Stock Price Factors Research and Classification

Zichun Hu [1,a †], Wenjie Wang[2, *, †]

[1]Beijing Normal University-Hong Kong Baptist University, United International College, Zhuhai, China
[2]University of Rochester, School of Mathematics, Rochester, USA
[a]p930018026@mail.uic.edu.cn
[*]wwang96@u.rochester.edu
[†]These authors contributed equally.

Abstract

Nowadays, to enter the stock market, company selection (fundamental analysis) and evaluation based on stock price (technical analysis) are significant parts of the preparation for a rational investor. Since there are multiple ways to do the stock analysis and stochastic elements in stock investment, none of these measures can be universally acknowledged infallible, but the result still can be a reference for investors. In this paper, 60 stocks are randomly picked from the American stock market and factor analysis is used to transform the 8 indicator variables of each stock chosen from Wind into 3 factors. Then by using the 3 factors of each stock, this paper figures out the score of each stock and rank the 60 stocks from the highest score to the lowest. Finally, this paper uses the cluster analysis to classify the stocks into 4 categories. The result of the score of each stock and the 4 categories of stocks can be a reference for investors.

Keywords: *stock price; factor analysis; cluster analysis.*

1. INTRODUCTION

With the prosperity of the financial market in recent years, more and more investors choose to believe the stock market and enter the financial market to make profit. For most investors, the most difficult part of investing the stock is that the stock price is so volatile that sometimes it is not able to predict [2]. However, most of them still believe that the movement of the stock price is not totally a stochastic process [4]. In fact, creating an intelligent prediction system to the stock price has always been a subject of great interest for many investors and financial analysts [7]. This would raise the attention of each individual stock price into the factors beyond the price that make an influence [9].

Connecting the big events in history and some companies' stock prices, it affects the stock price much more significantly and directly. For example, in 2020, equity markets in the European Union and the United States dropped around 30 percent between February and March due to covid-19 [8]. Making a general analysis to the stock price changing during this period, covid-19 causes more and more workers being unable to work as usual. Then, most manufacturing companies were not able to make the product which leads to the decreasing yield. Therefore, both their interests decreased, and investors' expectations became more negative. Reflecting on the stock markets, the decreasing of stock prices is predictable during this period [6]. Considering another sample, in 2016, China introduced the Korean Restriction Order to limit the Korean development of pan entertainment in China. Just three days after the order, JYP stock price decreased 5.4%, SM's stock price decreased 4.8%, CJ EM's stock price decreased 8.99% and YG's stock price decreased 11.98%. All these companies had the main business in pan entertainment, especially the outputting to China. The order restricts their entertainment business with China which directly cut down the interests. Therefore, if any investor got that information like the deterioration of covid-19 in Europe and America or the introduction of the order, they could guess the stock price more accurately in future. But the big events are not happening every day. Therefore, the useful module should not always use those events as one main factor to the changing of price. Then, a good stock price module requires the study of factors that would be affected by the price (Hassan & Natalya, 2016). Looking through the basic operation of each company, the following are the possible factors: EPS, ROE, EBITDA. The liabilities to assets ratio, asset turnover ratio, Gross revenue, Equity value per share, Current ratio [5].

G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 420–425, 2023.
https://doi.org/10.2991/978-94-6463-042-8_61

2. METHODOLOGY

The following includes the introduction of the factors and the basic meaning to the company operation which is used in this paper. All these factors are suggested in various of literatures. All the data sources are from WIND dataset.

2.1. Factors and data collection

TABLE 1. FINANCIAL EVALUATION INDICATOR

First indicator	Processing	Second indicator
Per share indicator	X1	EPS
	X2	EBITDA
Profitability indicator	X3	ROE
	X4	Gross revenue
Growth Capability	X5	Equity value per share
Operational Capability	X6	Asset turnover ratio
Financial risk	X7	Current ratio
	X8	Asset-liability ratio

