

Analysis and Categorization of Stock Price Factors via a Novel Framework based on Computer Science Technology

Chang Li, Bing Pan, Zheng Xiang, Lixuan Zhang, Lee Chen, Don Chen

Faculty of Computer Science and Information System, Universiti Teknologi MARA (UiTM), Malaysia

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: classification; stock price; factor analysis; cluster analysis, discrimination, adverse selection

1.0 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. However, most of them still believe that the movement of the stock price is not totally a stochastic process. 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. This would raise the attention of each individual stock price into the factors beyond the price that make an influence. 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. 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' 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 [1-23].

2.0 METHODOLOGY

The following includes the introduction of the factors and the basic meaning to the company operation

*This work is licensed under the Creative Commons Attribution International License (CC BY).
Copyright © The Author(s). Published by International Scientific Indexing & Institute for Scientific Information*

which is used in this paper. All these factors are suggested in various of literatures. All the data sources are from WIND dataset. 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 [24-36].

TABLE 1. FINANCIAL EVALUATION INDICATOR

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

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 [37-43].

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

And

$$X_{(p \times 1)} = A_{(p \times m)} F_{(m \times 1)} + \varepsilon_{(p \times 1)} \quad (2)$$

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.

3.0 RESULT

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.

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 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 identical matrix. The three factors are unrelated to each the score covariance matrix which is equivalent to the other and can represent different dimensions.

TABLE 4. TOTAL VARIANCE 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

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 5. ROTATED COMPONENT 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

4.0 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] Dionne, Georges, Christian Gouriéroux, and Charles Vanasse. "Testing for evidence of adverse selection in the automobile insurance market: A comment." *Journal of Political Economy* 109.2 (2001): 444-453.
- [2] Hinz, Oliver, II-Horn Hann, and Martin Spann. "Price discrimination in e-commerce? An examination of dynamic pricing in name-your-own price markets." *Mis quarterly* (2011): 81-98.
- [3] Birjandi, Alireza Komeili, Sanaz Dehmolaee, Reza Sheikh, and Shib Sankar Sana. "Analysis and classification of companies on tehran stock exchange with incomplete information." *RAIRO-Operations Research* 55 (2021): S2709-S2726.
- [4] Edelman, Benjamin, Michael Luca, and Dan Svirsky. "Racial discrimination in the sharing economy: Evidence from a field experiment." *American economic journal: applied economics* 9.2 (2017): 1-22.
- [5] Nazari Enjedani, Somayeh, and Mandar Khanal. "Development of a Turning Movement Estimator Using CV Data." *Future Transportation* 3, no. 1 (2023): 349-367.
