Research on Importance Fluctuation of Stock Selection Factors Based on the Perspective of COVID-19 Epidemic Normalization

Lei Shi[1, *], Yifei Ai[2], Ran Wang[3]

[1] Economics and Management, School of China University of Mining and Technology, Xuzhou, Jiangsu, 221116 China

[2] Finance school of Central University of Finance and Economics, Beijing, 100000 China

[3] Elite school of Harbin University of Commerce, Harbin, Heilongjiang, 150028 China

*Corresponding author. Email: shilei20220325@163.com

Abstract. Since the outbreak of COVID-19 epidemic, the efficiency of the Chinese A-share market has been severely decreased, so it is worthwhile for investors to set up a new stock selection system. This article proposes a novel factor quantitative method based on information entropy and Critic-weight for the Chinese A-share market before and after the normalization of epidemic. Then we use the combination machine learning method including Stochastic Forest, XGBoost, LightGBM, and mutual information to fit the stock return and evaluate factor importance. Under the condition that the model has stability and effectiveness, the results by using paired T-test method show that the impact of volatility factors on stock return is significantly enhanced, and the importance of factors such as scale and leverage is significantly descended.

Keywords: A-share market, normalization of the epidemic, stock selection factor, quantitative weighting, combination machine learning.

1. Introduction

1.1 Significance of topic selection

Since the outbreak of the Covid-19 epidemic in late 2019, it has spread worldwide, posing unprecedented challenges to the world economy. As the barometer of the economy, the stock market will be impacted greatly, which finally reflects in stock return. During the first quarter of 2020, China's GDP has experienced negative growth of 6.8%, and SSEC has plunged 9.83%. Since April 29, 2020, the Covid-19 epidemic has entered the normalization stage, indicating that the stock market is gradually emerging from the era with negative emotion spread. At the moment, it is significant to re-stimulate the vitality of the stock market and improve the confidence of investors, facilitating the fund of capital market financing efficiently.

As we all know, an investor's stock selection strategy will be adjusted to the social-economic background, and the importance of the factors that affect the stock returns will change simultaneously. Therefore, it is of great practical significance to explore and construct a new stock selection system under the new environment. This article focuses on the fluctuation of the importance of stock selection factors in the A-share market after normalization of the epidemic. Through the intuitive reflection of the different importance of stock selection factors between BE(before epidemic) and AE(after the normalization of the epidemic), we know that the Covid-19 epidemic has caused a "sequel effect" on the Chinese A-share market. The influence of some factors concerning stock return becomes more and more important, and investors gradually neglect some factors after the impact of the Covid-19 epidemic.

1.2 Literature review

Before this article, many research scholars have studied the external impact on the stock market, which draws some realistic and instructive conclusions: Xu Hong (2021) [1]proposed the event research method to analyze the overall impact of the Covid-19 epidemic on the stock market in China.

Zhang Yuehua (2021) [2]used the fixed-effect model to study the impact of the return on the manufacturing stock market in China. Li Shuai (2021) [3]compared and improved the three-factor model to analyze the changes in 49 US stock markets before and after the epidemic. However, the current articles focus more on the time points before and after the Covid-19 epidemic or pay more attention to some specific industries. However, the articles that research the influence on the stock market from the perspective of epidemic normalization are few and scattered.

In terms of factor quantification, most of the articles mainly include the following methods: For example, Li Wenying (2009) [4] proposed the hierarchical analysis method for the risk evaluation of engineering projects. Chen Qiao (2006) [5] adopted the AHP method to analyze the quantitative evaluation of the ecological environment of mines. Kerong (2006) [6] proposed the information entropy method to evaluate multiple indicators of five listed industrial companies comprehensively. However, in the traditional articles, the methods mentioned above can not effectively take a variety of quantitative weighting methods into account.

