

Empirical Analysis of Multi-factor Stock Selection Model Based on Weight Assignment of Momentum and Discrete Degree

Li Li^a, Sui Tian^b

School of Economics, Hangzhou Dianzi University, Hangzhou, China.

^a 928262440@qq.com, ^b tiansui0312@hdu.edu.cn

Abstract. Through the stock selection of a large number of literature, we found that factors contribute differently at different market styles or stages, and the factor weight is different. Therefore, in this paper, the factors were preliminarily screened by IC mean and IR equal weight scoring method and T test, and then the screened factors were tested by correlation test and sorting method to obtain the final factors. After five methods, such as factor equal weight, momentum and dispersion comprehensive average weight, were tested and compared, we found that the method of momentum and dispersion comprehensive average weight on the effect of the stock selection is more excellent than other methods, so we chose the momentum and discrete degree of comprehensive weighted average approach to dynamic weighting of each stock, the last stock through empowerment scoring method selection. In this way, the speed of market change and the impact of industry on factors are better considered, and the te-test effect is better.

Keywords: Momentum; Discrete degree; Dynamic weighting; Multi-factor stock selection; Quantitative investment.

1. Introduction

One of the most widely used models in quantitative investment strategy is the multi-factor model. The basic principle of the multi-factor model is to build the model by using a series of effective factors. Stocks with high scores or expected high risk-adjusted returns are bought, and vice versa. The multi-factor stock selection model was first proposed by Ross's [1] arbitrage pricing theory. Later, Fama et al. [2] proposed the famous three-factor model and pointed out that the differences in stock returns could be explained by market factor, scale factor and value factor. Carhar [3] added momentum factor on this basis and constructed a four-factor model that exceeded the explanatory power of three-factor model. In 2014, Fama et al. [4] put forward a five-factor model to further explain the excess return rate of individual stocks.

At the present research stage, the core differences of different multi-factor models lie in two aspects: one is the different effective factors selected from numerous candidate factors; the other is the different specific methods of constructing multi-factor models based on effective factors. Liang Xiaoying [5] thought that the advantages of the linear regression method is the ability to observe and adjust the sensitivity of the various factors in time, and the drawback is that it's easily affected by outliers and extreme value. Scoring method has the advantage of relatively stable and not easily affected by extreme value, but the weight of each factor in the need for subjective set is relatively difficult to do, which depends on the experience of the researcher. Lu Kaichen et al. [6] believed that the multi-factor scoring model was a stock selection model widely used in the industry, which could obtain relatively robust returns. In factor weighting, factors with higher IC absolute values were given greater weight, but the direct relationship between the weight value and the absolute value of IC was not established. Yue Shuning et al. [7] proposed a multi-factor scoring model using FP-growth association rule algorithm and improved the equal-weight scoring into weighted scoring. The improved FP-growth algorithm was used for factor weighting, which had strong profitability and universality. But continuity analysis cannot be performed. Gao Zhiqiang [8] proposed to use the coefficient β of the significant factor obtained by adjusting Fama-Macbeth as the weight of each factor, and to recalculate the β of the significant factor and rank the stocks in a three-month portfolio cycle, so as to carry out the rotation until the end of the term, achieving good strategic returns. But only the textile, transportation and financial industries were tested back. Dong Xiaobo et al. [9] used

the IC time series of the first 6 months to calculate the IC mean vector and IC covariance matrix to calculate the factor weight of each day, and constructed a multi-factor stock selection model that achieved good returns, but the maximum drawdown was relatively high. Liu Jiaqi et al. [10] used half-life IC weighting to weight important factors. Dynamic weighting method was effective in the test period of the model, and the back-test results were good, but there was no comparison with other models.

Based on this, due to the different factors influence in the market, this strategy is mainly focus on scoring method for objective and timely correction factor weight problems. By comparing the Weight method of momentum and dispersion with the remaining four weighting methods, we finally chose the method of weighted assignment of both momentum and dispersion. The advantage of this method is that it can solve the subjective problem of traditional scoring method in assigning factor weight.

2. Principle of danamic weighted scoring of momentum and dispersion

Momentum is a dynamic IC coefficient, which has two momentum effects. One is the temporal momentum effect, that is, the past performance of a factor can well predict the future returns of a factor; the other is the cross-sectional momentum effect, that is, the current advantage of a factor compared with another factor can always be maintained. In the time period when the two effects reach the optimal balance (here, we take it as N), we weight the inclusion factors with the IC mean of the past N period.

