The time series clustering of stock price in LQ45 index and its financial performance analysis

Afrimayani, D Devianto

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Andalas University, Limau Manis Campus, Padang 25163, Indonesia

Corresponding author: ddevianto@sci.unand.ac.id

Abstract. Stock price in time series data can be analyzed with the clustering method by using the autocorrelation function distance measurement method. The purpose of this study is to cluster stock prices with the same characteristics and analyze companies' financial performance in each cluster and provide a reference to investors in making choices to develop their investments. This study uses time-series data from stock prices in the LQ45 index, which is continuously available and registered from January 2010 to December 2019, as many as 32 companies. The results of this study are obtained 3 clusters, where the first cluster contains 17 stocks, the second cluster contains six stocks, and the third cluster contains nine stocks. After clustering, the financial performance of each cluster is analyzed in 2019. The companies' financial performance in the first cluster shows that the company has proper inventories, total assets, profit for the period, and can get great benefits. The third cluster shows that the company has a relatively good current ratio and demonstrates its ability to generate profits from the high assets and equity used. Meanwhile, the second cluster has quite high receivables.

1. Introduction

The capital market is an intermediary institution between those who have excess funds and those who need funds, and as an institution that encourages efficient fund allocation [1]. Other than that, the capital market is a place for the long-term financial asset. Capital market instruments that have been traded very often are stocks. Stock is proof of participation as owners in a company. The stock index is a statistical measure that reflects the overall price movement of a set of stocks selected based on specific criteria and methodologies, and it is evaluated regularly [2]. One of the stock indexes is LQ45. LQ45 stock is indices that measure the price performance of 45 stocks that have high liquidity and large market capitalization and are supported by good corporate fundamentals [3].

The advantage of having stock for investors is to get the capital gains that are obtained every year. Investors can obtain all these profits if the company has excellent financial performance [4]. The company's financial performance can be seen from the high value of the company. High company value will be able to maximize shareholder wealth. If the stock price is high, the company's value will be high too, and this also reflects the high level of investor confidence in the company that issued the stock. The demand and supply of shares will also significantly affect fluctuations in stock prices [5].

Number of stock investors in Indonesia is still low recently, only 1 percent of Indonesia's population. Increasing the number of stock investors is expected by the Indonesia Stock Exchange [6]. This is because an increase in the name of stock investors can withstand the turmoil of the economic...
crisis. The lack of public knowledge about the valuation of shares is a significant problem, which is the cause of the minimal number of stock investors in Indonesia. Therefore, it is necessary to cluster companies to find clusters of stock that have the same characteristics.

In the statistical concept, and analysis used to classify objects such as stock price is cluster analysis. That is a statistical analysis technique to sort objects into clusters based on similarities of objects based on various characteristics [7]. If the data used is time-series data, researchers must use cluster analysis specifically for time series data [8]. In this paper, the study focuses more on how the cluster analysis process uses time-series data because previous researchers were more dominant in conducting cluster analysis using cross-section data. For stock prices, research has been conducted on nonlinear modeling of IHSG with artificial intelligence [9] and the model of artificial neural network and nonparametric MARS regression for Indonesian composite index [10]. Besides that, the study also assessed companies by using financial performance analysis after the cluster analysis.

The reason for researching the analysis of financial performance is because the information published by the company is the easiest, cheapest, and most sufficient information to be able to describe how the company's condition so far. Investor's consideration in choosing a stock investment can also be seen from the company's financial statements that can estimate the company's state or position and direction. Besides, researchers previously have reviewed the clustering of financial time series data on the international stock market, time series clustering via community detection in networks [11], forecasting the stock price index of the financial company LQ45 using the Autoregressive Integrated Moving Average (ARIMA) method and Vector Autoregressive (VAR), cluster analysis of shares based on the profitability ratio at crisis period, and effects of clustering stock price techniques in portfolio management [12], time series clustering based on linear fuzzy information granules [13], probabilistic representational structures for clustering the time series data [14], Toeplitz inverse covariance based on clustering of multivariate time series data [15], a model based on multivariate time series clustering algorithm [16], and recent techniques of clustering of time series data survey [17].