Regarding the importance of factors, the relevant articles are as follows: For example, Zhou Jiahao (2022) [7] put forward a multivariate linear regression method to score 29 characteristic indicators extracted from stock research reports. Cao Wen (2021) [8] proposed a machine learning method combining XGBoost with stochastic forest to construct a multi-factor stock selection model based on the Shanghai and Shenzhen 300 Index. Cao Zhengfeng (2014) [9] proposed a stochastic forest algorithm model to set up the multi-factor model based on value growth. However, in the traditional articles, the relationship between factors and stock return is relatively complex, not a simple linear correlation, and it is challenging to solve the multi-collinearity problem caused by this relationship. Moreover, the authenticity of the conclusions has not been deeply tested in these articles, so the reliability of the model needs to be further examined.

1.3 Article contribution and innovation points

From the new perspective of AE, we propose the method based on information entropy and Critic-weight to quantify the factors of 29 industries in Chinese A-share markets. Secondly, we judge the importance of each factor by the combination machine learning method, which fully considers the valuable information on each factor's contribution to stock return. Finally, we use the paired-T test to examine the importance fluctuation of each stock selection factor BE and AE, showing investors' differences in selecting stocks between two stages.

The innovation of the article can be summarized into two parts: the innovation of empirical topic and the innovation of model methods. The innovation of empirical topics is as follows: Firstly, we try to integrate the staged progress called "normalization of the epidemic" on social hygiene work with the stock return time series. The article explores the influence caused by the Covid-19 epidemic on the Chinese A-share market from a new perspective, "BE and AE," not only limited on the time point before and after the epidemic outbreak. This research has a more realistic guiding significance to the construction of the stock selection system by investors currently.

The innovation of model methods is in the following three aspects: Firstly, to quantify and empower stock selection factors from the original data, we use the combination method based on the information entropy and Critic-weight efficiently, considering the correlation and volatility of the data. Secondly, we propose the method of combination machine learning to fit the quantified factors with a stock return based on different data structures and criteria and calculate the final average factor scores of 29 industries. Finally, from the perspective of statistics, we propose the paired-T test to provide a statistical guarantee for determining fluctuation in the importance of stock selection factors.

2. Theoretical Elaboration

2.1 Factor quantification weight

The entropy method(see Zhang ChunYan (2021) for detail) [10] is a common synthetical evaluation method. The greater the information value E_j of an index is, the smaller the variation of its index value, and the smaller its role in the synthetical evaluation, the smaller its weight should be. And vice versa. The core formula is as follows:

$$W_{1j} = \frac{1-E_j}{n-\sum_{i=1}^{n} E_j} \qquad j=1,2,\ldots\ldots,n$$

Where E_j represents the amount of information contained in the j-measure; W_{1j} represents the weight of the j-measure; n represents the amount of sample.

The Critic-weight method (see Yu HouQiang, Li Ling(2012) for details) [11] is another objective weighting method proposed by Diakoulaki. Its basic idea is to embody the fluctuation intensity and conflict of the evaluation index. The core formula is as follows:

$$W_{2j} = \frac{\sigma_j(1-r_j)}{\sum_{j=1}^{n} \sigma_j(1-r_j)}$$

Where C_j represents the amount of information contained in the j index; W_{2j} is the weight of the j index; r_{ij} evaluates the correlation coefficient between i and j; n represents the number of the experiment sample.

The entropy weight method is based on the probability of assigning the weight to the index according to the variation of the index. It is and considered the possibility of the development of each index. The Critic-weight method is based on the full trust of the existing data, considering the discreteness and conflict of the data in each index. Because of the perfect complementarity between these two methods, the new comprehensive evaluation method can consider the existing characteristics of each index data and the variability of the data, which makes the evaluation of index weight more scientific. The core formulas are as follows:

$$T = \frac{W_{1j}}{\sum W_{1j}} * \frac{W_{2j}}{\sum W_{2j}} \qquad T\_\mathrm{new} = \frac{T}{\sum T}$$

Where T represents the new weight of indexes that combines two methods. T_new is the weight normalized from T.

