Multi-Factor Stock Selection Strategies Based on Different Securities Markets

Shuiyingzi Hu[3, †], Pai Peng[1, *, †], Mingwei Xu[2, †]

[1]Department of Economics and Trade, Guangzhou Xinhua, China

[2]Department of Physics, 2023 Brown University, Providence RI, USA

[3]School of Accounting, Chongqing Technology and Business University, China

*Corresponding author: pai_peng@brown.edu

[†]These authors contributed equally

Abstract. In the context of the transformation stage of the epidemic and the 20th National Congress, we selected factors such as Market Capitalization, ROE, EPS, P / E, and Turnover to analyze the trading data of stock markets such as Shanghai and Shenzhen 300 and Shanghai Securities last year. Use quantitative methods to construct multi-factor stock selection strategies. In terms of factor selection, 8 different types of factors were selected with reference to research literature in related fields and research reports of securities companies. Next, the validity of the factors in the factor pool is tested, and then the correlation analysis is carried out on the effective factors that have been screened out. Finally, the method of equal-weight synthesis is used to synthesize a large class of factors for the highly correlated factors for subsequent preparations. strategy. Three effective factors were screened out through IC test and hierarchical method test, and redundant factors were eliminated. Finally, based on the research process and conclusions, this paper puts forward relevant suggestions for quantitative investment.

Keywords: Multifactor Model; IC Value; rank IC; SSE 50; CSI 300; CSI 500.

1. Introduction

Since 2022, the spread of COVID-19 in China has become more complicated. China has strengthened its macroeconomic control and support for the real economy. In the report of the 20th National Congress of the People's Republic of China, a new policy for the medical industry was introduced, and the new formulation of "Medicare, Medical, and Pharmaceutical Coordinated Development and Governance" pointed out the direction for the next step of medical reform. With the introduction and in-depth promotion of a series of national policies in the "Healthy China 2030" goal and the "Fourteenth Five-Year Plan", medical health has risen to the national strategic level. In the process of the digitalization and intelligent transformation of the medical health industry, smart medicine has become the backbone. Under the impact of the epidemic, people's demand for high-quality, convenient, intelligent and efficient medical care services is more urgent. At the same time, it has promoted the research and development and popularization of medical wisdom digital products.

In western countries, the study on quantitative investment has been developed for more than 40 years, while in China, the study on quantitative investment started late. The difference between quantitative investment and traditional investment is that quantitative investment uses high-speed computers carry out programmed transactions, which can grasp opportunities, discover information, respond to market changes and complete transactions in a very short time, and effectively control risks. Excellent quantitative investment is the right investment concept after quantitative research numerical, standardized. In recent years, the stock market in China has fluctuated. Most of the time, the A-share market in China is in a downturn stage. Many investors' actual stock returns cannot reach the expected returns. The multi-factor quantitative stock selection model can comprehensively analyze the stock fundamentals, market transactions and macroeconomic at the same time.

2. Literature Review

In “A theory of market equilibrium under conditions of risk”. The Capital Asset Pricing Model states that systematic risk and expected returns are correlated, and taking on more risk is required to achieve a larger return.

Markowitz used quantitative methods for the first time to measure, subverting the previous situation that there was no quantitative calculation of securities and bonds investment [1]. In order to lay the foundation of quantitative investment theory, Sharpe, Lintner and Mossin's series of developments based on Markowitz theory laid the foundation for the rise of multi-factor theoretical models [2]. The research and application of multi-factor stock selection model is not very mature, but many scholars have begun to build a multi-factor stock research model suitable for the Chinese market. Wu Shinong et al. took listed companies in Shenzhen and Shanghai markets as data objects for a period of time. Comparing with different models, the three-factor model is the most suitable for China's stock market. Du Weiwang improves the Fama-French five-factor model to make it more suitable for the operation of China's stock market.In the book, Ding Peng improved the multi-factor stock selection model, using correlation to complete a multi-factor model containing nine effective factors, and through practice, proved that it has good investment returns [3]. In Harnessing Multi-Factor Strategies Close to the Core state that factors that do well over time are also prone to underperforming for extended periods of time, which can be difficult to time. Multi-factor strategies may be more suitable than single factors for investors seeking exposure to factor risk premia but with better diversification and lower cyclicality.

3. Basic Theory

In quantitative investment, the multi-factor model is one of the most widely used models. It states that returns on various securities are affected by various of factors. Multi-factor model chooses the closely related factors for fitting and statistical analysis of historical data to Establish the relationship between Stock Returns and Factors.