In general, cluster analysis and financial performance analysis are not the only ways to determine stock price movements and specifications of each company in the LQ45 stock. An example is the company's characteristics can be seen by doing a regression analysis technique, which can be used to estimate stock prices. The same thing in the world of statistics has a role in the distribution of distribution, be it an exponential distribution, where exponential distribution is a unique form of the gamma distribution [18] [19], Poisson distribution [20], Gamma distribution [21], and other distributions. From the placement of these distributions, it can also be seen the characteristics of an object to be examined.

Based on the background described, the problems examined in the development of a company can be identified through cluster analysis to determine the company's financial condition. This can be a problem for investors if they do not know the prospects of future profits to learn the right investment to survive in changing economic conditions. Not all investors have access to see the development of a company's financial situation and obtain the easy to understand this. Therefore, from this description, the researcher will process the stock price data on the LQ45 index with time series clustering and then analyze each cluster's financial performance to obtain the characteristics of each stock.

2. Data and methods
This research will discuss the clustering of companies that are in LQ45 stock based on the stock price from time to time by using time series data cluster analysis. Time series clustering is the most-used approach as an exploratory technique, and also as a subroutine in more complex data mining algorithms, such as rule discovery, classification, indexing, and anomaly detection [22]. Time-series clustering is a challenging issue because first of all, time series data are often far larger than memory size and consequently they are stored on disks. This leads to an exponential decrease in the speed of the clustering process [22].
Next, an analysis of financial performance is carried out in each cluster. Each company in each cluster is calculated financial performance in terms of aspects of inventory, receivables, total assets, profit for the period, operating profit margin, net profit margin, current ratio, return on assets, return on equity, and gross profit margin. The data used is the stock price data on each company. This data is monthly secondary data from January 2010 to December 2019 obtained from publications on the website https://finance.yahoo.com/. The cluster analysis process involved in clustering companies is processed with the help of RStudio. The companies included in the LQ45 stock category are 32 companies, where the company was chosen with consideration of the completeness of the stock price data from January 2010 to December 2019. Company codes that have data that are available continuously as data processing objects include EXCL, GGRM, BBCA, BBNI, BBRI, BMRI, BRPT, BSDE, CTRA, KLBF, PWON, TPIA, UNVR, HMSP, ICBP, INDF, ITMG, TKIM, UNTR, INCO, INKP, JPFA, ASII, BBTN, ADRO, JSMR, MNCN, SCMA, WIKA, LPPF, INDY, and INTP.

In this study, the cluster analysis used is a method with distance measurement Autocorrelation Function (ACF) using the complete linkage method. The steps of this research stage are as follows:

1. Conduct descriptive analysis to describe the movement of stock prices in each company.
2. Conduct cluster analysis using ACF distance on the stock price data as follows:
   a. Get the ACF distance matrix using the following formula [23]:
      \[ d_{ACF}(U,V) = \left( \hat{\rho}_U - \hat{\rho}_V \right)^T \left( \hat{\rho}_U - \hat{\rho}_V \right) \]^1/2 \]  
   \( d_{ACF}(U,V) \) is the autocorrelation distance between U data and V data.
   \( \hat{\rho}_U = (\hat{\rho}_{1U}, \ldots, \hat{\rho}_{LU})^T \) is an autocorrelation vector whose elements are the estimated value of the time series data autocorrelation U for Lag to 1,2,...,L.
   \( \hat{\rho}_V = (\hat{\rho}_{1V}, \ldots, \hat{\rho}_{LV})^T \) is an autocorrelation vector whose elements are the estimated value of the time series data autocorrelation V for Lag to 1,2,...,L.
   b. Combine clusters that have the closest distance.
   c. Compute distance matrix repair with complete linkage method using the following formula [22]:
      \[ d_{(UV)} = \max\{d_{UW}, d_{VW}\} \]  
   \( d_{UW} \) is the distance between object U and object W,
   \( d_{VW} \) is the distance between object V and object W,
   \( d_{UVW} \) is the distance between object UV and object W.
   d. Repeat step b-c, N-1 times, so we get one cluster that contains all objects. N is the number of objects.
   e. Obtain a number of clusters.
   f. Interpret cluster analysis according to the specified method.
3. Conduct an analysis of financial performance in each cluster by calculating the value of the financial performance of each company [24].
   a. Return on Assets (ROA) is measured with formula as follows.
      \[ \text{ROA} = \frac{\text{Profit for the period}}{\text{Total Assets}} \times 100\% \]  
   b. Return on Equity (ROE) is measured with formula as follows
      \[ \text{ROE} = \frac{\text{Profit for the period}}{\text{Total Equity}} \times 100\% \]  
   c. Gross Profit Margin (GPM) is measured with formula as follows
      \[ \text{GPM} = \frac{\text{Gross Profit}}{\text{Total Sales}} \times 100\% \]  
   d. Operating Profit Margin (OPM) is measured with formula as follows
      \[ \text{OPM} = \frac{\text{Operating Income}}{\text{Total Sales}} \times 100\% \]  
   e. Net Profit Margin (NPM) is measured with formula as follows
4. Interpretation of financial performance analysis on each cluster formed.

3. Results and discussions

3.1. Cluster analysis of the company

This section will discuss the data description and cluster analysis process related to clustering the company after it has been processed with the help of RStudio. The following is the development of the stock prices of each company in Indonesia. This chart aims to see the pattern of stock price movements in each company.

![Figure 1. Development of stock prices company in Indonesia.](image)

Figure 1 shows that the two companies, namely GGRM and ITMG, have a pattern of movement and high stock prices compared to other companies. So, it can be said that the two companies are different from other companies in terms of stock prices. It is still difficult for other companies to find commonalities in stock price patterns if only viewed from a chart. Therefore, in this study grouping companies is done by cluster analysis.

The clustering process is begun by considering each company as one cluster. Interpretation of the size of the distance between other companies is the smaller the distance obtained between the two companies, the more similar the characteristics of the two companies. Thus, there will be 32 clusters, each of which consists of one company. Of the 32 clusters, the closest distance to the cluster is seen first. The determination of the distance of each company is calculated based on the ACF distance between companies by the formula given in equation (1).

Based on the distance measurement matrix obtained from the processed RStudio data, the results of the calculation of the ACF distance between BBCA companies and BBNI companies is 0.0774, while the ACF distance between BBCA companies and BBRI companies is 0.0232. This shows that BBCA companies have characteristics that are more similar to BBRI companies compared to BBNI companies. So that the two companies are combined into one cluster; thus, there will be 31 clusters. Then from the merging of the two clusters above the distance, the complete linkage method is improved so that a new distance matrix is obtained. The distance matrix repair is done using the complete linkage method of equation (2). Then the same process is carried out until a cluster containing all the companies is obtained.

The results of the distance calculation for all companies in the distance measurement matrix from the processed RStudio. The interpretation of the size of the distance between other companies is the same as shown previously, where the smaller the distance obtained between the two objects, the more
similar the characteristics of the two objects. The result of the cluster analysis process can be illustrated in the form of a dendrogram. The following is a clustering dendrogram for companies incorporated in LQ45 stock based on time series stock price data:

![Cluster Dendrogram](image)

**Figure 2.** Dendrogram of cluster analysis on ACF distance.

From the dendrogram, it appears that the clusters formed are as many as three clusters. The first cluster results consisted of 17 companies, namely companies with the code EXCL, GGRM, BCCA, BBNI, BBRI, BMRI, BRPT, BSDE, CTRA, KLBF, PWON, TPIA, UNVR, HMSP, ICBP, INDF, and ITMG. The results of the second cluster consisted of only six companies, namely companies with the code TKIM, UNTR, INCO, INKP, JPFA, and ASII. The results of the third cluster consist of 9 companies, namely companies with codes BBTN, ADRO, JSMR, MNCN, SCMA, WIKA, LPPF, INDY, and INTP. Next, look at the pattern of stock price movements in each cluster formed.

