 in the equation represents the non-diversifiable risk and $0.0137*k^{-1}$ represents the diversifiable risk. For different portfolio sizes K, the smaller the residual diversifiable risk ratio is, the better the risk diversification is $\frac{0.0137}{0.0255k}$. Figure 1 shows the basic residual diversifiable risk ratio for portfolios with portfolio size between [0.300]. It can be seen from the figure that the residual diversifiable risk ratio decreases steeply at the beginning as the portfolio size increases, and then decreases at a slower rate when the portfolio size reaches a certain size.

Therefore, when the size of the portfolio reaches 58 stocks, the residual diversifiable risk ratio of the comprehensive portfolio decreases to within 0.01, after which the residual diversifiable risk ratio decreases at a significantly slower rate; moreover, in foreign countries, the positions of investment funds are generally controlled at about 60 stocks, of which 50 stocks are used as long positions [7]. Therefore, in this paper, it is appropriate to build a portfolio of 6 effective factors with 50 stocks in the A-share market, which can effectively diversify the unsystematic risk, reduce the transaction cost and improve the portfolio return.

4.2 Portfolio Construction and Testing

This paper uses data from January 2017 to December 2020 for an additional test period of 48 periods. The CSI 300 is used as the benchmark, where the number of up months is 20 and the number of down months is 28. The construction process is the same for all portfolios, except for the different effective factors in the portfolio and the calculation of the composite score, which is as follows:

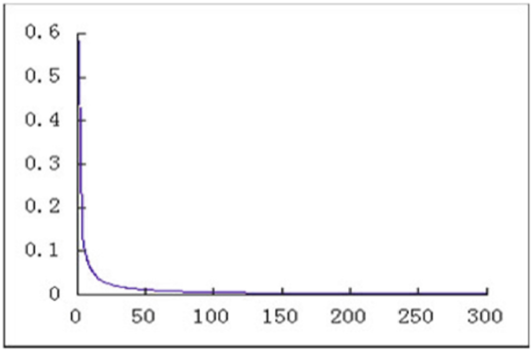

Fig. 1. Percentage of residual diversified risk of different size stock portfolios

- At the beginning of each month, all normally traded stocks are scored according to the scoring method on the stock selection date, and the score of factor k for stock i is recorded as $Z_{i,k}$.
- Sum the scores of all factors in the portfolio according to formula a to obtain the composite score Z_i for stock i.
- Based on the composite score Z_i, all stocks in the market are ranked and the 50 stocks with the highest scores are selected to enter the portfolio.
- Invest $1000 in this portfolio and CSI 300 index respectively at the beginning of the period. The net value is counted at the end of each month.
- Test the validity of the stock selection model. The effectiveness of the stock selection model will be examined in terms of return, risk index and frequency of beating the market benchmark.

4.2.1 Single-Factor Model Stock Selection Ability Test

Value-based portfolio: In this paper, the value-based factors that pass the validity test are earnings yield (EP), book-to-market ratio (BM) and cash return (CR), so K = 3 in the equation to obtain the comprehensive scoring formula of the value-based portfolio.

$$Z = \frac{1}{3}Z_{i,EP} + \frac{1}{3}Z_{i,BM} + \frac{1}{3}Z_{i,CR}$$

Growth portfolio: In this paper, the value-based factors that pass the validity test are ROA change and PEG, so K = 2 in the equation, and the comprehensive scoring formula of the value-based portfolio is obtained.

$$Z = \frac{1}{2}Z_{i,ROAC} + \frac{1}{2}Z_{i,PEG}$$

Technical indicators portfolio: In the technical indicators factor screening, only one valid and non-redundant factor is obtained, so it is equivalent to a single-factor model, and its composite score is the score of the single factor of change in turnover rate on stock i.

