Proceedings of the 2nd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/10/20230468

Research on the Value of Stock Investment Based on Factor Analysis and Cluster Analysis

Yihang Yuan[1,a,*,†], Yucheng Hou[1,b,†], Zongyang Zhang[2,c,†]

[1]Macau University of Science and Technology, Business college, Macau, China
[2]Bachelor of Commerce Programme with a Major in Finance between Beijing Normal University Zhuhai, China and Saint Mary's University, Canada Zhuhai, China
[3]Beijing Normal University at Zhuhai, Huitong college, Zhuhai, China
a. 1909853db011072@student.must.edu.mo
**corresponding author*
[†]These authors contributed equally

Abstract: This paper mainly explores the important practical significance of stock investment factor analysis. The factor analysis method has been applied in economics, sociology, management, and other research fields, its main purpose is to describe the basic hidden variables hidden in a group of measured variables, which cannot be directly measured. This paper takes 62 listed companies of the North Exchange as examples for analysis. The formation mechanism of a stock price is very complex, which needs to be analyzed by many factors. Several different factors are analyzed by IMO and Bartlett tests and the orthogonal rotation method. Finally, the price factor, trade factor, debt factor, earnings factor and growth factor are determined. Factor scores are then calculated and compared for each stock. Descriptive analysis was used to get the final conclusion. Through the use of factor analysis, investors can make quantitative analysis of each index and extract the factors affecting the stock, so as to make a relatively objective evaluation of the listed company's stock, and provide an evaluation criterion for investors.

Keywords: stock investment factor analysis, IMO and Bartlett tests, orthogonal rotation method, descriptive analysis

1. Introduction

The stock market is called an "economic barometer", and its fluctuations are closely related to the rise and fall of the market economy. However, the formation mechanism of stock prices is very complicated, being affected by many factors. Beijing Stock Exchange was registered and established on September 3, 2021. It is China's first corporate stock exchange approved by the State Council [1]. According to the literature, on December 11, 2021, a total of 82 shares have been listed on the Beijing Stock Exchange [2]. The biggest difference between Beijing Stock Exchange and Shanghai or Shenzhen Stock Exchange lies in the service object. Compared with the other two stock exchanges, Beijing Stock Exchange provides more services to small and medium-sized enterprises. From the perspective of the size of listed companies on the Beijing Stock Exchange, represented by the manufacturing industry, information transmission, software, and information technology services account for the highest proportion [3]. At the same time, the Beijing Stock Exchange expanded the

DOI: 10.54254/2754-1169/10/20230468

rise and fall to 30%. The risks faced by investors should not be underestimated [4]. There are great differences in the investment style and capital size of investors. So different investors have different risk preferences. Some investors are stable, some investors are radical and some investors have more liquidity [5]. Therefore, how to choose an investment plan that is suitable for your own situation is very important. The purpose of this case study is to provide some suggestions for different types of investors by dividing stocks into different categories and summing up their characteristics with the Beijing Stock Exchange as the core.

The factor analysis of stock investment value has practical significance in reality [6]. And the most important significance is to put forward corresponding investment suggestions for investors with different financial strengths and risk preferences through this factor analysis [7]. For example, investors can be divided into stable investors, aggressive investors, and rich and capricious investors. After that matching the characteristics of investors in order to provide the corresponding results of stock investment value factor analysis [8]. Factor investing is to find out the return risk of trade-related factors. Factor analysis based on stock investment value is helpful to improve portfolio results, reduce volatility risk and increase portfolio diversification [9]. In the face of numerous indexes in the financial tables of listed companies, investors can use factor analysis of stock investment value to replace the original indexes with several factor score indexes. These indexes also contain information reflecting the differences between stocks which can help investors have a clear understanding of the financial situation of listed companies [10]. In addition, investors aim to select the listed companies with good economic benefits. A large number of data makes it difficult for investors to analyze the listed companies, so the use of investment value factor analysis will be a good solution to this problem. The research results and models of factor analysis made by Han are used for reference, in order to make the factor analysis made later more reasonable and scientific. [11]

2. Methodology

2.1. Factor Analysis

Prior to doing a factor analysis, the writers examined stock quote data as well as the company's fundamentals. Prior to choosing the data, every available information was examined. Instead of using the amount of the transaction, the average price of a transaction was chosen since it can be more closely correlated with other price factors. This average price is calculated by dividing the volume by the amount of the transaction. 19 variables were selected by the author from the basics and share list. RESSET database is the source of the data. The author of this publication referred to them as Table 1, which is as follows. They'll be put to use next.