Discrete degree can greatly affect the core of asset pricing. A factor with a large degree of dispersion has the potential to become the dominant, while for small discrete degree factor, it will lose its effect on the choice of stock. Therefore, the dispersion degree is used to carry out the quadratic weighting of the factors. We start by classifying industries and calculate each industry discrete degree of various factors. Then the market dispersion is the mean value of industry dispersion. Finally, the market dispersion of each factor is compared. The factors with smaller dispersion are removed, and the factors with larger dispersion are treated with equal weight.

In the end, we give the weight to the factor after the comprehensive average of the two kinds of weight, and then on this basis, score the corresponding factor of each stock, so that we can get the total score of each stock after the factor weighting. After ranking the stocks according to their scores, we can select the stocks that meet the return criteria. In this way, because the way of calculating weight is changing with time, it can achieve the goal of building dynamic stock selection model.

3. Strategy construction

3.1 Processing of factor data

(1) The establishment of candidate factor database

We selected 115 factors from common indicators related to finance and technology as the initial candidate factor concentration pool, with factor data from the flush MindGo platform.

(2) Screening period and detection period of factors

Factor screening period: July 2005 to June 2013; Factor test period: July 2013 to December 2018.

(3) Exception handling

We adopted the Winsorize method commonly used in the industry, that is, 95% of the data is taken as the critical value M, the data greater than or less than 95% of the data is regarded as outliers, and the value is replaced by 95% of the data.

(4) Standardized treatment

This paper adopts z-score method for standardization.

3.2 Screening of factors

(1) Scoring method

Information coefficient IC is the main index to measure the effectiveness of each factor. The size and direction of IC value reflect the prediction effect of each factor on income. The larger its absolute value is, the more accurate its factor will be in predicting the return of the stock in the next period, indicating that the effect of this factor is better. IR, also called information ratio, is calculated by dividing the mean of excess returns by the standard deviation, in which the situation of excess returns can be seen as the stock selection ability of the factor, and dividing by the standard deviation makes IR's evaluation ability more stable at the same time. Therefore, IR takes into account both kinds of abilities and has good evaluation effect.

During the factor screening period (July 2005 -- June 2013), we obtained the IC sequence mean, standard deviation, IR ratio and the proportion of $IC > 0$ of each factor in the candidate factor bank. We scored according to the absolute value of IC mean and IR mean, and the weight was set as 1:1. A factor can only be selected if its score is greater than 60% of the total score (i.e., greater than 133) and it has passed the T-test (t-threshold is 2).

(2) Removal of redundancy factor

The correlation coefficients between the above selected factors were calculated, and the critical value was 0.5, within which they were regarded as independent.

(3) Tests of factor efficiency

First, we sorted the stocks according to the selected factors, divided the stocks into 10 groups on average, and calculated the weighted average return and annual compound return of each group during the period. Valid factors meet the following quantitative criteria:

① the correlation coefficient between the size of factor and portfolio return rate

There should be a strong correlation between the size of the factor and the return on the portfolio. In other words, the expected return of the portfolio is largely affected by the value of the factor. The relationship is expressed by the following formula:

$$\text{Abs}(\text{Corr}(X_{i,i})) \geq \text{Mincorr}$$

Where, $X_{i,i}$ is the annual compound return rate of the portfolio with sequential value of I , Corr is the correlation coefficient between the two in parentheses, Abs is the absolute value of the value, Mincorr is the minimum range of correlation coefficient set by the model, and its critical value is 0.75.

② the yield on the portfolio

Regardless of the order in which we're building the model, the yield difference between the first and last portfolio should be very large, or at least much higher than the benchmark yield.

(4) Final factor pool

To sum up, the last factors that enter stock selection phase are accounts payable turnover, William index, the weighted dividend yield over the past 12 months, belonging to the parent company shareholders - deducting non-recurring gains and losses net profit year-on-year growth rate, net cash flows /interest-bearing debt from operating activities, year-over-year growth rate of net cash flows from operating activities.

3.3 Stock selection based on dynamic weighting scoring method

(1) Construction of a preliminary stock pool

This paper decided to CSI 300 stocks as our stock selection range.

(2) Factor scoring method for stock selection

This paper mainly considers the direction of factors, so as to sort and score all stocks. The calculation steps are as follows:

① Using the final six factors we have screened out above, we grade the stocks in the market by the following method. First, the relationship between the factor and the stock is judged. If there is a positive correlation, the stock corresponding to the largest factor value gets the highest score, that is, the total number of all stocks in the market. The smallest corresponding stock gets the lowest score of 1, and other stocks also get corresponding scores. If there is a negative correlation, then the stock corresponding to the smallest factor value gets the highest score, which is also the total number of all stocks in the market. The largest factor deserves the lowest score of 1, and so on.

