Quantitative Multi-factor Stock Selection Strategy Based on Internet Service Industry

Yue Zhuang *

School of Economics, Fujian Normal University, Fujian, China

* Corresponding Author Email: 101112020087@student.fjnu.edu.cn

Abstract. As a matter of fact, with regard to stock investment practice, quantitative strategies have become popular with the rapid development of China's capital market. Plenty of quantitative strategies have been proposed and implemented into the Chinese markets, where multi-factor stock selection is one of them. Based on the multi-factor model commonly used in quantitative investment, this article selects relevant data of the Internet service industry in the A-share market from 2018 to 2022, conducts an industry-level empirical test on 15 factors in 5 categories that are widely used, and constructs investments accordingly combination. According to the analysis, the research found that the stock portfolio constructed based on the above method can outperform the market, far exceeding the performance of the CSI 300 during the same period, indicating that the multi-factor model is still effective in the Internet service industry, and this industry has high investment value. Overall, these results shed light on guiding further exploration of quantitative analysis of stock market.

Keywords: Internet service industry; quantitative investment; multi-factor stock selection model.

1. Introduction

Quantitative trading was born in the United States and has a history of nearly sixty years. In the 1980s, James Simons founded the Medallion Fund, which adopted quantitative high-frequency and multi-trading strategies, and achieved an average annualized rate of return of 35% within 20 years of its establishment, far exceeding the return on investment of traditional financial instruments. At present, the operation of foreign quantitative funds has been relatively mature. As of the end of November 2021, the total global investment scale of quantitative funds has exceeded 200 billion yuan. The trading strategies included in it involve quantitative stock selection, quantitative timing, position management and stop in terms of profit and stop loss, etc. It covers almost the entire investment process of investors, and has become the mainstream investment tool of asset management companies around the world.As the domestic capital market is more open to overseas,the concept and strategy of quantitative investment have begun to flourish in China.At the same time,the favorable situation of the increasing expansion of the A-share market also created a good environment for its development.On February 1,2023, the reform of the registration system of Chinese stock issuance market was fullyimplemented.The improvement of the efficiency of corporate listing and issuance made the stock market more and more powerful. Therefore, how to choose stocks with great investment value and high growth from many listed companies for investment and gaining profitability has become one of hottest issues widely concerned by investors. Compared with the traditional empirical stock selection method, quantitative investment has the advantages of discipline, system, timeliness and decentralization. It can better adapt to this unprecedented situation, overcome the irrational behavior of investors, and obtain relatively stable excess income. Among them, the multi-factor stock selection model is the most commonly used quantitative stock selection method by institutional investors in Chinese country's securities market. Its basic principle is to use various effective factors that affect the stock return rate as stock selection indicators, and buy stocks that meet the requirements at the same time sell non-compliant stocks.For the inefficient or weakly efficient A-share market, information asymmetry leads to many price deviations in the market, which provides a lot of opportunities for using multi-factor model sets to obtain Alpha returns [1].

The Internet service industry is an emerging industrythat provides Internet-related services to users based on the Internet technology, and is the product of the combined development of the Internet and the information content service industry. In recent years, benefiting from the accelerated advancement of digital technologies such as big data, artificial intelligence and cloud computing, the Internet service industrycontinues to develop well, with a growth rate higher than the market average, and has become one of the most active industries in the market today. Large fixed investment, high operating risk, poor sustainability and stability, and high profit level are the distinctive features of this industry: the industry's product replacement speed is very fast, and enterprises need to invest a lot of money to innovation, which is called"blood-making". Once a new product is developed, it will be produced on a large scale in order to quickly occupy the user market to form a first-mover advantage and realize a "winner takes all" situation. Due to the information attribute of the product, the marginal cost of production is almost negligible. If people can find a company with investment value in this industry and choose the right time to invest, there is a certain chance of obtaining excess returns.Hence, more and more investors believe that this industry has a broad market prospect and the significance of asset allocation, and is an important sector for stock investment. This article attempts to combine fundamental analysis with technical analysis, construct a quantitative stock selection strategy based on a multi-factor model with industry characteristics, and apply quantitative analysis methods to the research of the Internet service industry, in order to provide investors with suggestions and references , which has certain practical research significance.

