Using Factor Analysis to Rank Several Chinese Public Stock

Zuming Jia*

University of California, Santa Barbara

*Corresponding author: zumingjia@ucsb.edu

Abstract. The stock market currently plays an important role in the whole society and people ranging from a child to the old are all starting to explore and invest in the stock market. Since the stock market is unpredictable and fluctuates, people need to master more information about the stock in order to prevent the risks of investing and profit from the investment competition. In this paper, several indicators will be proceeded factor analysis by using SPSS, which can help people better understand how the stocks go and prevent them from losing too much because of ignorance. This method firstly performs correlation analysis on stock price, and quantitatively explains the correlation between stocks. Then, researchers will use the SPSS to process the linear regression analysis. Finally, by using the function got from the process to calculate the final scores of each stock, researchers get the results that 300201 (stock symbol) scores the highest and know the scores of the other stocks, which greatly contribute people to finding and choosing a potential stock.

Key words: SPSS, stock market, factor analysis.

1. Introduction

Under the influence of the pandemic for the past few years, the global economy system around the world has been hit hard. Almost every industry is facing severe problems. People are trying to seek ways to turn the situation into a right way. Hence, people pay their attention to the market and try to benefit from the current situation. However, as more people get access to the Internet, investing in stocks seems to be a trend for more groups of people, like housewives, office workers, or even students. Except for the professional investors, some of the other people may have little knowledge about investing and little experience in investing in stocks. Most people were just interested in making extra money to begin to invest in stocks. Therefore, investing in stocks seems to be a common way for every family that has stable income and daily life to earn some extra money in their free time. Since there are lots of risks in the stock market, people without professional knowledge and experience in investing in the stocks have a large possibility of losing money in the stock. Under this circumstance, if people can analyze some factors of the stocks they want to invest in, it is undoubtedly that investors can get much more return and reduce the possibility of being in the risk. However, in the stock market, high return and high risk are directly proportional, which again shows the importance of getting to know the whole aspects of the stock. The reason why the stock market is diverse can be relevant to many factors. Since the stock market is the microcosm of the economy system, stocks and the economy system as a whole are closely related. Hence, the fluctuations and changes of the society can also be the factors that make the stock market unpredictable and puzzling. The factors, like the investment environment, governmental policy, celebrity events, and other changes in the external environment can all cause the investors to gain or lose. Investors need to collect information through various channels in order to correctly make the infestation judgments [1].

To select a stock with excellent conditions and potentials, people have explored many ways and methods from the origin of the stock. Besides some traditional ways, people now can use the Internet and some machines to evaluate a stock's comprehensive data. Some statistical analysis has been invented to be used as the tools which are suitable for people to evaluate some data. For example, to be more specific, Wu, by using the ARIMA Model (time series analysis) to build the ARIMA Model, can predict the fluctuation price, development trend, and laws of Chinese public stock, which can roughly predict the stock market by using the historical information of stocks [2]. Liang uses the LSTM (Long Short-term Memory) Model which is a special kind of RNN and is an artificial neural

network used in the fields of artificial intelligence and deep learning and human's emotion to forecast stock price development [3]. Liu used the principal component analysis method to extract the financial indicators of 15 listed securities companies, select the most influential comprehensive financial indicators, and rank the companies' competitiveness and influence [4]. In this paper, researchers are going to use factor analysis to evaluate and score several stocks of Chinese public companies.

2. Methodology

The factor analysis method is a statistical analysis method applicable to multiple variables. This method conducts scientific research on the dependencies existing within the variables, and scientifically summarizes the related variables with complex relationships to form a small number of comprehensive factors [5]. Factor analysis combines multiple variables into several factors, and on the basis of in-depth study of the internal correlation in the form of correlation matrix or covariance rectangle, it reveals the original correlation between variables and factors [6]. To be more simplified, people can just categorize the data which have high relevance as a category. For every group of variances, they can individually be a factor. By the scientific calculation, people can figure out that a few factors can have great significance on the whole dataset and tell the relevance among them.

In this paper, totally sixteen stocks of public companies will be analyzed by using factor analysis through the usage of SPSS software. SPSS software is a software product and is related service widely used in statistical analysis operations, data mining, predictive analysis and decision support tasks [9]. The relevance of these sixteen stocks will be discussed in the following paper, and a table will be built to see the score of each stock, which can greatly help people better know about various indicators of the stock they choose when they are going to invest in the stocks.