② the score of each factor of each stock is weighted by the comprehensive average of momentum and dispersion to calculate the final score of the stock. According to the size of the final score, the highest score is ranked at the first place, and the lowest score is ranked at the last place, and the top ten stocks are selected as the final stock.

(3) The determination of factor weight

① Equal weight of factors

Under the method of equal-weight weighting of factors, each factor is given the same weight and its sum is 1.

② Static IC weighting method

The IC coefficient of the factor is calculated during the test period and then divided by the mean of all IC coefficients to obtain the weight of the factor. Because here the calculated weight is fixed, including an assumption that the weight of each factor is constant during the test period, we call it as static IC weighting method.

$$w_i = \frac{IC_i}{\sum IC_i}$$

Where IC_i is the IC coefficient of each factor and w_i is the weight of the factor.

③ Momentum weighting method

Compared with static IC weighting method, momentum weighting method has the advantage of dynamic weight transformation, which makes the determination of factor weight more accurate and better prediction effect. The factor weight is the IC mean of this factor in the last N months. N is usually 12 or 24. M is the number of factors, and t is the time of the current period.

From the beginning to the end of the back-test, the weight of each factor in each period is calculated as follows:

$$W_f^t = \frac{\text{mean}(IC_f^t: IC_f^{t-N})}{\sum_{f=1}^m \text{mean}(IC_f^t: IC_f^{t-N})}$$

Where W_f^t represents the weight of f factor at time t, IC_f^t represents the IC value of f factor at time t, IC_f^{t-N} represents the IC value of f factor before a period.

The specific steps are:

- 1) In time t, we obtain the IC value IC_{it} for each factor.
- 2) For the positive factor, if $IC_{it} > 0$, then $w_{it} = IC_{it}$, otherwise $w_{it} = 0$.
- 3) For the negative factor, if $IC_{it} < 0$, then $w_{it} = -IC_{it}$, otherwise $w_{it} = 0$.
- 4) Normalize the weights so that the sum of their weights equals 1.

After a large number of data studies, when N is equal to 12, the effect of all aspects is superior when the period is one month. Therefore, when weighting factors according to momentum, we set the trading period as one month.

④ Dispersion weighting method

Principle of discreteness weighting method: by comparing the discreteness of different factors, the factor with high discreteness is given a larger weight, and the weight with low discreteness will be correspondingly smaller.

The steps to calculate the market dispersion of the category factors are as follows:

Step 1: Subclass factor industry dispersion = the median of the top 20% subclass factors in the industry — the median of subclass factors in the bottom 20% of the industry;

Step 2: Subclass factor market dispersion = the mean value of industry dispersion of subclass factors;

Step 3: In order to make the market dispersion of different subclass factors comparable, we standardized it by subtracting the mean value from the dispersion and then dividing it by the standard deviation.

⑤ Weight method of momentum and dispersion

After each factor is weighted by momentum and dispersion respectively, we synthesize and average the weighted weight of momentum and the weighted weight of dispersion to obtain the average weight of each factor, namely, the weighted weight of momentum and dispersion.

4. Model back testing and analysis

Table 1. Five weighting methods back-test comparison table

Strategy	Strategic return (annualized)	benchmark yield (annualized)	Alpha	Beta	Sharpe Ratio	Maximum drawdown
Equal weight of factors	30.53%	23.49%	0.2	0.38	2.38	9.46%
Static IC weighting method	100.88%	23.49%	0.79	0.94	3.38	18.91%
Momentum weighting method	53.87%	23.49%	0.4	0.55	2.89	13.01%
Dispersion weighting method	72.39%	23.49%	0.55	0.69	3.25	14.06%
Weight method of momentum and dispersion	124.37%	23.49%	1	1.02	3.79	20.38%

In order to evaluate the stock selection performance of this strategy, we back-test the five weighting methods respectively and analyze the five weighting methods in Mind Go platform. The benchmark is set as CSI 300 index, and the back-test time is selected from 2019-01-01 to 2021-07-25. With momentum and dispersion weighting as the final weight of the factor, the stocks are scored according to the weight, and the top ten stocks are selected to constitute the portfolio. In terms of turnover, on the last trading day of each natural month, we calculate the factor value and sell the stocks we hold for that month. On the first trading day of the next natural month, positions are traded according to the top 10 stocks, with equal weight of funds for each stock. The specific back-test results are shown in the table and figures.