The study of quantitative multi-factor models began overseas. In the early 1960s, "Modern Finance Pioneer" Markowitz revolutionaryly introduced mathematical methods into financial research, quantified risks and returns, established a mean-variance model, and proposed the “Portfolio Theory”that was regarded as a standard in the financial world. In 1964, on the basis of portfolio theory, the famous economist William Sharpe and others discovered the relationship between the expected rate of return and risk of securities under the market equilibrium conditions - the Capital Asset Pricing Model, which demonstrates all stocks’return of the investment portfolio can be decomposed into two parts: systematic risk (that is, market risk) and idiosyncratic return.In the 1970s, investors and financial scholars realized that there were still major shortcomings in explaining stock performance only by market factors, and carried out a large number of researches with the goal of improving these shortcomings.Among them, the economist Stephen Ross proposed the Arbitrage Pricing Theory in view of the shortcomings of the CAPM model, and made outstanding contributions to the formation of the investment strategy construction idea of "multi-factor explanation of stock performance". In the early 1990s, after Fama and French conducted a series of studies on the factors that may determine the difference in the return rates of different stocks in the stock market, they added the two factors of value (HML) and market capitalization (SMB) to the traditional CAPMmodel, the "Fama-French three-factor model", known as the originator of the multi-factor model, was constructed. With the further deepening of the capital market, the power of the Fama-French three-factor model to explain the new financial anomalies in the market gradually weakened, so scholars begun to enrich and improve this model. In 1997, Carhart added the momentum factor (WML) to the three-factor model and formed the Carhart four-factor model; in 2015, Fama and French added the factor RMW representing profitability and the factor CMA representing investmentmode to the three-factor model and established a five-factor model,which had been well verified by using the market data of the United States in the past 50 years. From then on, multi-factor models gradually attracted the attention of all parties. Researchers have explored thousands of factors that can reflect various information about stocks, and frequently apply multi-factor models into quantitative practice.

Compared with the foreign research that started earlier, the research on the multi-factor stock selection model in domestic academic circles started relatively late. After the quantitative multi-factor stock selection model was introduced into China, scholars firstly verified whether the model was still universal in China's capital market. Chen et al. used the relevant data of Chinese A-share market for empirical testing, the results showed that whether industry factors were included or not, the regression results of the multi-factor model were significant, which could explain the common changes in stock returns to a certain extent [2]. Lin used The multi-factor model to predict the return rate of constituent stocks, trying to build a constituent stock portfolio that outperformed the Shanghai and Shenzhen 300

Index [3]. Wang et al. based on the multi-factor model of the regression method for quantitative model stock selection, the empirical results revealed that based on multi-factor quantification, the security portfolio constructed by the stock selection model had a return rate better than the market benchmark return rate [4].

In terms of factor mining, Fan found that adding a price-earnings ratio factor on the basis of the Fama-Franch three-factor model could well explain the market value effect, book-to-market value ratio effect, price-earnings ratio effect and price effect of the stock market [5]; Ding analyzed more than 500 stock factors and constructed two new main factors for stock selection: technical factors and value factors [6]. In addition, Cheng introduced the individual investor sentiment index as a timing strategy factor stock selection model [7].