![Average Stock Price](image)

**Figure 3.** Average of stock price in each cluster
Based on the graph, it can be seen that in the 1st to 70th months, the average of the stock price in clusters 1 and 2 has fluctuations, and the average of stock price tends to decline. But in cluster 3, the average stock price tends to increase. In the 71st month to the 120th month, the average stock price in clusters 1 and 2 increased, but cluster 3 decreased. The average stock price in cluster 1 tends to be high compared to other clusters. Then, followed by cluster 2 and cluster 3. But in the last 20 months, the average price of cluster 2 stock has increased beyond the average value of cluster 1.

3.2. Analysis of financial performance in each cluster
After the company cluster is obtained, then financial performance analysis is performed on each cluster formed. The general description of the financial performance of the stock price in the LQ45 index in 2019 can be seen in Table 1 and Figure 4.

| Financial Ratios | Cluster 1 (%) | Cluster 2 (%) | Cluster 3 (%) |
|------------------|---------------|---------------|---------------|
| Current Ratio    | 5.26          | 5.31          | 5.42          |
| ROA (%)          | 1.53          | 0.97          | 1.73          |
| ROE (%)          | 2.42          | 1.78          | 2.59          |
| GPM (%)          | 3.71          | 2.74          | 3.74          |
| OPM (%)          | 3.34          | 0.85          | 2.93          |
| NPM (%)          | 2.96          | 2.06          | 2.67          |

| Financial Distribution | Cluster 1 (million) | Cluster 2 (million) | Cluster 3 (million) |
|------------------------|---------------------|---------------------|---------------------|
| Receivables            | 14.85               | 16                  | 15                  |
| Inventories            | 16.09               | 16                  | 14                  |
| Total Assets            | 19                  | 19                  | 18                  |
| Profit for the period  | 15                  | 15                  | 14                  |

This data is the result of natural logarithms from the original data to standardize the data. Based on the table and graph, it can be seen that the companies in cluster 3 have quite high financial performance in the aspect of current ratio, ROA, ROE, and GPM. However, in the element of OPM and NPM, companies in cluster 1 have high values. The analysis shows that the most significant financial performance is in cluster 3. This shows the company’s ability to generate profits from high assets and equity used. This can be seen from that representing the value of the current ratio that is still relatively high from other clusters, which is 5.42%. Seen from the side of OPM and NPM, it is still
classified as low in cluster 3, meaning that companies in cluster 3 cannot yet get high profits. All companies in cluster 3 can maintain their performance, indicated by the value of the current ratio.

Based on the table and graph, it can be seen that the companies in cluster 1 have quite high financial aspects in terms of inventory, total assets, and profits for the period. Then, in the issue of accounts receivable, the company in cluster 2 has a high value. The analysis shows that the most significant financial performance is in cluster 1. This can be seen from what represents the relatively high inventory value of other clusters, namely 16.09 million. In terms of receivables, it has a low cost in clusters 1 and 3. It is mean that companies in clusters 1 and 3 can produce a good performance to the fullest. All companies in cluster 1 can maintain their performance.