There are totally sixteen stocks chosen to be evaluated. They are individually: Shenzhen JT Automation Equipment Co, Hangzhou Huaxing Chuangye Communication Technology Company, Zhejiang Meili High Technology Company, Konfoong Materials International Company, and so on. All of these stocks are Chinese public companies' stocks. These sixteen stocks will be the subjects of the factor analysis in this paper. Below is the stock investment evaluation index that will be discussed in the following paper.

To analyze these stocks, there will be two indicators to evaluate the factors. There are four components in the first-level indicator, which are the Index of the Stock, Profitability Index, Potential Development index, and Financial Index, as figure 1 shows.

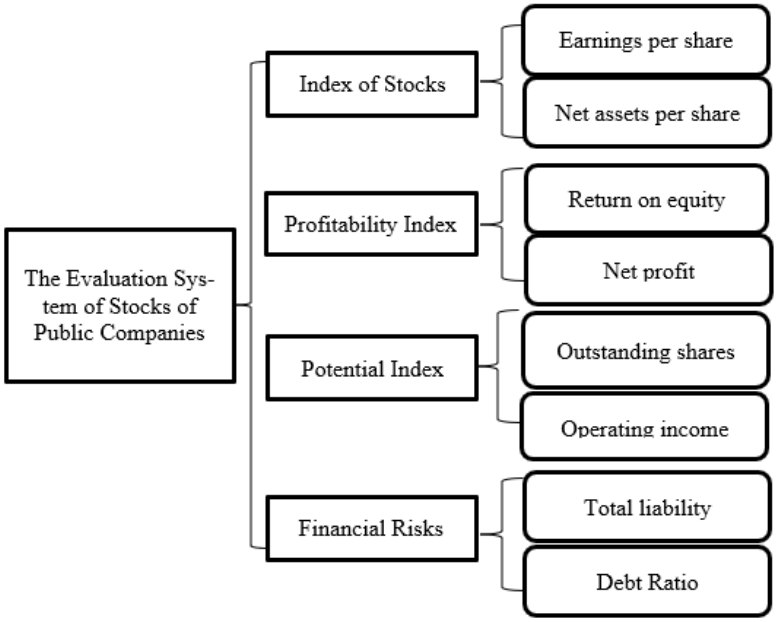

Figure 1. Indicators of the evaluation

For the index of the stock, it means the data of the stock itself, which includes earnings per share(X_1) and net assets per share(X_2). Profitability index can tell readers the profit and loss of the stock, which has return on equity(X_3) and net profit(X_4). Potential development index refers to the future development trend of the stock and whether it is a potential stock that deserves people to invest in, which are built by the outstanding shares(X_5) part and operating income part(X_6). Financial risk can show the debt of the stock, which includes total liability(X_7) and debt ratio(X_8).

Based on the indicators above, there are some professional words that are unfamiliar for people. So below is the definition of each indicator, as table 1 shows.

Table 1. Definition of each factor

Variable	Definition of each factor
X_1	Earnings per share is calculated as a company's profit divided by the outstanding shares of its common stock. EPS is the indicator of how much money a company makes for each share of its stocks, and EPS is a great and widely used tool to evaluate the value of a company.
X_2	Net assets per share is a real indicator of estate metric. The more expensive the net assets per share is, the more benefit holders can get from it.
X_3	Return on equity is measured by using the company's net income to divide its shareholders equity. It can show the income of holders.
X_4	Net profit is the most intuitive value to show the benefit and income. It can clearly reflect the profitability of the company.
X_5	Outstanding shares it means a company's whole stocks held by all shareholders, including share blocks held by institutional investors and restricted shares owned by the company's officers and insiders [7].
X_6	Operating income refers to the income made by the company's main business.
X_7	Total liability means the combined debts and obligations of a company owed to outsides parties.
X_8	Debt ratio is a financial ratio that indicates the percentage of a company's assets that are provided via debt [8].

Following, the data from the website named *Eastmoney in 2022* will be used as the resources to proceed the analysis. The results people got from this will be useful for people's reference for the future investments [9, 10].