With the development and application of computer technology in recent years, the current academic research in multi-factor models field mainly focuses on how to use machine learning algorithms to screen effective factors and build stock selection models. Machine learning uses nonlinear methods to find and mine potential relationships between factors and results. Commonly used machine learning stock selection algorithms include neural networks, deep neural networks, support vector machines, random forests, and XGBoost algorithms, etc. [8], which can be more sensitive to capture investment opportunities brought about by unreasonable or irrational factors in the market. Furthermore, in terms of constructing investment portfolios, Li made predictions based on the parameter regression of the logistic model, and utilized the clustering stock selection strategy on the basis of selecting the stock pool by the regression method, so as to construct a reasonable investment that can obtain excess returns portfolio [9]. Wang et al. found that the introduction of the Knockoff method on the basis of Logistic regression was conducive to improving the accuracy of factor selection.Compared with the comparison method, this method could better take the security and profitability of the investment portfolio into account, which has reference significance for optimizing asset allocation [10].

To sum up, more and more scholars are trying to use the indicators in the financial market for investment research, and the multi-factor stock selection model becomes the most widely used model in China with a good momentum of development. At the same time, some scholars began to conduct research from the perspective of industry, and found that there are differences between the model constructed for a certain industry and the model of the whole industry, that is, the effective factors between industries are different, which shows that it is necessary to conduct research at the industry level. Besides, industry characteristics will also affect the effectiveness of the factor.

2. Data and Method

Taking into account the impact of multiple factors on security returns, the multi-factor model is produced. Compared with pricing models such as threefactors, the multi-factor model considers various factors that may affect stock price fluctuations, such as macro, micro and market, which can better realize the complete risk exposure of stock returnsand separation of various influencing factors so as to improve the accuracy of measurement. By this means, it is more helpful for investors in formulating strategies as well. Based on the multi-factor model, the following multiple linear regression equation is established:

$$Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_k X_{ki}+\varepsilon_i \tag{1}$$

Among the equation, Y_i stands for the quarterly return rate of stock i, X_{ji} is the size of the $j-$th factor of the i-th stock, β_j is the regression coefficient of the j-th factor, indicating the sensitivity of the stock return rate to this factor, and ε_i is a random error item .

In this paper, considering the actual situation of Chinese financial market and the specificity of the Internet service industry, combined with the effective factors involved in previous literature, 15 indicators are selected from five categories: growth factor, scale factor, quality factor, value factor and technical factor to construct industry candidate factor bank. According to the industry classification standards announced by the China Securities Regulatory Commission in 2021, the stock pool is established with all individual stocks included in the "Internet service industry" in A-share market. The sample period is set from January 4, 2018 to December 30, 2022, a total of 5 years, the research data are all collected from wind information. In order to improve the accuracy of the regression equation and make the research results more in line with the investment logic, the collected data are preprocessed as follows: Firstly, the stocks of ST and *ST listed companies are excluded, because their market pricing is largelydepends on uncertain factors such as reorganization, which is difficult to predict; secondly, delete the companies that contain missing data and companies listing after 2018, and get 61 stock samples; thirdly, for the individual outliers contained in the remaining individual stock data values and extreme values are winsorized.

Table 1. Validity test results of candidate factors.

category	factor	T-value	Linear relationship
Value factor	P/E ratio	-2.79	Negative
	P/B ratio	-2.80	Negative
	P/S ratio	-4.80	Negative
Quality factor	ROE	0.39	Not obvious
	ROA	-0.03	Not obvious
	Asset-liability ratio	1.42	Not obvious
	Total assets turnover	3.83	Positive
Growth factor	Total assets growth rate	-2.34	Negative
	Operating revenue growth rate	0.94	Not obvious
	Net profit growth rate	-0.34	Not obvious
	ROIC	1.88	Not obvious
Technical factor	Quarterly trading volume	-2.20	Negative
	Quarterly average turnover rate of circulating stocks	-4.00	Negative
	Quarterly average turnover rate of total shares	-3.68	Negative
Size factor	Total market value	-3.68	Negative