In Table 1, EPS represents earnings per share. It is counted by the outstanding shares of its common stock divides company's profit. The resulting number serves as an indicator of a company's profitability. ROE, return on equity, is the variable through dividing net income by shareholders' equity. It is considered a gauge of a corporation's profitability and how efficient it is in generating profits. EBITDA is called earnings before interest, taxes, depreciation, and amortization, one measurement of a company's overall financial performance. It is usually used as an alternative to net income. Asset-liability ratio, the liabilities to assets (L/A) ratio, is a solvency ratio that examines how much of a company's assets are made of liabilities. Companies in signs of financial distress will often also have high L/A ratios. Asset turnover ratio measures the value of a company's sales or revenues relative to the value of its assets, which can be used as an indicator of the efficiency. Gross revenue, also known as gross sales. When it is recorded, all income from a sale is accounted for on the income statement. Equity value per share is counted by the total amount of money including share's value and other properties divided by the total number of outstanding shares. It usually is considered as the total value of the company that is attributable to equity investors. Current ratio, the current ratio, measures a company's ability to pay short-term obligations or those due within one year [3]. It provides information about how a company can maximize the current assets on its balance sheet for debts. Based on the factors considering about, this paper selected 60 stocks and recording those prices from WIND dataset. Importing the sorted dataset into SPSS Statistics 22.0. All the data are standardized by SPSS.

2.2. Factor analysis

Factor analysis is the extension of the main factor analysis which is a measure to simplify the multiple variables to fewer variables. The basic idea of factor analysis is to extract the key variables which can influence all the variables or samples by analyzing the correlation matrix to describe the correlation between multiple samples or variables [1]. However, in this case, the few random variables are not observable. Therefore, they are called factors. The next step is dividing those variables into groups according to the values of their correlation, which needs to make sure that the correlation between variables in the same group is large and the correlation between variables from the different groups is small [10]. In general, there are two types of method of factor analysis, the R factor analysis and Q factor analysis. The factor analysis in the paper is based on the variables of each sample (R factor analysis).

By using factor analysis, the number of analysis variables can be reduced. Then by using the correlation between variables and variables, the original variables can be classified. The model can be written as:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \ldots + a_{pm}F_m + \varepsilon_p \end{cases} \tag{1}$$

And

$$\underset{(\mathrm{p}\times\mathrm{l})}{X} = \underset{(\mathrm{p}\times\mathrm{m})}{A}\underset{(\mathrm{m}\times\mathrm{l})}{F} + \underset{(p\times\mathrm{l})}{\mathcal{E}} \tag{2}$$

where $m \leq p$; and $\mathrm{Cov}\,(\mathrm{F}, \varepsilon) = 0$. In another word, $m \leq p$; and F is independent to ε. And $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{p})'$ is the random vector of p dimensional variables. $F = (F_1, \ldots, F_p)'$ represent the vector unmeasurable which is also the common factor of X. The a_{ij} is the load of ith variable on the jth common factor. And ε is the special factor of X. The purpose of the factor analysis is to use the X to substitute F by using $X = A * F + \varepsilon$.

2.2.1. Clustering analysis

Cluster analysis is a statistical method which classifies a batch of samples or variables according to their affinity degree. The common clustering methods are two-step clustering, K-means clustering and system clustering. In this project, system clustering which uses the Euclidean distance clustering to define the sample distance is used.

Since the samples or variables researched have similarity in different levels, according to the multiple variables of samples, a statistical measure can be found to measure similarity between the variables or samples. Based on that, the clustering method is used to aggregate the samples or variables into different classes to make sure that the variables or samples in the same class are more similar and variables or samples in the different class are less similar. The type of clustering used by this article that aggregates the samples is Q type clustering. The method used to aggregate the samples is the systematic clustering.

1) Data transformation

Suppose there are m cluster objects, and each cluster object is composed of multiple elements. In general, there are different dimensions and different ranges of value. To make the data of different dimensions and different ranges of value can be compared together, data standardization is required.

2) Distance

In the process of cluster analysis, the most common way to describe whether the samples are closed is using the distance to explain. Suppose there are n samples, and each sample has p variables. Therefore, the original matrix is:

$$x = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \tag{3}$$

The X_{ij} is the measured variables of the jth variable on ith sample. d_{ij} represents the distance between ith sample and jth sample. The smaller d_{ij} is, the closer the sample is.