$$Z = Z_{i,CC}$$

Table 1. Single factor stock selection ability test

Various statistical indicators	Value type	Growth type	Technical index type
Cumulative Return (%)	24.9	27.6	27.1
Compound Annualized Return (%)	5.77	6.29	6.15
Annualized Excess Return (%)	13.41	13.94	13.76
β Coefficient	0.88	0.94	1.091
Sharpe Ratio	0.38	0.42	0.42
Information Ratio	0.59	0.65	0.64
Months of outperformance (%)	71	74	69
Outperforming Month (%)	67	78	66
Months outperforming benchmark (%)	73	63	56
Positive Months (%)	57	59	53

The results of the three single factor tests described above are tabulated below. It is clear from Table 1 that all three single-factor portfolios have relatively good returns and also have good performance in terms of the three risk metrics. All exceeded 50% in terms of frequency of beating the benchmark, and all reached a relatively good level. In addition, the value portfolio has a higher frequency of beating the benchmark in down months than the other two portfolios, reflecting the superiority of value investing. The growth portfolio has shown its superiority in recent years, as the price fluctuations of stocks compensate for the time value of capital, and high growth companies are bound to gain more capital. These factors have been shown to be effective in selecting a suitable portfolio for the Chinese A-share market in several ways.

4.2.2 Multi-Factor Model Stock Selection Ability Test

In this paper a has 6 valid and non-redundant factors, so Eq. K = 6.

$$Z = \frac{1}{6}Z_{i,EP} + \frac{1}{6}Z_{i,BM} + \frac{1}{6}Z_{i,CR} + \frac{1}{6}Z_{i,ROAC} + \frac{1}{6}Z_{i,PEG} + \frac{1}{6}Z_{i,CC}$$

Table 2 clearly shows that the annualized return of this model is 7.25%, which is much higher than the CSI 300 index, with an excess return of 14.85%; even if we take the one-year deposit rate of 3.5% as the benchmark, the excess return of this model is nearly 5%. The model outperformed the CSI 300 Index in about 72% of the months, including 75% of the months when the CSI 300 Index was up, and 70% of the months when it was down.

Table 2. Multi-factor stock selection model portfolio return

Various statistical indicators	Numerical value
Cumulative Return (%)	32.2
Compound Annualized Return (%)	7.25
Annualized Excess Return (%)	14.85
Sharpe Ratio	0.52
Months of outperformance (%)	72
Outperforming Month (%)	75
Months outperforming benchmark (%)	70
Positive Months (%)	62

5 Conclusion

The essence of multi-factor quantitative stock picking is based on statistical analysis of historical data to screen stocks by finding those factors that are most correlated with stock returns and constructing a comprehensive stock picking index by combining multiple influencing factors based on arbitrage pricing theory. Any multi-factor stock selection model is time-sensitive, risky and needs to be adjusted and updated by the user according to market conditions, which is the shortcoming of multi-factor stock selection strategy. The biggest advantage of the multi-factor stock selection model as a quantitative investment strategy is that its results are based on objective data and a complete model, which can avoid the interference of traders' personal subjective ideas and has a certain objectivity. The object of multi-factor stock selection research is mainly the factors, so the backtesting and validity testing of single factors is an important part of the whole multi-factor model.

References

1. Deng J (2021) GBDT multi-factor stock selection model based on factor deviation. Softw Guide 20(1):109–112
2. Ding Q (2020) Research on multi-factor quantitative investment strategy for stocks based on principal component analysis. Time Finance 17:74–76
3. Dong X, Chang Y (2019) Multi-factor quantitative stock selection model based on factor IC and performance analysis. J Changchun Univ Technol (Soc Sci Edn) 32(6):82–87
4. Liu N, Zhang Q (2019) Research on multi-factor quantification strategy supporting vector regression. Fujian Comput 35(8):1–4
5. Wang C, Wang X (2021) Research on quantitative investment strategies based on investor sentiment. China Price 3:82–85

6. Zhou C, Xu J (2020) Empirical analysis of pairing factors in Chinese and American stock markets. J Syst Manag 29(3):417–424
7. Zhou H, Tang Q, Yang X (2019) Research on optimization of multi-factor quantitative investment strategy for stocks based on big data. Bus News 24:86–87