- [6] Tirole, Jean. "Overcoming adverse selection: How public intervention can restore market functioning." *American economic review* 102.1 (2012): 29-59.
- [7] Saadat, MohammadReza, and Benedek Nagy. "Cellular automata approach to mathematical morphology in the triangular grid." *Acta Polytechnica Hungarica* 15, no. 6 (2018): 45-62.
- [8] Richardson, Alan J., Michael Welker, and Ian R. Hutchinson. "Managing capital market reactions to corporate social responsibility." *International Journal of Management Reviews* 1.1 (1999): 17-43.
- [9] Saadat, Mohammad Reza, and Benedek Nagy. "Generating Patterns on the Triangular Grid by Cellular Automata including Alternating Use of Two Rules." In *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 253-258. IEEE, 2021.
- [10] Amini, Mahyar, and Zavareh Bozorgasl. "A Game Theory Method to Cyber-Threat Information Sharing in Cloud Computing Technology ." *International Journal of Computer Science and Engineering Research* 11.4 (2023): 549-560.
- [11] Nazari Enjedani, Somayeh, and Mahyar Amini. "The role of traffic impact effect on transportation planning and sustainable traffic management in metropolitan regions ." *International Journal of Smart City Planning Research* 12.9 (2023): 688-700
- [12] Jahanbakhsh Javidi, Negar, and Mahyar Amini. "Evaluating the effect of supply chain management practice on implementation of halal agroindustry and competitive advantage for small and medium enterprises ." *International Journal of Computer Science and Information Technology* 15.6 (2023): 8997-9008
- [13] Amini, Mahyar, and Negar Jahanbakhsh Javidi. "A Multi-Perspective Framework Established on Diffusion of Innovation (DOI) Theory and Technology, Organization and Environment (TOE) Framework Toward Supply Chain Management System Based on Cloud Computing Technology for Small and Medium Enterprises ." *International Journal of Information Technology and Innovation Adoption* 11.8 (2023): 1217-1234
- [14] Amini, Mahyar and Ali Rahmani. "Agricultural databases evaluation with machine learning procedure." *Australian Journal of Engineering and Applied Science* 8.6 (2023): 39-50
- [15] Amini, Mahyar, and Ali Rahmani. "Machine learning process evaluating damage classification of composites." *International Journal of Science and Advanced Technology* 9.12 (2023): 240-250
- [16] Amini, Mahyar, Koosha Sharifani, and Ali Rahmani. "Machine Learning Model Towards Evaluating Data gathering methods in Manufacturing and Mechanical Engineering." *International Journal of Applied Science and Engineering Research* 15.4 (2023): 349-362.
- [17] Sharifani, Koosha and Amini, Mahyar and Akbari, Yaser and Aghajanzadeh Godarzi, Javad. "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model." *International Journal of Science and Information System Research* 12.9 (2022): 20-44.
- [18] Amini, Mahyar, et al. "MAHAMGOSTAR.COM AS A CASE STUDY FOR ADOPTION OF LARAVEL FRAMEWORK AS THE BEST PROGRAMMING TOOLS FOR PHP BASED WEB DEVELOPMENT FOR SMALL AND MEDIUM ENTERPRISES." *Journal of Innovation & Knowledge*, ISSN (2021): 100-110.
- [19] Amini, Mahyar, and Aryati Bakri. "Cloud computing adoption by SMEs in the Malaysia: A multi-perspective framework based on DOI theory and TOE framework." *Journal of Information Technology & Information Systems Research (JITISR)* 9.2 (2015): 121-135.