The matrix requires $\mathrm{d_{ij}} \geq 0$, when $\mathrm{d_{ij}} = 0 \Leftrightarrow \mathrm{X_{(i)}} = \mathrm{X_{(j)}}$,$\mathrm{d_{ij}} = \mathrm{d_{ji}}$and $d_{ij} \leq d_{ik} + d_{kj}$, for every $i,\ j,\ k$.

There are different distances with different definitions, and in this article, the Euclidean distance is used.

$$d_{ij} = \left(\sum\nolimits_{i=1}^{p} \left|X_{ik} - X_{jk}\right|^{2}\right)\hat{}\,(1/2) \tag{4}$$

2.2.2. Factor rotation

Operating the factor analysis is used to both find out the main factor and the objective meaning of each main factor. Therefore, to explain each factor, the component matrix should be rotated. Generally, the expression of the factor model is:

$$X_i = a_{i1}F_1 + \cdots a_{im}F_m + a_i u_i \tag{5}$$

where $F_1, \ldots, F_m$ represent the main factors, a_{ij} is the loading of the factors, which is the loading of ith variable in the jth main factor. u_i is the special factor and a_i is the loading of the special factor.

For the initial component matrix, if the loading of the factors doesn't differ much from each other, it is difficult directly to find out the explanation of each factor. Therefore, the rotation of the component matrix is needed that each loading of the factor can be polarized toward 0 and 1 by column and row. This article chooses to maximize the variance, which is the most common way to rotate the component matrix. After the rotation, the original variable only has a large loading on one factor, while the loading on the other factors is small. Therefore, the meaning of the main factor will be clearer that each factor can be intuitively divided by each factor so that the model can be simplified.

3. RESULTS AND DISCUSSION

TABLE 2. KMO AND BARTLETT'S TEST

KMO and Bartlett's Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.594
Chi-Square.	213.391
Df.	28
Sig.	0

According to table 2, the KMO value is larger than 0.5 and the Sig is less than 0.001. Therefore, doing main factor analysis is useful in this research.

According to table 3, it is obvious that all the extractions of normal variables are larger than 0.5 and most of them are close to or larger than 0.8. which means the variables can be expressed well by the main factor.

TABLE 3. COMMON FACTOR VARIANCE

Variables	Initial	Extraction
EPS	1	0.567
ROE	1	0.867
EBITDA/Gross revenue	1	0.768
Asset-liability ratio	1	0.571
Total asset turnover	1	0.783
Gross revenue	1	0.793
Equity value per share	1	0.84
Current ratio	1	0.863

From table 5, by observing the value of load on the three factors of every variable, some general claims for the factors are: (1) EBITDA/Gross revenue, current ratio, equity value per share and asset liability ratio has more loading in factor 1. Because the variables have the key words about liability, equity, and asset, it can be concluded that the factor 1 is the company fundamental factor. (2) ROE and total asset turnover respectively the efficiency of using equity and asset. Therefore, factor 2 is the company operating factor. (3) EPS and Gross revenue conquer the factor 3, which can be given the meaning more flexible, and this author decided to give the name stock potential profit factor.

To make sure the factors are extracted on different dimensions during the evaluation, the correlation among factors shouldn't occur. According to table 2-6, it shows the score covariance matrix which is equivalent to the identical matrix. The three factors are unrelated to each other and can represent different dimensions.

TABLE 4. TOTAL VATIANCE EXPRESSED

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	*% Variance*	*Cumulative %*	*Total*	*%Variance*	*Cumulative %*	*Total*	*% Variance*	*Cumulative %*
1	3.116	38.944	38.944	3.116	38.944	38.944	2.567	32.089	32.089
2	1.575	19.681	58.626	1.575	19.681	58.626	1.839	22.993	55.082
3	1.363	17.043	75.669	1.363	17.043	75.669	1.647	20.587	75.669

TABLE 5. ROTATED COMPONET MATRIX

Variables	Component		
	1	2	3
EBITDA/Gross revenue	0.855	-0.009	-0.194
Current ratio	-0.822	0.007	0.432
Equity value per share	-0.794	-0.384	-0.249
Asset-liability ratio	0.688	0.266	0.164
ROE	0.213	0.905	0.049
Total asset turnover	0.086	0.839	-0.268
Gross revenue	-0.011	0.072	0.888
EPS	0.058	0.304	-0.687