Table 1: Variables and meanings.

Variable Name	Meaning	Variable Name	Meaning
Op-pr	open price	ROA	ROA
Hi-pr	high price	Np-r	net profit ratio
Lo-pr	low price	Gp-m	gross profit margin
Cl-pr	close price	Ct-r	current ratio
Wk-at	week amplitude	Qk-r	quick ratio
Tr-ol	trade volume	Na-gr	net asset growth ratio
Me-Pr	mean price	Ta-gr	total asset growth ratio
Da-at	day amplitude	It-to	inventory turnover
Da-tn	day turnover ratio	Da-r	debt asset ratio
ROE	average ROE	-	-

DOI: 10.54254/2754-1169/10/20230468

The factor analysis can then be applied. To start, the KMO and Bartlett tests are run to see if factor analysis can be of any value. For Bartlett's test (p 0.05, strictly p 0.01), if the significance is less than 0.05 or 0.01, rejecting the null hypothesis, suggests that factor analysis can be effective. Verifying the number of factors by examining the variance explained table and the scree diagram The gravel diagram displays the Eigen's downward trend, which can be used to determine how many factors to consider, while the variance explained table indicates the factor's contribution to the explanation of the variable. It is possible to generate a matrix of importance and composition of the hidden variables in each factor by examining the factor load factors and heat plots. Multi-factors are reduced to a few factors based on the factor loading diagram. The component matrix is examined to determine the factor composition formula and weight. Detailed output factor analysis technique score.

Table 2: KMO and Bartlett's test.

KMO test and Bartlett's test		
KMO		0.683
Bartlett Test	Approx. Chi-Square	3014.976
	df	171.000
	P value	0.000***

Note: ***, **, * represent the significance levels of 1%, 5%, and 10%, respectively

The findings of the KMO test and the Bartlett spherical test to determine whether factor analysis is possible are shown in Table 2 above. The Bartlett spherical test findings indicate that the significance P value is 0.000***, which is statistically significant and rejects the null hypothesis, and the KMO test results indicate that the value of KMO is 0.638. There is a relationship between the variables. The factor analysis works well. The table of explained total variance then looks like this.

Table 3: Total variance explained.

Factors	Eigen			% of Variance (Rotated)		
	Eigen	Variance	Cumulative	Eigen	Variance	Cumulative
1	6.062	31.906	31.906	5.263	27.699	27.699
2	4.262	22.430	54.336	3.574	18.809	46.508
3	2.580	13.577	67.913	2.758	14.516	61.024
4	1.687	8.881	76.794	2.393	12.594	73.618
5	1.283	6.751	83.545	1.886	9.926	83.545

Table 3 shows the total variance explained in different numbers of factors. In general, the higher the variance explained rate, the more important the principal component, and the higher the weight ratio. The authors decide to use 5 factors depending on the gravel diagram. Feature root is 1.28 when using 5 factors and that is 0.98 when using 6 factors. It is at a useful and reliable level. At the same time, variance interpretation is 83.545%, which is at a high level (Figure 1).

DOI: 10.54254/2754-1169/10/20230468

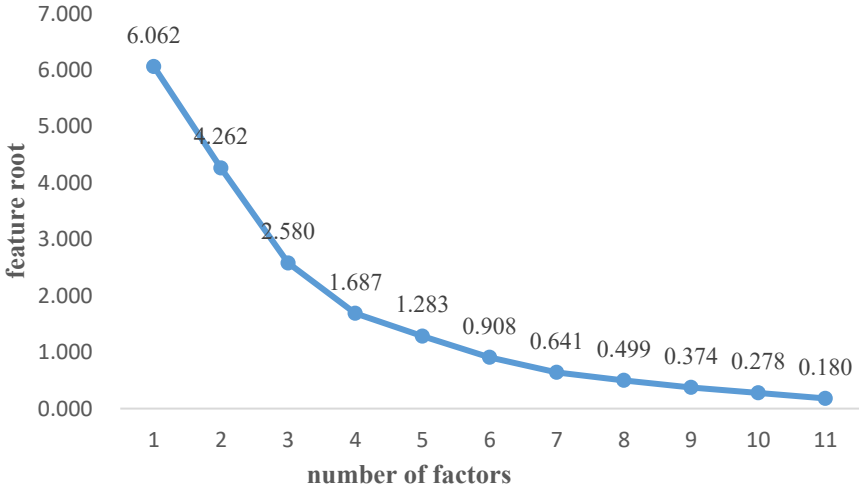


Figure 1: Gravel diagram.