2.2 Machine learning methods

This article uses the Stochastic Forest Model, XGBoost, and LIGHTGBM, all decision tree-based ensemble learning methods. However, they have different algorithms when they search for the importance of features. Please refer to Cao Wen (2021) [8], Tang Jiazheng (2020) [12], Le Bin (2019) [13], Yu Juqi (2020) [14] for details. In addition, we use a mutual information method, Darbellay G (1999) [15], to measure the interdependence of random variables in the amount of information.

2.3 Modeling process

Using entropy and Critic-weight comprehensive evaluation methods, 42 small factors in each industry stage were combined, weighted, and quantified into the final stock selection factor.

The importance degree of the stock selection factors in different periods of each industry is calculated by using the method of combinatorial machine learning, and the average importance of the stock selection factors of 29 industries is calculated by stages.

Finally, based on statistical significance, we propose whether there is any difference in the importance between before and after using the paired-T test. See Guo Hongfei (2019) [16] for details. According to this article, the T statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Where $\bar{x}_1$ and $\bar{x}_2$ are mean values. S_1 and S_2 are standard errors; n1 and n2 are sample amounts.

When the t statistic is greater than the t critical value, P < 0.05, it rejects the original hypothesis, indicating a significant difference between the importance of factors BE and AE. According to the changes of each stock selection factor, 29 industry data were integrated as a sample. BE and AE have conducted Two stages of double sample tests based on the significance level.

The modeling process of the article is shown in Figure 1:

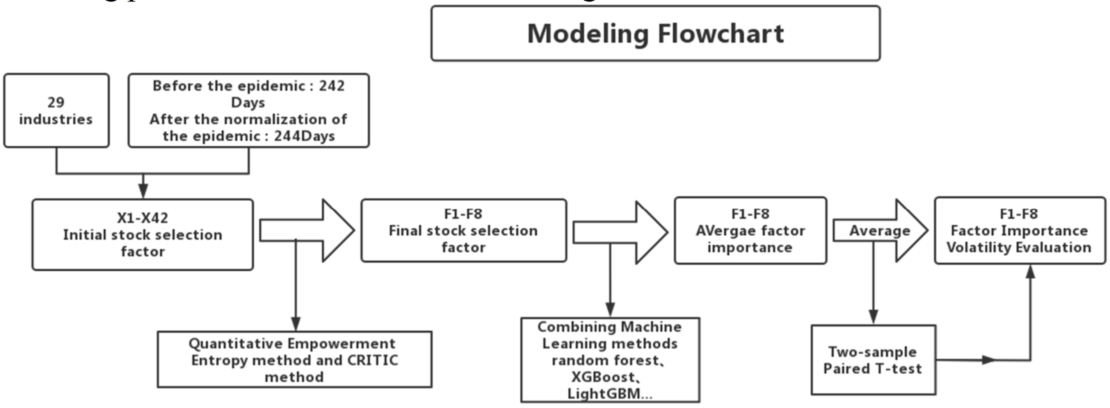

Figure 1 modeling process flowchart

3. Empirical Analysis

3.1 Data sources and indicators description

3.1.1Data sources

The data in this article is selected from the Wind financial terminal. Referring to the classification basis of the securities company CITIC, China's A-share market can be divided into 29 industries such as Construction, Chemical, etc. Based on the time point definition of 《Fighting COVID-19 China Action》, we extracted daily data including P/E, ROA, and other 42 financial indicators of 242 trading days from January 2, 2019, to December 27, 2019, BE and 244 trading days from April 29, 2020, to April 29, 2021, AE for each industry.