3. Results and Discussion

To proceed the factor analysis, people need to first import the sorted data into SPSS. Then, in order to better exhibit the data, the random data need to be normalized within the use of SPSS. Through the SPSS, the standard deviation, variance, range, and some other information can be shown clearly.

In order to see whether this dataset is feasible, what researchers needs to do is to check the KMO and Bartlett tests results. In the Dimension Reduction part, researchers can analyze the given factors. By the calculation of SPSS, the results below are the KMO and Bartlett tests results.

Table 2. KMO and Bartlett's Test Results

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	-	0.675
Bartlett's Test of Sphericity	Approx. Chi-Square	135.29
-	df	28
-	Sig.	<0.001

In this test, for the Kaiser-Meyer-Olkin Measure of Sampling Adequacy, it gets 0.675. In KMO's metric, when it scores above 0.9, it means that the data are very suitable. The suitability decreases as the score decreases. As the score of KMO gets close to 1, it indicates that the relevance among the variance is stronger. On the contrary, as the KMO scores close to 0, it means that the relevance among the variance is getting weaker and make the data unsuitable to be used to process factor analysis. Hence, since KMO score= 0.675, it means that these data can be used to perform factor analysis, even though it's not perfect. In the Bartlett's Test of Sphericity, there are three components, which are individually Approx. Chi-Square, df, and Sig. They are all devote to prove the feasibility of proceeding factor analysis of these data. Take the Sig. as an example. Through table 2, it's clearly that the score of the significance is less than 0.001 but not equals to 0. Hence, this significance is failed to reject the hypothesis. Therefore, the conclusion is that these data are suitable to be used to conduct the factor analysis.

By processing these data through SPSS, like dimension reduction, people can get the covariance matrix, correlation matrix, total variance including their explanations, scree plot, and a series of data which can be useful for the analysis. Based on the calculation above, people get a lot of results that can provide them with more information.

Table 3. Correlation Matrix

Correlation Matrix								
	EPS	NAPS	ROE	Netprofit	TSO	BI	DT	DTS
EPS	1.000	0.418	0.854	0.866	-0.026	0.622	-0.513	-0.587
NAPS	0.418	1.000	0.333	0.212	-0.551	0.075	-0.576	-0.734
ROE	0.854	0.333	1.000	0.936	-0.013	0.625	-0.695	-0.723
Netprofit	0.866	0.212	0.936	1.000	0.205	0.819	-0.542	-0.570
TSO	0.026	-0.551	0.013	0.205	1.000	0.627	0.579	0.403
BI	0.622	-0.075	0.625	0.819	0.627	1.000	-0.053	-0.181
DT	0.513	-0.576	0.695	-0.542	0.579	0.053	1.000	0.851
DTS	0.587	-0.734	0.723	-0.57	0.403	0.181	0.851	1.000
EPS	-	0.054	<0.001	<0.001	0.406	0.005	0.021	0.008
NAPS	0.054	-	0.104	0.215	0.013	0.391	0.010	0.001
ROE	0.000	0.104	-	0.000	0.481	0.005	0.001	0.001
Netprofit	0.000	0.215	0.000	-	0.223	0.000	0.015	0.011
TSO	0.461	0.013	0.481	0.223	-	0.005	0.009	0.061
BI	0.005	0.391	0.005	0.000	0.005	-	0.423	0.251
DT	0.021	0.010	0.001	0.015	0.009	0.423	-	0.000
DTS	0.008	0.001	0.001	0.011	0.061	0.251	0.000	-

The table above is the Correlation Matrix chart which can show the relevance of each factor. By comparing the values in each row, only the same indicator has the same score for the correlation. For the rest of indicators, they do not equal to 1, which mean that the correlation among them is very low and can be assumed that there is no relation. Hence, it's clear that there is no relevance for this dataset, which means this is completely random data that can be used for the factor analysis.