The final factor loading matrix is then generated in this study using the orthogonal rotation method. As shown in figure 2, each factor's characteristics are obtained by employing two colors that represent -1 and 1, respectively.

Da-r	0.326	-0.105	-0.752	0.106	0.082	0.700
It-to	-0.094	0.337	-0.401	-0.115	-0.197	0.335
Ta-gr	0.306	0.180	-0.036	0.121	0.839	0.846
Na-gr	0.129	0.166	0.134	0.234	0.863	0.862
Qk-r	0.012	0.468	0.727	-0.289	-0.009	0.831
Ct-r	0.009	0.466	0.741	-0.264	-0.002	0.836
Gp-m	0.072	-0.126	0.805	-0.056	0.015	0.673
Np-r	-0.023	0.059	-0.089	0.775	0.074	0.617
ROA	0.180	-0.045	-0.125	0.930	0.076	0.921
ROE	0.378	-0.069	-0.208	0.789	0.122	0.828
Da-tn	-0.006	0.906	0.156	-0.014	0.039	0.847
Da-at	-0.035	0.892	0.072	0.082	0.019	0.810
Me-Pr	0.987	-0.025	-0.025	0.096	0.100	0.995
Tr-ol	0.004	0.379	0.169	0.335	-0.462	0.498
Wk-at	-0.074	0.791	-0.043	-0.026	0.232	0.687
Cl-pr	0.987	-0.025	-0.021	0.097	0.106	0.996
Lo-pr	0.987	-0.044	-0.023	0.097	0.099	0.996
Hi-pr	0.987	-0.028	-0.015	0.099	0.103	0.996
Op-pr	0.988	-0.048	-0.019	0.096	0.096	0.996
	factor 1	factor 2	factor 3	factor 4	factor 5	Communality

Figure 2: Heat map of the factor load matrix.

2.2. Determine the Factors

Combining the heat map of the factor loading and the factor loading itself, the individual factors are named as follows:

Price factor: factor 1 is loaded higher on price variables such as Open-price, High-price, Low-price, Close-price, and Mean-price. It can reflect the fluctuation of the stock price. Therefore, in this paper, the authors name it the Price factor.

Trade factor: factor 2 is loaded higher on week amplitude, day amplitude, and day trade turn rate. Day turnover rate and amplitude can reflect the trend of exchange and the stock's popularity. Therefore, in this paper, the authors name it the Trade factor.

Debt factor: factor 3 is loaded higher on gross profit ratio, current ratio, quick ratio, and debt asset ratio. It can reflect liquidity. The current ratio, quick ratio, and gross profit ratio are usually used to

DOI: 10.54254/2754-1169/10/20230468

judge a company's short-term liquidity. And debt asset ratio can reflect the company's long-term liquidity. Therefore, the authors name it the Debt factor.

Earnings factor: factor 4 is loaded higher on ROE, ROA, and net profit margin. It can evaluate a company's profitability. Therefore, in this paper, the authors name it the Earnings factor.

Growth factor: factor 5 is loaded higher on the net asset growth rate, total asset growth rate, and trading volume. The net asset growth rate and total asset growth rate reflect the development of a company. Therefore, in this paper, the authors name it a growth factor.

3. Results and Discussion

3.1. Descriptive Analysis

The first batch of 81 companies listed came from 51 prefecture-level cities. Among them, the largest number of companies listed in Beijing is 11. It is followed by Guangzhou, Shanghai, and Suzhou, which are tied for second, with 4 companies. A lesser-known city Nantong has 3 listing companies the same as Shenzhen, which is included in "Third" in the figure. The "Fourth" in figure 3 includes the cities: Bengbu, Changzhou, Huizhou, Qingdao, Shiyan, and Chongqing. As for the last 38 cities, each of them keeps a listing company.