- [20] Amini, Mahyar, and Nazli Sadat Safavi. "A Dynamic SLA Aware Heuristic Solution For IaaS Cloud Placement Problem Without Migration." *International Journal of Computer Science and Information Technologies* 6.11 (2014): 25-30.
- [21] Amini, Mahyar. "The factors that influence on adoption of cloud computing for small and medium enterprises." (2014).
- [22] Amini, Mahyar, et al. "Development of an instrument for assessing the impact of environmental context on adoption of cloud computing for small and medium enterprises." *Australian Journal of Basic and Applied Sciences (AJBAS)* 8.10 (2014): 129-135.
- [23] Amini, Mahyar, et al. "The role of top manager behaviours on adoption of cloud computing for small and medium enterprises." *Australian Journal of Basic and Applied Sciences (AJBAS)* 8.1 (2014): 490-498.
- [24] Amini, Mahyar, and Nazli Sadat Safavi. "A Dynamic SLA Aware Solution For IaaS Cloud Placement Problem Using Simulated Annealing." *International Journal of Computer Science and Information Technologies* 6.11 (2014): 52-57.
- [25] Sadat Safavi, Nazli, Nor Hidayati Zakaria, and Mahyar Amini. "The risk analysis of system selection and business process re-engineering towards the success of enterprise resource planning project for small and medium enterprise." *World Applied Sciences Journal (WASJ)* 31.9 (2014): 1669-1676.
- [26] Sadat Safavi, Nazli, Mahyar Amini, and Seyyed AmirAli Javadinia. "The determinant of adoption of enterprise resource planning for small and medium enterprises in Iran." *International Journal of Advanced Research in IT and Engineering (IJARIE)* 3.1 (2014): 1-8.
- [27] Sadat Safavi, Nazli, et al. "An effective model for evaluating organizational risk and cost in ERP implementation by SME." *IOSR Journal of Business and Management (IOSR-JBM)* 10.6 (2013): 70-75.
- [28] Safavi, Nazli Sadat, et al. "An effective model for evaluating organizational risk and cost in ERP implementation by SME." *IOSR Journal of Business and Management (IOSR-JBM)* 10.6 (2013): 61-66.
- [29] Amini, Mahyar, and Nazli Sadat Safavi. "Critical success factors for ERP implementation." *International Journal of Information Technology & Information Systems* 5.15 (2013): 1-23.
- [30] Amini, Mahyar, et al. "Agricultural development in IRAN base on cloud computing theory." *International Journal of Engineering Research & Technology (IJERT)* 2.6 (2013): 796-801.
- [31] Amini, Mahyar, et al. "Types of cloud computing (public and private) that transform the organization more effectively." *International Journal of Engineering Research & Technology (IJERT)* 2.5 (2013): 1263-1269.
- [32] Amini, Mahyar, and Nazli Sadat Safavi. "Cloud Computing Transform the Way of IT Delivers Services to the Organizations." *International Journal of Innovation & Management Science Research* 1.61 (2013): 1-5.
- [33] Abdollahzadegan, A., Che Hussin, A. R., Moshfegh Gohary, M., & Amini, M. (2013). The organizational critical success factors for adopting cloud computing in SMEs. *Journal of Information Systems Research and Innovation (JISRI)*, 4(1), 67-74.
- [34] Khoshraftar, Alireza, et al. "Improving The CRM System In Healthcare Organization." *International Journal of Computer Engineering & Sciences (IJCES)* 1.2 (2011): 28-35.
- [35] Amini, Mahyar and Ali Rahmani. "How Strategic Agility Affects the Competitive Capabilities of Private Banks." *International Journal of Basic and Applied Sciences* 10.01 (2023): 8397-8406.
- [36] Amini, Mahyar and Ali Rahmani. "Achieving Financial Success by Pursuing Environmental and Social Goals: A Comprehensive Literature Review and Research Agenda for Sustainable Investment." *World Information Technology and Engineering Journal* 10.04 (2023): 1286-1293.
- [37] Yun, Chidi, et al. "The Significance of Information Systems in Enhancing Strategic Agility within Supply Chain Context: A Case Study of Telecommunications Industry." *International Journal of Engineering and Applied Sciences* 11.02 (2023): 67-74.
- [38] Motalo, Kubura, et al. "The Competitive Edge of Strategic Agility in Airlines through Embracing Novel Technologies ." *International Journal of Technology and Scientific Research* 12.04 (2023): 789-796.
- [39] Olutola, Tomiloba, et al. "The Intermediate Function of Sustainable Dynamic Capabilities in the Association between Social Customer Relationship Management and Sustainable Competitive Advantage." *Asian Journal of Basic and Applied Sciences* 10.04 (2023): 97-104.
- [40] Balen, John, et al. "Essential Determinants for Assessing the Strategic Agility Framework in Small and Medium-sized Enterprises (SMEs) ." *European Journal of Scientific and Applied Sciences* 10.04 (2023): 2124-2129.
- [41] Chen, Lee, et al. "Categorization of Surgical Complications using Computer Vision Technique ." *American-Eurasian Journal of Scientific Research* 11.04 (2023): 5994-6001.
- [42] Zhang, Lixuan, et al. "A Hybrid Forecasting Method for Anticipating Stock Market Trends via a Soft-Thresholding De-noise Model and Support Vector Machine (SVM) ." *World Basic and Applied Sciences Journal* 13.03 (2023): 597-602.
- [43] Xiang, Zheng, et al. "Efficiency of Risk Classification and Insurance Market Regulation through the Integration of Advanced Technology ." *World Engineering and Applied Sciences Journal* 14.02 (2023): 3986-3992.