Table 4. Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
Component	Total	Variance	Cumulative	Total	Variance	Cumulative	Total	Variance	Cumulative %
1	4.488	56.098	56.098	4.488	56.098	56.098	3.906	48.829	48.829
2	2.440	30.503	86.601	2.440	30.503	86.601	2.039	25.487	74.317
3	0.501	6.265	92.867	0.501	6.265	92.867	1.484	18.550	92.867
4	0.327	4.090	96.957	-	-	-	-	-	-
5	0.122	1.520	98.477	-	-	-	-	-	-
6	0.060	0.744	99.221	-	-	-	-	-	-
7	0.050	0.631	99.852	-	-	-	-	-	-
8	0.012	0.148	100.000	-	-	-	-	-	-

Table 4 is the variance rate of this dataset. By simply browsing this chart, the first three components together in the chart make up 92.867% of the assembly, which take a very big and crucial part for the whole dataset. Besides, the cumulative ratio of the first three factors stays almost the same after rotation even if the value of each factor changes a little bit, the small change will not influence the estimation that the top of three factors play an important role in verifying and representing these data. Hence, this dataset perfectly met the standards of evaluating stocks and can be used to process factor analysis. The Screen plot is the diagram form that displays the characteristic roots. It's mainly used to assist in determining the number of factors.

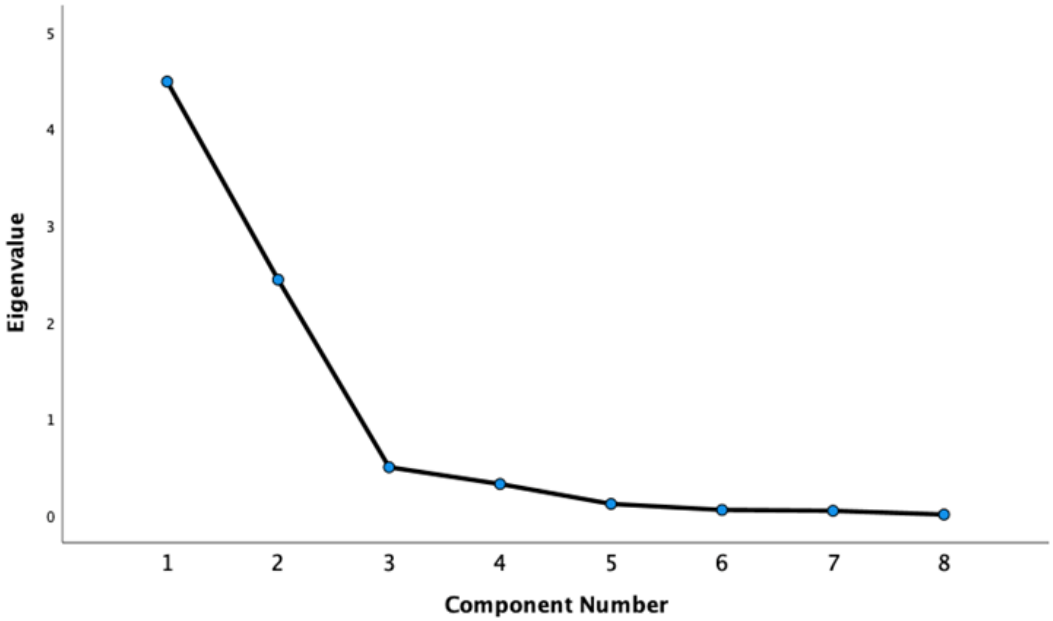

Figure 2. Scree Plot

Based on the Screen Plot above, the coordinate horizontal axis refers to the component number which is from 1 to 8. Each number can represent each factor researchers used in this analysis. For the vertical axis, it represents the eigenvalue that. Through this picture, the eigenvalue of the first three factors greatly changes and the trend is gradually leveling off for the rest of factors. The first three factors are almost all larger than 1, and the rest five factors are all less than 1, which means that these factors cannot express eigenvalue for the dataset. Therefore, the first three factors have significant influence on the data and will be used for the future calculation.

Table 5. Rotated Component Matrix

Rotated Component Matrix			
	Component		
	1	2	3
Netprofit	0.989	-0.042	0.071
ROE	0.930	-0.281	0.131
EPS	0.849	-0.086	0.324
BI	0.846	0.473	-0.039
TSO	0.280	0.840	-0.343
DT	-0.493	0.797	-0.261
NAPS	0.126	-0.322	0.931
DT	-0.536	0.532	-0.549

The table above is the rotated component matrix. Component 1 has larger loads on net profit, return of equity, earnings per share, and operating revenue, which means these variances cause great influence on component 1, so it can be categorized as an efficiency factor. For component 2, outstanding shares and total liability take up a big part, so it can be the scale factor. In component 3, net assets per share and debt ratio are mainly dominated so that it can be the development factor.