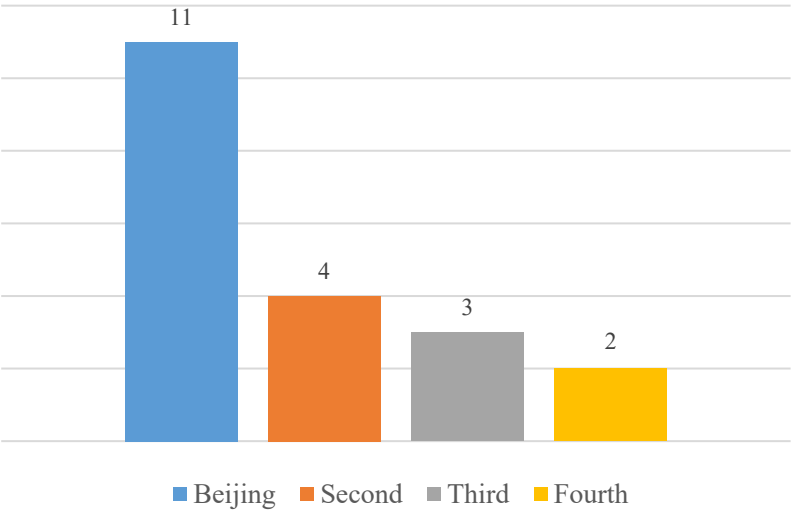


Figure 3: The number of listing companies in different cities.

Software and information technology services are short for SITS. Software and information technology services are what PM is. The special equipment manufacture is called SEM. Instrument manufacture is referred to as IM. The manufacturing of computers, communications, and other electronic devices is known as CEEM. AM stands for auto manufacturing. Rubber and plastic products are known as RP. The manufacturing of electrical machinery and equipment is known as EMAE. Professional and technical services are PATE. NMM stands for Non-Metallic Mineral Products. The initial batch's 81 listed enterprises feature high-tech components in over half of them. They cover the core areas of the country's economy, primarily represented by high-tech manufacturing, contemporary services, advanced manufacturing, high-tech services, and key emerging sectors. These 81 businesses made a combined net profit of 3.815 billion yuan in 2018, 4.609 billion yuan in 2019, and 5.62 billion yuan in 2020, respectively. The covid-19 does not affect the overall increasing trend of these 81 enterprises, and their operating conditions have improved steadily over time. Additionally, the first group of 81 businesses listed following the CSRC's industry

DOI: 10.54254/2754-1169/10/20230468

classification placed among the top five in the manufacturing of pharmaceuticals, special equipment, instrumentation, computer, communication, and other electronic equipment, and software and information technology services.

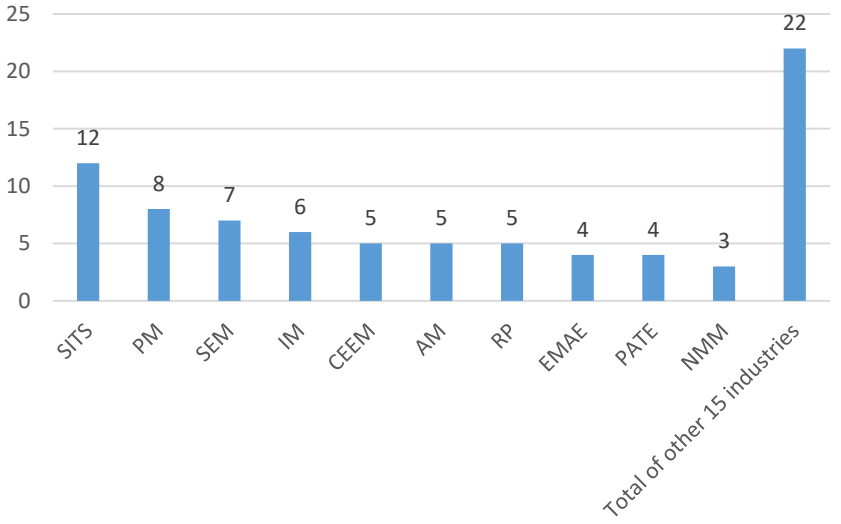


Figure 4: The industry distribution of listed companies in BSE.

The index of PE among in star market, second board, BSE, SZSE, and SHSE were shown on it. The index of the star market and second board are bigger than others. The trailing PE ratio in BSE is 45, which is a medium level. The trailing PE-ration in SZSE and SHSE are as follows.