Figure 1. Equal weight of factors (left) and Static IC weighting method (right) back-test



Figure 2. Momentum weighting method (left) and Dispersion weighting method (right) back-test



Figure 3. Weight method of momentum and dispersion back-test

In order to comprehensively compare the effects of the five weighting methods, according to the scoring rules of quantitative trading group of Zhejiang Securities Investment Competition, this paper selects the weighted comprehensive score of strategy annual return, win rate, return volatility, maximum drawdown and Sharpe ratio, with weights of 30%, 10%, 20%, 20% and 20% respectively. The results are shown in the table below.

Table 2. Weight method back-test results analysis

Strategy	Strategic return (annualized)	benchmark yield (annualized)	Alpha	Beta	winning probability	Return Volatility	Sharpe Ratio	Maximum drawdown	Total score
Weight method of momentum and dispersion	124.37%	23.49%	1	1.02	53.54%	0.32	3.79	20.38%	1.29
Static IC weighting method	100.88%	23.49%	0.79	0.94	52.73%	0.29	3.38	18.91%	1.13
Dispersion weighting method	72.39%	23.49%	0.55	0.69	52.41%	0.22	3.25	14.06%	0.99
Momentum weighting method	53.87%	23.49%	0.4	0.55	51.61%	0.18	2.89	13.01%	0.85
Equal weight of factors	30.53%	23.49%	0.2	0.38	51%	0.12	2.38	9.46%	0.66

It can be clearly seen that the momentum and dispersion weighted back-test results are generally superior to the other four methods, and the return rate, Alpha and Sharpe are also higher than the other four strategies. However, the maximum retracement is relatively large, indicating that risk control should be added to reduce the maximum retracement.

5. Summary

We take CSI 300 component stocks as the stock pool, and the factors that pass the correlation and validity tests are selected as the final factors. In the comparison of the five weighting methods of factor equal weight, static IC, momentum, dispersion and momentum and dispersion, we choose the more excellent weighting method of factor momentum and dispersion to carry out dynamic weighting for each stock, and finally select stocks through scoring method to buy the top ten stocks. This method better considers the speed of market change and the impact of industry on factors, and has better results than other strategies in terms of return rate, Alpha and Sharpe ratio, indicating that the strategy

is initially effective. However, the maximum retracement is slightly worse than the previous strategy, indicating that risk control needs to be done in the later stage of the strategy.

References

- [1] STEPHEN A R. The arbitrage theory of capital asset pricing [J]. *Journal of Economic Theory*, 1976, 13: 341-360.
- [2] EUGENE F F, KENNETH R F. The cross-section of expected returns on stock returns [J]. *Journal of Finance*, 1992, 47:427-465.
- [3] CARHART M M. On Persistence in Mutual Fund Performance [J]. *Journal of Finance*, 1997, 52(3): 57-82.
- [4] FAMA E F, FRENCH K R. A five-factor asset pricing model [J]. *Journal of Financial Economics*, 2014, 116(1): 1-22.
- [5] Liang Xiaoying. Research on quantitative stock selection method based on multi-factor model [J]. *China Market*, 2021(25): 31-32. DOI:10.13939/j.cnki.zgsc.2021.25.031.
- [6] Lu Kai-chen, Yan Hong-fei, Chen Chong. Research on quantitative investment strategy based on CSI 300 stocks [J]. *Journal of Guangxi Normal University (Natural Science Edition)*, 2019, 37(1); 1-12.
- [7] Yue Shu-ning, Shao Bo, Wang Jian. Research on the selection and weighting of model factors of multi-factor scoring method based on FP-Growth association rule algorithm [J]. *Modern Marketing (Business Edition)*, 2020(03): 88-89. DOI:10.19921/j.cnki.1009-2994.2020.03.055.
- [8] Gao Zhiqiang. Multi-factor quantitative stock selection and empirical research based on A-share market [D]. Xiamen University, 2017.
- [9] Dong Xiaobo, Chang Yuqi. Multi-factor quantitative stock selection model and performance analysis based on factor IC [J]. *Journal of Changchun University of Science and Technology (Social Sciences Edition)*, 2019, 32(06): 82-87.
- [10] Liu Jiaqi, Zhang Jian. Multi-factor stock selection model based on machine learning [J]. *Times Finance*, 2020(17): 99-103.