3.1.2 Description of indicators

Based on the guiding methods for the classification of stock selection factors in literature, such as Cao Wen (2021)[8] and Xie Mingzhu (2021)[17]. These articles combine the initial factors to construct a factor library. For the scale factor F1, indicators are selected, such as total share capital etc. For the Valuation factor F2, indicators such as price-earnings ratio PE (TTM) etc are selected. For the Sentiment factor F3, we selected indicators such as consensus forecast net profit (FY1) etc. For the Growth factor F4, we selected indicators such as return on equity (diluted) (year-on-year growth rate etc. For the Profitability factor F5, we selected indicators such as Earnings per share EPS (TTM) etc. For the Leverage factor F6, we selected indicators such as gearing ratio etc. For the Volatility factor F7, we selected indicators such as amplitude etc. For the Technical factor F8, we selected indicators such as MA5 daily moving average etc. Finally we use Y to represent the stock return rate.

3.2 Factor quantization results

In the above analysis, we obtain the data needed to research the importance fluctuation of stock selection factors and normalize the whole index with its positivity and negativity. F-factor Quantization Matrices before and after normalization of the Covid-19 epidemic were obtained from 29 industries based on CITIC classification.

The heatmap of the correlation BE, and AE of the stock selection factor F1-F8 are shown in Figure 2:

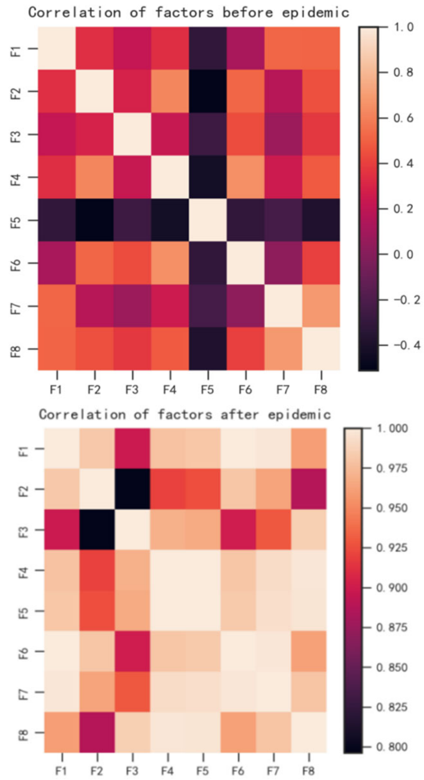

Figure 2 Heatmap of the correlation between stock selection factors BE and AE

From the correlation results, the correlation between the factors BE and AE is very different, which indirectly shows the necessity of comparing the two stages' factors. In addition, the correlation between the factors existed BE and AE, especially after the normalization of the epidemic, the correlation of the factors increased significantly and even had multiple collinearities. It is shown that the influence of factors on stock return is a complex system. When we do not pursue causal inference but only stress the relative importance of factors, we can use a combination machine learning method to research, which can greatly avoid the error caused by the high correlation of factors.

3.3 Factor importance fluctuation result

The average importance of factors for F1-F8 BE and AE are shown in Table 1:

Table 1. Average importance for F1-F8 BE and AE

F	F1	F2	F3	F4
BE	0.0975	0.1094	0.1090	0.0937
AE	0.0874	0.1001	0.0962	0.0987
F	F5	F6	F7	F8
BE	0.0713	0.0384	0.3431	0.1382
AE	0.0634	0.0203	0.3925	0.1408

The average importance of the F1-F8 factor BE and AE is illustrated as shown in Figure 3 (Since the volatility factor F7 is significantly more important than other factors, we scaled it four times smaller so that it can be put in a map with other factors).

For the factor importance distribution, we know that except for F7, the other indicators follow symmetrical distribution, which is basically in line with the actual situation. In addition, it is worthing that importance fluctuation and distribution of some factors whether there are significant difference needs to be judged by other tools. Therefore, we will further analyze the fluctuation of these factors BE and AE by using paired T-test.