The theory of the multi-factor model is mainly constructed by: the combination theory by Markowitz, capital asset pricing model (CAPM) by William Sharpe, and three-factor pricing model by Fama-French.

From the perspective of the development of the factor model, in the classic CAPM theory, the excess return rate of the market is the driving factor for the excess return rate of individual stocks [4-5]. later, Fama and French combined previous studies and found that a three-factor model including market, size and value can more effectively explain the excess return of stocks [6].

3.1 Portfolio Theory

The investment portfolio theory proposed by Markowitz adopts the analysis method of variance "return” and proposes the concept of effective risk-return investment and investment portfolio. The average return rate of the portfolio is the expectation of the above-mentioned investors, expressed by variance. By looking at the variance analysis of the mean asset price, it is possible to determine how much risk investors must take in order to earn returns and the ideal investment portfolio.

3.2 Capital Asset Pricing Model (CAPM)

Capital Asset Pricing Model is the cornerstone of modern financial economy. It states that systematic risk and expected returns are correlated, and taking on more risk is required to achieve a larger return. The model can accurately predict the relationship between asset risk and expected return. This relationship has two important functions:

a. Provide a fund rate of return so that different investments can be evaluated.

b. Make a reasonable estimation of the expected return on potential assets.

3.3 Fama-French

According to a 1992 study by Fama and French, market value, book-to-market ratios, and price-earnings ratios of listed companies are the only variables that can adequately explain the differences in the rates of return of various stocks on the US stock market. The beta value of the stock market is unable to do so. variance in the rate of return. According to Fama and French, the excess return described above compensates for the risk elements that the CAPM fails to take into account.

4. Multi-Factor Stock Selection Strategy

At present, there are already mature research and many successful cases on multi-factor stock selection. However, due to changes in the market itself or lack of cognition by researchers, traditional multi-factor pricing bias exist. In order to improve the power of the multi-factor model, scholars are committed to finding new factor indicators that can explain the excess return of stocks. So far, There are more than one hundred effective pricing factors proposed in the literature [7].

Although the number of effective factors is huge, it is not appropriate to take all factors into consideration at the same time. On the one hand, according to the traditional asset pricing theory, the stochastic pricing kernel is a low-dimensional random variable [8]. Therefore, in the case of high-dimensional factor sets, we need to explore which factors contain more pricing information. On the other hand, according to the theory of limited attention, the risk factors considered by investors when valuing assets are limited [9]. Therefore, how to select the truly explanatory factors from a large number of factors and realize the compression of the model is an urgent need in the factor "big data" era.

4.1 Selection and Introduction to factors

The steps to build a multifactor model are as follows:

1. Select candidate factors; 2. Factor validity test; 3. Eliminate residual factors; 4. Comprehensive scoring model construction.

4.1.1 Fundamental factors

First, use fundamental factors to build a scoring model, and select effective factors by industry for stock selection scoring, select stocks with higher scores to form a stock selection stock pool, and update the stock pool monthly.

At present, machine learning methods are very popular as the core algorithm of multi-factor stock selection models. However, most of the fundamental factors come from annual corporate financial data, which is characterized by low update frequency and small amount of data.

Therefore, the fundamental factor decided in factor scoring model, although it is relatively traditional, it has strong stability and requires a small amount of data.

Although the fundamental factor is suitable for predicting the long-term trend of stock prices, it cannot give an accurate time length and time interval and has a certain degree of instability. And because it relies on financial data, it has a lag.

4.1.2 Other factors

(1) Market capitalization is an important indicator for investors to choose investment varieties. The stoch; on the market is divided into large market, medium cap, small cap, etc., according to the market value, the total market value = the stock price of the day x the total share capital of the day.

(2) EPS is known as earnings per share or profit per share, refers t2the open market, each share to investors or shareholders of the earnings. For businesses with stock trading on the public market, the earnings per share and share price are somewhat correlated. As a result, one of the most important factors used by both current and potential investors to gauge a company's earnings is its earnings indicator.

(3) ROE, also known as return on equity, is the relationship between net profit and average net assets, which can show the rate of return on owner equity and gauge how well the business can use its own capital to generate net profit.

(4) One of the most popular metrics for determining if the stock price level is appropriate is the P/E ratio. That is to say, the genersl assessment of whether the stock value is overvalued or undervalued is derived by splitting the stock price by the share's yearly profits.