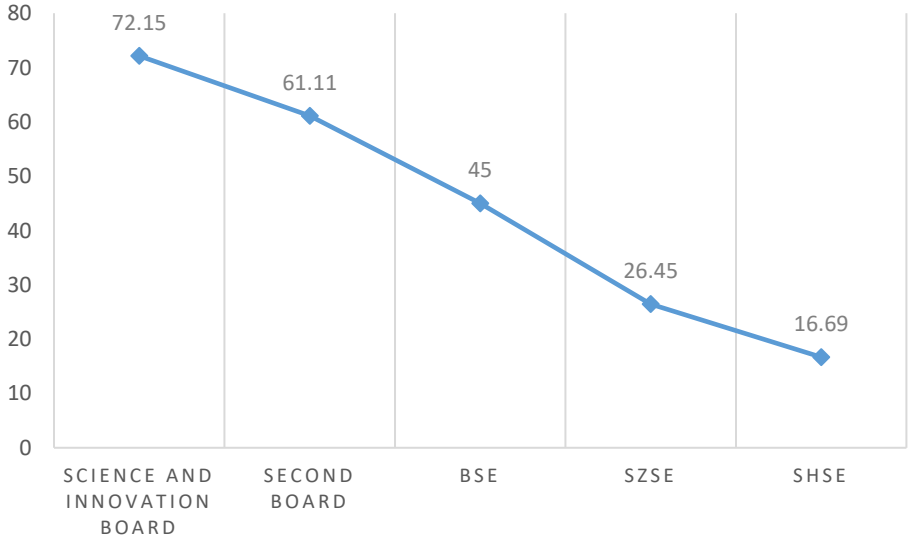
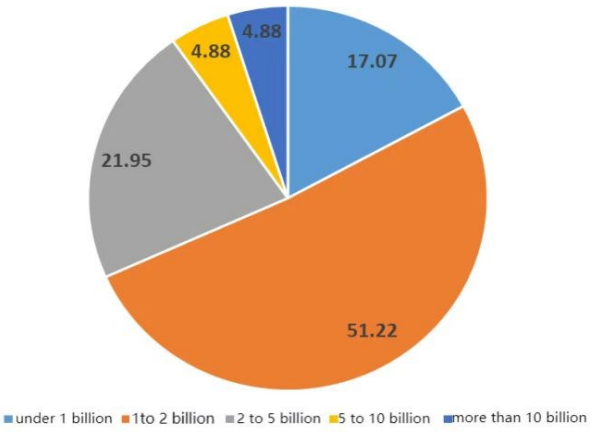


Figure 5: The trailing PE in science and innovation board, second board, Beijing.

DOI: 10.54254/2754-1169/10/20230468



Unit: %

Figure 6: Distribution of total market capitalization of listed companies on the north exchange in 2021.

The total number of companies listed on the North Exchange reached 100. Official data showed that as of June 23, the total share capital of 99 listed companies on the North Exchange was 14.543 billion shares, with 8.164 billion shares in circulation. The total market value is 205.929 billion yuan, and the outstanding market value is 131.238 billion yuan. As a whole, the market capitalization of the companies listed on the North Exchange varies greatly. Among them, Betree has the highest total market value of 61.6 billion yuan. Corresponding to this, there are many small and medium-sized market capitalization companies in the North Exchange. There are 67 listed companies whose total market capitalization is less than 1 billion yuan, 21 listed companies whose total market capitalization is between 1 billion and 2 billion yuan, and 9 listed companies whose total market capitalization is between 2 billion and 10 billion yuan.

3.2. Stocks' Factor Score

After the final factor loading matrix is obtained, the factor loading is calculated according to the heat map. The authors identified five factors, multiplied the original data by the factor loadings matrix, and then standardized it to eliminate the influence of the dimension, and obtained the following results. The top ten tickers are as follows in Table 4.

Table 4: Stocks' factor score and rank.

Ticker	Price	Trade	Debt	Earning	Growth	Rank
831305	0.530	5.080	2.515	0.248	1.893	1
836077	2.429	-0.381	-0.359	2.831	1.091	2
835185	5.122	-0.157	-0.707	-0.271	-0.297	3
832566	-0.168	-0.585	1.244	1.168	1.191	4
872925	-0.347	-0.781	0.495	-0.289	3.758	5
835179	-0.050	-0.441	0.928	0.000	2.187	6
835670	0.853	1.370	3.220	-1.838	-1.198	7
835174	-0.206	0.892	-0.140	1.117	0.335	8
833523	-0.178	-0.057	-0.140	-0.220	2.157	9
831832	-0.173	-0.080	0.216	-0.303	1.896	10

DOI: 10.54254/2754-1169/10/20230468

3.3. K-mean Cluster Analysis

This article takes a K-mean value of 3, which means that the factor scores of the 62 stocks are classified into 3 categories using cluster analysis as an indicator. Subsequently, the coordinates of the obtained clustering centroids are shown below in Table 5.