3. Results and Discussion

3.1. Factor Screening

This paper uses the multiple linear regression test method to test a single factor, and its advantage is that it can observe and adjust the sensitivity of each factor in time [11]. The specific operation is: carry out the significance test on the candidate factors one by one, that is, use the stock return rate of the next period and the value ofthe factor to be tested in this period for time series regression in each period (every quarter), and the factor with stronger correlation can select stocks better.The specific time series regression equation is as follows:

$$y_{t+1}=\alpha_i+\beta_i X_t+\varepsilon_i \tag{2}$$

Here, y_{t+1} represents the stock return rate in period $t+1$, X_t represents the value of the single factor in period t, β_i represents the corresponding regression coefficient, and ε_i is the random error term. This article uses stata software to obtain the regression coefficient$\beta_i$of each single factor in each period through linear regression, and conducts t-statistic test on the effective value at the 5% significance level (H_0: β=0, $H_1 \neq 0$). The test results are shown in Table 1. According to the calculation results, from 2018 to 2022, among listed companies in Chinese Internet service industry, a total of 9 factors including price-earnings ratio, price-to-book ratio, price-to-sales ratio, total asset turnover rate, and total asset growth rate passed the test, indicating that the above-mentioned single factors are more effective in predicting changes in stock price returns. Generally speaking, the positive factors are mainly quality factors, while the negative factors include value factors, growth factors, technical factors and scale factors. Among them, technical and price factors have a significant negative effect on the quarterly return of individual stocks [12]. Since there is often a strong correlation between the factors that have a high explanatory effect on the rate of stock return, this paper then conducts the Pearson correlation coefficient test on the remaining 9 variables. The result shows that there is a high positive correlation between the quarterly average turnover rate of circulating stocksand the quarterly average turnover rate of total shares (the correlation coefficient value is 0.9312). After comparing the regression test results, the quarterly total share turnover rate is excluded, and the remaining 8 factors are used for the next stepwise regression and model construction.

3.2. Construction of Multi-factor Stock Selection Model

Taking the stock return rate as the dependent variable, the stepwise regression method is used to introduce the remaining eight variables into the model one by one to estimate the stock price return rate of the A-share Internet service industry during 2018-2022. The estimation equation of the model is:

$$Y = 0.0798 - 0.0021PS - 0.8854CSTR + 0.0759TAT - 0.0055PB - 0.0001PE \tag{3}$$

After stepwise regression, the three factors of total asset growth rate, quarterly trading volume and total market capitalization in the model were eliminated because the test was not significant, and the final factor set included price-earnings ratio (PE), price-to-book ratio (PB), price-to-sales ratio (PS), total asset turnover rate (TAT) and quarterly average turnover rate of outstanding shares (CSTR).

3.3. Model Adaptability Testing

Substitute the values of related factors in the first quarter of 2023 into the above regression equation to obtain the estimated return rate of each stock in the Internet services industry in the second quarter. The stocks are sorted according to the size of the estimated value, and the top 10 stocks are selected. And based on these stocks, the investment portfolio for the second quarter of 2023 is constructed. By testing whether this securities portfolio can stably outperform the CSI 300 Index in the market from April 3, 2023 to June 30, 2023, one can test the actual investment effect of the multi-factor quantitative stock selection model established above (seen from Table 2) [13]. It can be seen from Table 2 that the average compound rate of return obtained by the portfolio selected using the multi-factor quantitative stock selection model in the second quarter of 2023 outperformed the benchmark quarterly rate of return of the Shanghai and Shenzhen 300 Index, with an excess rate of return of 18.79%. Therefore, the above-mentioned multi-factor quantitative stock selection model constructed based on the data in the first quarter of 2023 passes the adaptability test of the model.

Table 2. Candidate stock portfolio for the second quarter of 2023.