(5) Turnover, presented as a percentage, is the proportion of all shares traded during a certain year. Refers to how often stock is traded in the market over a specific time period, which can reflect the strength of stock liquidity, and is also among the most significant technical metrics used to measure market activity.

4.2 Test and Standard of Valid Factors

4.2.1 Correlations between multiple single factors

After an alpha factor is well known by the market and becomes effective, it will become a beta factor and be used for asset pricing to explain portfolio returns. However, only one factor is not sufficient to explain the portfolio return, because its return is affected by multiple factors and is the result of multiple factors, which requires multiple factors as stock selection criteria. There will be a complex interaction effect among multiple factors, which is called "correlation", and the test results simply superimposed with multiple factors may have errors [6]. There is no mechanical "1+1=2" relationship between the performance of a single factor and the multi-factor model. A single-factor test does not perform well, but it may also play its role in a multi-factor model. A test effect good factors may perform poorly in a multifactor model. Therefore, when building a multi-factor model, it is necessary to consider the linkage between factors and do more tests, to better select factors and assets, and then improve the income of the entire multi-factor model.

4.2.2 Information coefficient

The IC value is short for information coefficient, which is the person correlation coefficient between the t-period factor of the stock and the t-1 period return rate of the stock and is used to detect the stock selection ability of the factor. Correlation reflects the comprehensive ability of factors to contribute to the next period's rate of return [9]. Usually, the IC value will be between -1 and 1. The larger the IC value, the better the stock selection ability of the factor. A positive IC value means that the factor has a positive correlation with stock returns, and a negative IC value means that the factor has a negative correlation with stock returns. Use the IC method to extract effective factors, and then build a scoring model based on the effective factors to form a candidate stock pool.

The usual practice of IC analysis is to take a period as the rebalancing period and calculate the factor exposure (factor value) of this period and the corresponding return of the next period. For example, if 20 trading days are taken as one period, then the next period will be the next 20 trading days. In this way, there are 20 factor values in the current period (if Rank is the factor value ranking) and 20 corresponding returns (or rankings) in the next period, a correlation coefficient is calculated for these two series, and this coefficient is the IC of the period. From a certain historical time, point to the present, there are multiple rebalancing cycles, and there are multiple IC values. From this, the average value of IC is calculated to measure the effectiveness of the factor for the entire past time period.

The IC value is between -1 and 1, reflecting the predictive ability of the factor on the return of the next period. The higher the IC value, the stronger the predictive ability of the factor on the stock return in this period. When IC is 1, it means the stocks selected by this factor, and the stocks with the highest scores will have the largest increase in the next rebalancing cycle. If the IC value of a factor is >0.05, it can be regarded as an effective factor. When the IC value is >0.1, the factor can be considered as a particularly good alpha factor; when the mean IC value is close to 0, it can be regarded as an invalid factor.

4.3 Results

Table 1. SSE 50

Turnover_ratio	1	0.16	-0.35	-0.3	-0.12	-0.093	0.09	0.16
Pe_ratio	0.16	1	-0.094	-0.082	-0.094	-0.21	0.14	-0.04
Market_cap	-0.35	-0.094	1	0.86	0.086	0.099	-0.038	0.3
capitalization	-0.3	-0.082	0.86	1	-0.12	0.045	-0.045	0.45
eps	-0.12	-0.094	0.086	-0.12	1	0.55	0.036	-0.15
roe	-0.093	-0.21	0.099	0.045	0.55	1	0.19	0.0063
Net profit year on year	0.09	0.14	-0.038	-0.045	0.036	0.19	1	-0.013
volume	0.16	-0.04	0.3	0.45	-0.15	0.0063	-0.013	1

Table 2. CSI 300

Turnover_ratio	1	0.027	-0.15	-0.13	-0.078	-0.0092	0.038	0.32
Pe_ratio	0.027	1	-0.02	-0.019	-0.025	-0.015	-0.036	-0.013
Market_cap	-0.15	-0.02	1	0.88	-0.17	-0.089	-0.0061	0.3
capitalization	-0.13	-0.019	0.88	1	-0.019	0.025	-0.0089	0.37
eps	-0.078	-0.025	-0.17	-0.019	1	0.56	0.082	-0.059
roe	-0.0092	-0.015	-0.089	0.025	0.56	1	0.2	-0.013
Net profit year on year	0.038	-0.036	-0.0061	-0.0089	0.082	0.2	1	0.037
volume	0.32	-0.013	0.3	0.37	-0.059	-0.013	0.037	1