Table 5: The cluster coordinates of the center point.

	Factors				
Type	Price	Trade	Debt	Earning	Growth
1	0.363	3.613	1.142	-0.650	-0.194
2	-0.209	-0.154	-0.169	-0.695	1.578
3	0.025	-0.190	-0.033	0.200	-0.349

Next, this paper converts the center of various types of stocks into a score between 1-3 using the "max-min Standardization" method. In order to highlight the characteristics of each class, a radar map of each type of center score is drawn in Figure 7.

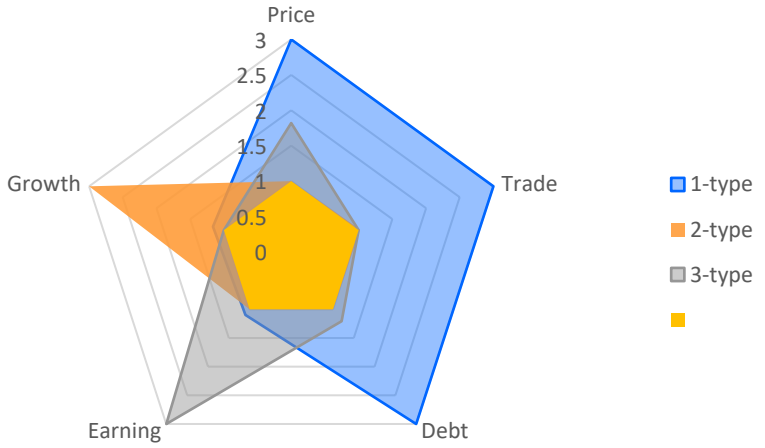


Figure 7: K-means cluster center radar map.

The above chart shows how the one type of stock fared better in relation to the three variables of price, trade, and debt. This demonstrates that even if the price of these stocks is high, they have good debt-paying capacity and excellent trade transaction performance. Their profitability does not stand out greatly. The Growth factor score indicates that they have a tendency to have a more steady development pattern. "High-Priced Stock" was given to it. The Growth component was the only one where the 2-type stocks performed well; all other factors were below average. "Potential Stock," so call it. The third class of equities outperformed on the Earn factor and underperformed on all other performance metrics. The third stock category, known as "Common Stock," has a flat overall impression.

DOI: 10.54254/2754-1169/10/20230468

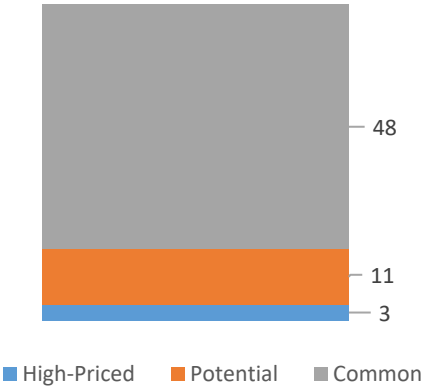


Figure 8: K-mean cluster result.

Figure 8 shows that the numbers of each type of stocks, it can find that most stocks belong to "Common Stock", and only a few stocks have their own outstanding advantages, which is in line with the current situation that most stocks in the stock market are mediocre, only a few stocks can stand out, indicating that the clustering results have a certain degree of reasonableness.

3.4. The Analysis of K-mean Cluster Result

Table 6: 1-type and 2-type ticker.

Ticker	Type	Ticker	Type
831305	1	835179	2
835670	1	837092	2
836263	1	837212	2
430198	2	838924	2
831768	2	871245	2
831832	2	871642	2
833523	2	872925	2

In the results of the cluster analysis, Table 6 above is the stocks that are considered worthy of investment in this article.