Score of the factors (Table 2-7) of every stock can be computed by using the variance contribution rate of the three main factors after the rotation (Table IV).

$$W_i = (0.321 \times F_1 + 0.230 \times F_2 + 0.2067 \times F_3)/0.757 \quad (6)$$

TABLE 6. FACTOR SCORE COVARIANCE MATRIX

Factor	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1

According to the score of each stock in Table 7, the evaluation combining fundamental factors and stock market factors of each stock can be figured out. In these stocks, the AERI.O gets the highest score 1.757. ABC.N(American source Bergen) takes the second which is 1.519. It also should be noticed that the ACRS.O (ACLARIS THERAPEUTICS) gets -1.715 which is the only stock with score less than -1. The score of other stocks with normal performance lies between negative 1 and positive 1.

According to tree diagram, figure 1(the figure shows only 40 stocks among the 60 stocks), this project divides the sample into 4 categories.

The first category includes only 1 stock, ABC.N. AmerisourceBergen Corporation company focuses on the pharmaceutical sourcing and distribution services, helping both healthcare providers and pharmaceutical. In addition, the biotech manufacturers improve patient access to products and enhance patient care. Based on the evaluation of the ABC.N, it performs well and gets the second highest grade. Because of the high grade in factor 2, the ABC.N has an excellent ability in maximizing the return on limited equity and using assets to run the company. However, compared with its excellent grade in factor 2, it gets a very low grade in factor 3 and common grade in factor 1, which means it lacks the ability to grow. Therefore, it is defined as the stock overvaluing in the short period and lack of the potential profit in the long period.

The second category includes the AKTS.O and ACRS.O. In the evaluation, both two stocks rank at the bottom with lowest grade especially ACRS.O, whose grade is even about -0.9 smaller than the last stock but one. An obvious problem of these two stocks is that the companies have trouble in companies' fundamental running. Therefore, the second category is defined as the stock which is of no worth based on the present condition. Since the data is 2021's third quarter report, it may be shown in the stock price that AKTS.O and ACRS.O are in the same category. According to the graphs from Baidu stocks, it supports the opinion above since the trend of their price is very similar.

The third category includes ALNY and AERI, two stocks with the highest grade in the evaluation. These two stocks have the same characteristics that the factor 3 score is very high, which means the stocks' potential profit may be very high. Therefore, the third category is defined as the potential stock, or the stock undervalued.

The fourth category includes the other stocks which rank between stocks from the other three categories. These stocks have the common score of the factor 1, 2 and 3. Therefore, the fourth category is defined as the stock which remains to be observed.

TABLE 7. SCORE OF THE STOCK

Rank	Code	Factor 1	Factor 2	Factor 3	Score
1	AERI.O	1.395	2.355	3.730	1.757
2	ABC.N	0.268	6.375	-0.159	1.519
3	ALNY.O	0.692	-0.816	4.375	0.935
4	ALEC.O	-0.204	-0.143	2.501	0.416
5	AAL.O	1.089	-0.475	0.547	0.352
6	ADSK.O	0.862	-0.081	0.266	0.312
7	AAPL.O	0.444	0.885	-0.286	0.287
8	AINC.A	0.468	-0.909	1.426	0.234
9	AES.N	0.858	-0.689	0.363	0.191
10	ALSN.N	0.624	0.027	-0.143	0.177
50	ALG.N	-0.468	0.351	-0.562	-0.185
51	ALGN.O	-0.105	0.019	-0.758	-0.185
52	ACHC.O	0.041	-0.608	-0.286	-0.185
53	AJRD.N	-0.142	-0.266	-0.425	-0.194
54	ACLS.O	-0.689	0.154	-0.129	-0.212
55	AAON.O	-0.502	-0.101	-0.296	-0.245
56	AAWW.O	0.061	0.141	-1.455	-0.247
57	ADNT.N	-0.318	0.489	-1.254	-0.247
58	ADI.O	-0.716	-0.960	-0.686	-0.591
59	AKTS.O	-3.980	0.063	1.990	-0.852
60	ACRS.O	-5.425	-0.023	0.150	-1.715