Consequently, these three factors are assigned with specific meanings, which simplify the whole dataset. These three components include all these eight factors and tell the proportion of the function.

Table 6. Component Score Coefficient matrix

	Component		
	1	2	3
Netprofit	0.186	0.114	0.198
ROE	-0.154	0.408	1.042
EPS	0.267	-0.217	-0.241
BI	0.283	-0.047	-0.157
TSO	0.109	0.489	0.091
DT	0.232	0.349	0.113
NAPS	-0.139	0.562	0.351
DT	-0.069	0.103	-0.247

Through table 6, people can know that factor scores can be obtained by regression. Based on all the information above, we can get the function of factor score:

$$F = (0.48829F1 + 0.25487F2 + 0.1855F3)/0.92876 \quad (1)$$

Therefore, by using the factor analysis, researchers can get the scores of these 16 stocks.

Table 7. Stocks' Scores and Ranking (Descending Order)

Stock Symbol	Comprehensive score	Ranking	Stock name	Comprehensive score	Ranking
300201	1.4947	1	300564	-0.2171	9
300735	0.9286	2	300611	-0.3569	10
300319	0.7015	3	300386	-0.3827	11
300666	0.4339	4	300309	-0.4477	12
300928	0.4122	5	300025	-0.4697	13
300978	0.1232	6	300436	-0.4738	14
300187	-0.004	7	300400	-0.7426	15
301120	-0.211	8	300108	-0.7679	16

From the table above, people can clearly see the score of each stock including their ranking, which is a great visualized way to help people screen the stocks. The stock, ST Helen, gets the highest comprehensive score and ranks at the first, which far exceeds the other stocks. For the rest 15 stocks, their comprehensive scores were stably decrease.

4. Conclusion

In this project, researchers use factor analysis to evaluate 16 public stocks and get the scores of each stock, which provides investors with some helpful information about nine different aspects of the stocks, making investors have a more comprehensive understanding of their desired stock before they invest in. The ranking above is the result researchers get to intuitively show the scores of each which can evaluate several indicators of the stocks. This process is crucially important since it can largely prevent the investors from being in the risk of investments. Since the stock market develops rapidly, a steady and rational approach should be used to conduct investment analysis. This method can standardize the investment market and increase the competitiveness of the public companies. In the meantime, the experimental results are not perfect and need to be further improved. Since the stock market is fluctuated, people need to more information before they start to invest in stocks.

Reference

[1] Huang. H, Research on the value of stock investment in my country's banking industry based on factor analysis. modern commerce, 2022.03.026. pp 86-88.

[2] Wu. Y, Wen Xin.Short-term stock price forecast based on ARIMA model. Statistics and Decision, 2016.23.051, pp 83-86.

[3] Liang. Y, S. Dongfeng, Stock Prediction Based on LSTM and Sentiment Analysis. Technology and Innovation, 2021.21.053, pp 126-127.

[4] Liu. Q, Research on the Competitiveness of Listed Securities Companies Based on Principal Component Analysis. Economic Research Guide, 2011, pp 74-76.

[5] Yang. X, Wang. Y, Analysis of stock investment value of listed commercial banks in China based on factor analysis. Special Economic Zone, December 2015, pp 67-69.

[6] Chen. Y, Empirical Research Based on Multivariate Statistical Analysis Method——Taking the Stock Market as an Example. Marketing, 2021, pp 40-43.

[7] J. Chen, Learn about shares outstanding. Investopedia, 2002, pp 1-5.

[8] Wikimedia Foundation. Debt ratio, Wikipedia, 2022, July 7, pp 1-2.

[9] Zan. J, Analysis of Factors Affecting Stock Transaction Amount Based on SPSS. business accounting, 2016, pp 67-69.

[10] Xu. C, Zhang. X, On the comprehensive performance evaluation of my country’s listed securities companies - based on principal component analysis. 2015, pp 19-21.