Table 3. CSI 500

Turnover_ratio	1	0.018	0.04	-0.084	-0.0032	0.0044	0.017	0.61
Pe_ratio	0.018	1	0.0097	-0.001	-0.029	-0.028	0.025	-0.0011
Market_cap	0.04	0.0097	1	0.3	0.24	0.26	0.065	0.2
capitalization	-0.084	-0.001	0.3	1	-0.21	-0.13	0.0088	0.34
eps	-0.0032	-0.029	0.24	-0.21	1	0.19	0.14	-0.055
roe	0.0044	-0.028	0.26	-0.13	0.19	1	0.19	-0.0066
Net profit year on year	0.017	0.025	0.065	0.0088	0.14	0.19	1	0.021
volume	0.61	-0.0011	0.2	0.34	-0.055	-0.0066	0.021	1

The first step is the correlation analysis between the factors. Mainly to examine the repeatability among factors. The purpose is Analyze the correlation between factors and avoid factors with high correlation from entering the factor combination. Factors with high correlation are for linear models, which often leads to a large error in the model, which reduces the predictive ability of the model.

The magnitude of the value reflects the strength of the correlation, and the larger the value, the stronger the correlation. Here we are using SSE 50 (IH) (see Table 1), CSI 300 (IF) (see Table 2) and CSI 500 (IC) (see Table 3) as the stock pool for analysis [10]. Taking a glance, the diagonal lines are all one, since the correlation with itself must be the strongest.

Firstly, from the perspective of the Shanghai Stock Exchange 50 Index (SSE50), the most correlated factor is equity and trading volume, reaching 0.86, followed by eps and roe, which is 0.55; Secondly from the perspective of the Shanghai and Shenzhen 300 Index (CSI 300), it is similar to the Shanghai Stock Exchange 50 (SSE50), but the correlation between turnover rate and market value is reduced; lastly, from the perspective of Shanghai and Shenzhen 500 Index (CSI 500), similar to the other two indices, the correlation between turnover rate and market capitalization is further reduced. The above is the mean value of the correlation.

The stock selection ability of a factor is judged by the IC of the rate of return of the factor, and IC is also a correlation coefficient. The difference from the above is that the profitability of the factor is

examined here. So, the higher the coefficient the better At first glance, the colors of this graph are relatively close, indicating that they have similar abilities in this indicator First of all, we show the average IC of each factor in the whole market and the three major stock indexes. The average IC value is above ±1%, which shows that the explanatory power of the factors is good. Among them, the turnover rate has the strongest stock selection ability, followed by the trading volume factor, and the equity capitalization factor has the worst effect among all factors.

5. Conclusion

The most correlated factors are stock capital and trading volume, and the average IC values are above ±1%, indicating that the explanatory power of the factors is good. Only B/m and eps passed the final group sort method. For the weight of the multi-factor model, we use equal weight and linear addition here, without considering different weight coefficients. Under the screening conditions of the above factors, a high-yield solution has not been found, and more factors need to be tested. Lastly, the report based its conclusions on historical price in 2022 and statistical laws. However, the secondary market is affected by immediate policies, so the report's conclusions may not always be accurate.

References

[1] Markowitz. Portfolio Selection [J]. Journal of Finance. March 1952, 7(1): 77-91.

[2] William F. Sharp. Capital Asset Prices: A Theory of Market Equilib brium under Conditions of Risk [J]. Journal of Finance, 1964, 19:425-42.

[3] Ding Peng. Quantitative Investment: strategy and Technology [M] .Electronic Industry Press, 2012.

[4] Sharpe W F. Capitial asset price: A theory of market equilibrium under conditions of risk. The journal of finance, 1964, (3): 425-442.

[5] LINTNER. Security Prices,risk,and maximal gains from diversification[J]. Thejournaloffinance, 1965, (4): 587-615.

[6] Cochrane J H. Presidential Address: Discount Rates[J]. The Journal of Finance, 2011, (4): 1047-1108.

[7] Mossin J. Equilibrium In a capital asset market[J]. Econometrica: Journal of Theeconometric society, 1966, (4): 768-783.

[8] Lucas R E. Asset Prices in an Exchange Economy. Econometrica, 1978, 46: 1429-1446.

[9] Andrew Innes, Jingwen Shi, Rupert Watts .Harnessing Multi-Factor Strategies Close to the Core.

[10] Li Haiting. Research on quantitative investment of new energy vehicle sector based on multifactor model [D]. East China Normal University, 2022.