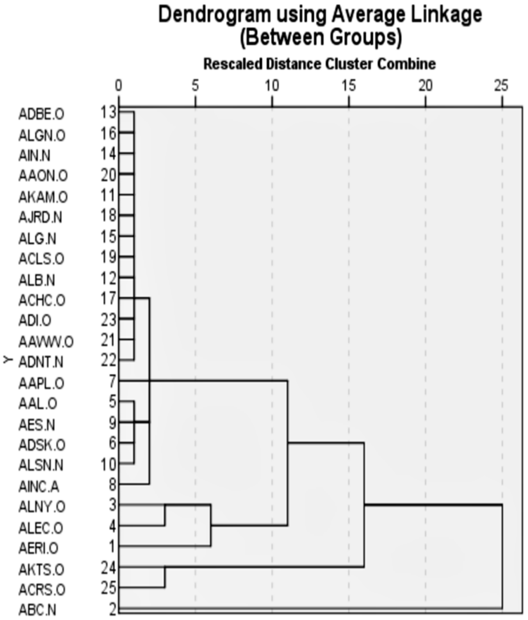

Figure 1. Dendrogram using average Linkage (Between group)

4.Conclusion

In this paper, 60 stocks with 8 variables of each stock from the American stock market are divided by factor analysis and clustering analysis. The result of the four categories concludes the investing condition of the stock companies, which can be the reference data provided to the investors.

However, it can't be determined that the 8 variables of the stock picked up is the best index to describe the investing condition, since according to the CAPM theory, before investing a stock, investors also need to consider the influence from the whole market. In the project, it only concludes the factor of the single stock. Therefore, the accurate result still needs to be explored in the future.

Multivariate statistical methods are used in this project, from the problems found, data. acquisition, index selection, the use of multivariate statistical methods and results in every link of operation, and hungry for more profound understanding. In the reliability of the experimental results and the actual economic significance on the interpretation of the result of the clustering, there is insufficient, needs further improvement.

REFERENCES

[1] B. Martin, E. Tomas, and W. Claudia, "The far-reaching implications of F0ama's efficient markets hypothesis: non-predictability of media investments". Taylor and Francis Group. vol. 27, 2020, 18, 1505–1508.

[2] D. Durusuciftci, et al., "Do stock markets follow a random walk? New evidence for an old question." International Review of Economics & Finance,2019, 64:165-175.

[3] D. Honfstrand, "Financial Ratio", Ag Decision Maker, 2015, 1-4.

[4] K. Chaudhuri, and Y. Wu, "Random walk versus breaking trend in stock prices: evidence from emerging markets", Journal of Banking and Finance, Vol. 27 No. 4, 2003, pp. 575-592.

[5] M. Flad, and R. C. Jung. "A common factor analysis for the US and the German stock markets during overlapping trading hours." Journal of International Financial Markets Institutions & Money,2008, 498-512.

[6] N. G. Ralph, and S. J. Koijen, "How the coronavirus affects stock prices and growth expectations". Chicago Booth Review. March 26, 2020.

[7] R. Hefezi, J. Shahrabi, and E. Hadavandi, "A bat-neural network multi-age system (BNNMAS) for

stock price prediction: Case study of DAX stock price". ScienceDirect. vol. 29, 2015, 196-210.

[8] R. Nelson, C, and C. Plosser, "Trends and random walks in macroeconomic time series", Journal of Monetary Economics, Vol. 10 No. 2, 1982, pp. 139-162.

[9] S. Hassan, D. Natalya, "The Random Walk in The Stock Prices of 18 OECD Countries". Journal of Economic Study. vol. 43, 2016, 598-608.

[10] W. Lo, A, and A. C. MacKinlay, "Stock Market Prices Do Not Follow Random Walk: Evidence from a Simple Specification Test." The Review of Financial Studies, 1988, 1: 41–66.

[11] Y. Campbell, et al, "An Intertemporal CAPM with Stochastic Volatility." LSE Research Online Documents on Economics, 2018, 207-233.