3.5. Suggestions for Different Types of Investors

In this situation, it is preferable for investors with sufficient startup capital and sound investment philosophy to invest in 1-type "High-Priced Stock," which has strong solvency and high trading volume, as long as the external environment, such as policies or war, acts as a force majeure to ensure that the type of stock is appropriate for investors of a sound philosophy. In this situation, choosing a "Potential Stock" of the second type would be more appropriate for investors that prioritize higher returns and have a higher tolerance for risk. They have a significant capability for expansion, and the business is already running at a higher level. Stocks of this kind are appropriate for risk-taking investors.

4. Conclusion

In conclusion, several different factors are analyzed by IMO and Bartlett tests and the orthogonal rotation method. This paper uses 62 listed companies of the North Exchange as examples for analysis.

DOI: 10.54254/2754-1169/10/20230468

The variance explained table reflects the contribution of the different kinds of factors to the explanation of the variable, while the gravel diagram shows the downward trend of Eigen which can be used to decide how many factors should be recorded. After that, the numbers of factors are five by analyzing the variance explained table and the gravel diagram. Combining the heat map of the factor loading and the characteristics of the factor itself, the individual factors are named as follows: price factor, trade factor, debt factor, earnings factor, and growth factor. Afterward, by analyzing the factor load factors and heat plots, it can obtain a matrix of importance and composition of the hidden variables in each factor. Based on the factor loading plot, multi-factors are reduced to several factors. Next, by analyzing the component matrix, the factor composition formula and weight are obtained. The comprehensive score of the output factor analysis method. All in all, several stocks with the highest comprehensive scores were calculated by factor analysis. And the comprehensive score of each factor was ranked. The comprehensive score of factor analysis will provide a reference for consumers with different investment styles. Value investment not only tests people's ability to analyze data but also tests value orientation and so on. In the selection of stock as a way of value investment, the listed companies with good economic benefits should be selected due to the consideration of multiple financial indicators of the listed companies. Data analysis is difficult due to a large number of miscellaneous data. The research purpose of this paper is to evaluate various value investment alternatives in an objective way based on various quantitative data, so as to help people make better investment decisions. Through the application of the factor method in multivariate statistics, this paper establishes the index system and evaluation standard for the economic situation of listed companies. Objectively, through a large number of data analyses, it provides a perspective and method for people to invest in stock value. Through the use of factor analysis, investors can make quantitative analyses of each index and extract the factors affecting the stock, so as to make a relatively objective evaluation of the listed company's stock, and provide an evaluation criterion for investors.

References

[1] *Zhu, G. J.: Promote Economic Transformation and Upgrading and Serve the High-quality Development of Small and Medium-sized Enterprises. In: China Academic Journal Electronic Publishing House, (2022).*

[2] *Zhang, J. B.: New Channels for Small and Medium-sized Enterprises to be Listed. In: Technology China (1), (2022).*

[3] *Cao, L.: The Anniversary of the Beijing Stock Exchange. In: China Business News, pp. Edition A08, (2022).*

[4] *Li, C.: Hsu Investment Decision Making Using a Combined Factor Analysis and Entropy-Based Topsis Model. In: Journal of Business Economics and Management 14(3), 448–466 (2013).*

[5] *Tian, H., Wang, W.R.: Research on Applicability of Stock Valuation Model. In: China Academic Journal Electronic Publishing House, (2022).*

[6] *Wang, Y.: Analysis of Stock Investment Value -- Taking Tianyu Stock as an Example. In: China Academic Journal Electronic Publishing House, pp. No.3x, (2022).*

[7] *Wu, J.: Analysis of Stock Investment Value of Listed Banks Based on Factor Analysis and Cluster Analysis. In: China Academic Journal Electronic Publishing House, (2022).*

[8] *Yang, S. Y., Guo, H. M., Li, J. G.: CNN-GRUA-FC Stock Price Forecast Model Based on Multi-Factor Analysis. In: Journal of Advanced Computational Intelligence and Intelligent Informatics 26(4), (2022).*

[9] *Gu, B. T., Zhu, J. M., Huang T. L.: Research on Investment Value of Innovation Board Based on Factor Analysis and Cluster Analysis. In: Journal of Jilin Agricultural Science and Technology University 29(3), (2022).*

[10] *Huang, H.: Research on the Value of Chinese Banking Stock Investment Based on Factor Analysis. In: China Academic Journal Electronic Publishing House, (2022).*

[11] *Han, F. C.: Comprehensive Evaluation Method of Initial Factor Analysis and its Application. In: Statistics and Decision Making, pp. No. 2x, (2022).*