Stock name	Q2 2023 Yield
Guohua Network Security Technology	54.04%
Montnets Cloud Technology	-7.71%
Dilong Culture Development.	-20.15%
Century Huatong	26.08%
Wangsu Science and Technology	-1.11%
Rastar Group	5.67%
Shunwang Technology	11.25%
Shanghai Ganglian	-20.74%
Everyday Network	11.44%
Ourpalm	80.56%
Average compound rate of return	13.93%
Quarterly benchmark return rate	-4.86%
Excess return rate	18.79%

4. Limitations and Prospects

The research results show that the constituent stock portfolio constructed based on the linear multi-factor stock selection model can not only stably outperform the benchmark, but also exceed the benchmark by 18.79% within 3 months. Therefore, choosing to build a multi-factor quantitative stock selection strategy can bring investors excess returns. However, it is worth noting that since the multi-factor model is built on the premise that the market is invalid or weak form efficiency, it is difficult to find long-term effective factors in real transactions due to the changes of market style, certain factors which performed well in the context of the past marketplace may become invalid, and other new factors may be verified to be effective and added to the model. In the process of using the quantitative multi-factor stock selection model, for the purpose of adapting to the changing market environments, it is crucial to continuously evaluate and improve the selected factors and the model.

5. Conclusion

To sum up, based on the transaction and financial data of the constituent stocks of the Internet service industry in the A-share market, this paper selects candidate factors for analysisand uses the historical market data of relevant stocks from 2018 to 2022 to conduct empirical research. The results indicate that the multi-factor stock selection model is still effective in the Internet service industry, and can predict stocks' return effectively. Besides, stocks in this industry with hightotal asset turnover rate, low price-earnings ratio, low price-to-book ratio, low price-to-sales and low average turnover rate of tradable shareshave higher investment value. In addition, the Internet service industry has obvious investment value, and with reasonable asset allocation, investors can obtain excess returns. In conclusion, in the future, quantitative investment will become more popular in the Chinese capital market, and it is expected to be continuously optimized and developed at the industry level.

References

[1] Gao Y, Yang H, Feng D, et al. Opportunities and Challenges of Quantitative Strategies in the A-share Market. China Collective Economy, 2019, 18: 107-108.

[2] Chen D, Sun C, Shi J. A Multifactor Model of the Securities Market and Its Application in the Shanghai and Shenzhen A-share Markets. Productivity Research, 2009, 21: 112-115.

[3] Lin D, Yang X. Construction of Component Stock Portfolios that outperform the Shanghai and Shenzhen 300 Index: An Empirical Analysis Based on a Multifactor Model. China Business, 2014, 2: 83-84.

[4] Wang C, Liu G, Wang Q. Multi factor quantitative stock selection model and timing strategy. Journal of Northeast University of Finance and Economics, 2018, 5: 81-87.

[5] Fan L, Wang H. Research on the Factors of Stock Return in Shanghai Stock Market. Journal of Management Science, 2003, 1: 60-67.

[6] Ding Q. Research on Multifactor Quantitative Investment Strategies for Stocks Based on Principal Component Analysis. Times Finance, 2020, 17: 74-76.

[7] Cheng H. Design of a Multifactor Quantitative Stock Selection Strategy Based on Individual Investor Emotions. Guangdong University of Technology, 2022.

[8] Ren X. A Study on Quantitative Multifactor Stock Selection Strategies in the A-share Market. Central University of Finance and Economics, 2022.

[9] Li Y. Empirical Study on Quantitative Stock Selection Based on Logistic Regression. Shandong University, 2018.

[10] Wang X, Zhou Y, Tang Ti, et al. Research on Multifactor Quantitative Stock Selection Based on Knockoff Logic. Statistical and Information Forum, 2023, 38(04): 19-32.

[11] Liang X. Research on Quantitative Stock Selection Methods Based on Multifactor Models. Chinese Market, 2021, 25: 31-32.

[12] Zou L. Quantitative Investment Research Based on Multifactor Models. China Business Review, 2021, 24: 100-103.

[13] Sun Y, Li H, Han M. The Application of Multifactor Models in Stock Selection in the Chinese Stock Market. Modern Marketing (Business Edition), 2020, 11: 242-243.