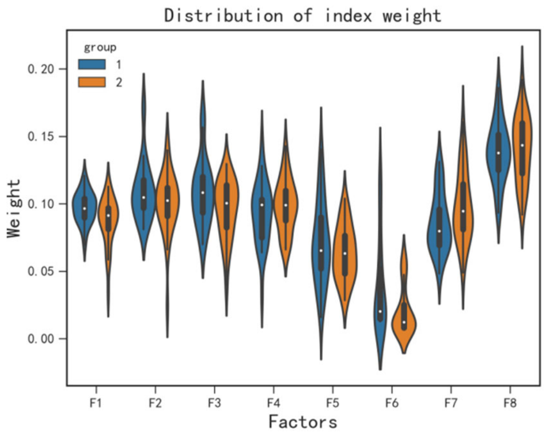

Figure 3 Violin Plot of Average Importance of Stock selection Factors in Two Stages

Finally, we illustrated the importance distribution of the F7 subdivision factor, which are shown in Figure 4 (X1, X2, X3 correspond to fluctuation).

The violin diagram describing the importance of subdivision factors shows a significant upward trend. Therefore, we can conclude an apparent importance promotion between BE and AE. It also confirms the previous view to some extent: the volatility factor of AE has a greater impact on stock return, the importance significantly increased. In the factor importance distribution, the subdivision factors of F7 show evident proper skew distribution, which has a good interpretation effect on the overall skewness of F7, and it accords with the reality.

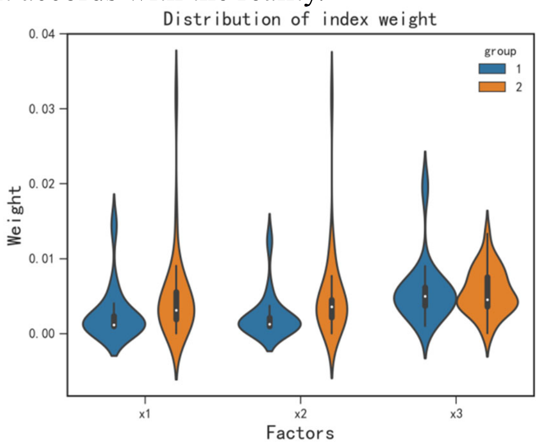

Figure 4 Violin Diagram of Two-Stage Volatility Subdivision Factor Importance

For the results of the paired-T test, the p-value of the volatility factor is less than the given significance level of 0.05, and the pairing difference between BE and AE is negative, indicating that the importance of the volatility factor has increased significantly. The p-value of leverage and the scale factor is less than or equal to the given significance level of 0.01, and the pairing-difference is positive, which indicates that the importance of leverage and scale factor decreases significantly.

T-test data for F1-F8 are shown in Table 2('*' means $p<0.05$; '**' means $p<0.01$):

Table 2. t test data for F1-F8(*mean

F	F1	F2	F3	F4
P	0.006**	0.206	0.017*	0.269
F	F5	F6	F7	F8
P	0.166	0.010**	0.023*	0.542

4. Summary and discussion

4.1 Conclusions and recommendations

In this article, we find that the Covid-19 epidemic makes the importance of multiple factors and investors' preferences change significantly, in which the importance of volatility factor increases substantially. Although the epidemic is getting better, the outbreak of the Covid-19 epidemic has had an important impact on the stock market's stability and investors' mood. Therefore, investors in China's A-share market should pay more attention to the technical indicators such as stock price volatility instead of focusing on company size, leverage, and other fundamental directions. If the epidemic in China disappears completely in the future, we can still re-obtain data and build a more adaptable stock selection system based on the model.

Based on the above conclusions, this article puts forward the following suggestions: Firstly, policymakers should adhere to the stability and effectiveness of fiscal and monetary policies to promote the calm of negative emotions in the post-epidemic period of the securities market. Secondly, the financial regulatory authorities should strengthen the compliance management of the securities market and stabilize systemic risks. Thirdly, financial institutions should actively participate in recovering the capital market in the post-epidemic period and stimulate market vitality under the premise of observing order. Fourthly, investors should focus on the technical indicators such as stock price amplitude in combination with the fluctuation of the importance of factors, take the fundamental indicators of companies into account, and build a new stock selection strategy system.

4.2 Future research directions

Enrich the research connotation at the empirical level by incorporating some unconsidered factors into stock selection factors.

Optimize the model at a deep level. Such as replacing more sensitive machine learning methods, testing the stability of the model based on a more extended period, expanding the capacity of the factor base, and selecting more indicators such as macroeconomic variables into the model.

References

[1] Xu Hong, & Pu Hongxia. (2021). The Impact of the New Crown Epidemic on China's Stock Market. Financial Forum.

[2] Zhang YueHua, Yang Lin&Xue WeiXian.THE IMPACT OF Covid-19 ON THE MANUFACTURING INDUSTRY OF OUR COUNTRY: A POSITIVE STUDY BASED ON STOCK MARKET RETURN.

[3] Li Shuai & Zhang Qiang. (2021). Empirical Research and Regulatory Thinking on the Impact of the New Crown Epidemic on the US Stock Market. Based on Fama-French Three-factor and Five-factor Model. Wuling Academic Journal (05), 48-57.

[4] Li Wenying. (2009). Application of Analytic Hierarchy Process (AHP) in Project Risk Management. Journal of Beijing University of Chemical Technology (Social Science), 2009 (1), 46-48.

[5] Chen Qiao, Hu Ke, Qu Kunli, Li Fulai, & Zhao Wei. (2006). Study of Doctoral Dissertation Based on AHP.

[6] Korong. (2006). Applied Research on Entropy Method in Comprehensive Evaluation of Multiple Indexes of Stock Market. Economic Forum, (14), 107-107.

[7] Zhou Jiahao, Zhang Mingfu, Leng Hongjie. Research on Stock Selection Strategy Based on Machine Learning Multifactor Quantization Model [J]. Science and Technology Innovation. 2022 (05): 161-164

[8] Cao Wen. Empirical Research on Multifactor Stock Selection Based on XGBoost Algorithm. Shandong University of Finance and Economics. 2021 (12)

[9] Cao Zhengfeng. Ji Hong. Xie Bangchang. Selection of Quality Stock by Using Stochastic Forest Algorithm [A]. Journal of Capital University of Economics and Trade. 2014 (02)

[10] Zhang Chunyan, Wu Ronghua, Zhuang Kezhang, Xu Jie, Li Xin, Li Jing... & Qi Xiaofeng. (2021). Comprehensive Evaluation of Different Maize Varieties in Southern Shandong by DTOPSIS Based on Entropy Weighting. China Agronomy Bulletin (23), 8-14.

[11] Yu Houqiang, Li Ling. Comprehensive Evaluation of Five Types of Enterprises Based on Entropy Weight Method and CRITIC Method [J]. Journal of Hubei Institute of Science and Technology, 2012,3212: 83-84.

[12] Tang Jiazheng. Study on quantitative multi-factor hedging model based on data mining and XGBoost algorithm [D]. Shandong University. 2020 (10)

[13] Le Bin, Cai Zhijie, Hu Weichao. Machine learning multi-factor stock selection model based on factor context [J]. Mathematical modeling and its application. 2019 (04): 10-19

[14] Yu ShuQi.Design of Multifactor Stock Selection Trading Strategy Based on Time Varying Weighting LightGBM [D]. Shanghai Normal University. 2020 (07)

[15] Darbellay G.A. An estimator of the mutual information based on a criterion for independence [J]. Computational Statistics & Data Analysis, 1999, 32: 1-17

[16] Guo Hongfei. Application of F Test and T Test in Method Verification [J]. Shanxi Metallurgy, 2019,4204: 114-116.

[17] Xie Mingzhu. (2021). Multi-factor quantization stock selection strategy under machine learning algorithm. Journal of Jilin Business College (06), 90-97.