4. Conclusion
The results of clusters formed at the distance of ACF obtained as many as three clusters. The results of the first cluster consisted of 17 companies, namely companies with the code EXCL, GGRM, BBCA, BBNI, BBRI, BMRI, BRPT, BSDE, CTRA, KLBF, PWON, TPIA, UNVR, HMSP, ICBP, IND, and ITMG. The second cluster results consisted of only six companies, namely companies with the code TKIM, UNTR, INCO, INKP, JPFA, and ASII. The third cluster results consist of 9 companies, namely, companies with codes BBTN, ADRO, JSMR, MNCN, SCMA, WIKA, LPFF, INDY, and INTP. The first cluster from January 2010 to December 2019 had the highest average stock price compared to the second cluster and the third cluster. In the second cluster, from January 2010 to December 2019, the average value of the stock price was below the average cost of the stock price in the first cluster. Furthermore, the third cluster has the lowest average stock price compared to the previous two clusters. The financial performance of the companies in the first cluster shows that the company has good investment, total assets, profit for the period, and can get great benefits. The financial performance of the companies in the third cluster shows that the company has a relatively good current ratio and shows the company's ability to generate profits from the high assets and equity used. Meanwhile, the financial performance of companies in the second cluster has quite high receivables. Based on its financial performance, the companies in the first cluster have high average stock prices. For companies in the second and third clusters, they should improve their financial performance to have an average stock price high enough to remain in the LQ45 index stock. This study only examines the distance method of time series data cluster analysis, namely the ACF distance using the distance improvement method, namely, complete linkage. For readers who wish to further investigate cluster analysis, other cluster analysis methods can be used, including the single linkage method, the average linkage method, and the ward's method, for other distance methods such as the Complexity Invariant Distance (CID) and the Pearson distance. Besides, it can also use other factors that are thought to affect the financial performance of the LQ45 index.

References
[1] Suryani A and Herianti E 2015 Procedia - Soc. Behav. Sci. 211 634–43
[2] Devianto D and Fadhilla D R 2015 Appl. Math. Sci. 9 2779–87
[3] Devianto D, Maiyastri, Randy, Hamidi M, Maryati S and Ahmad A W 2018 Proc. ICAITI 2018 - 1st Int. Conf. Appl. Inf. Technol. Innov. Towar. A New Paradig. Des. Assist. Technol. Smart Home Care 78–83
[4] Utami W B 2017 Int. J. Econ. Bus. Account. Res. 1
[5] Satryo A G, Rohkmania N A and Diptyana P 2017 Indones. Account. Rev. 6 55
[6] S, Ratnawati T and Moehaditoyo S H 2019 Arch. Bus. Res. 7 108–15
[7] Nazar M F, Maiavasti, Devianto D and Yozza H 2018 Proc. ICAITI 2018 - 1st Int. Conf. Appl. Inf. Technol. Innov. Towar. A New Paradig. Des. Assist. Technol. Smart Home Care 108–13
[8] Alonso A M, Berrendero J R, Hernández A and Justel A 2006 Comput. Stat. Data Anal. 51
762–76

[9] Yollanda M, Devianto D and Yozza H 2018 Proc. ICAITI 2018 - 1st Int. Conf. Appl. Inf. Technol. Innov. Towar. A New Paradig. Des. Assist. Technol. Smart Home Care 85–90

[10] Devianto D, Permathasari P, Yollanda M and Ahmad A W 2020 IOP Conf. Ser. Mater. Sci. Eng. 846

[11] Ferreira L N and Zhao L 2016 Inf. Sci. (Ny). 326 227–42

[12] Malhotra N and Tandon K 2013 IRACST-International J. Res. Manag. Technol. 3 2249–9563

[13] Duan L, Yu F, Pedrycz W, Wang X and Yang X 2018 Appl. Soft Comput. J. 73 1053–67

[14] C J, Prasad M V N K, Nickolas S and Gangadharan G R 2020 Expert Syst. Appl. 145

[15] Hallac D, Vare S, Boyd S and Leskovec J 2018 IJCAI Int. Jt. Conf. Artif. Intell. 2018-July 5254–8

[16] Zhou P-Y and Chan K C C 2014 Springer Int. Publ. Switz. 805–17

[17] Rani S and Sikka G 2012 Int. J. Comput. Appl. 52 1–9

[18] Devianto D, Maiyastri, Oktasari L and Anas M 2015 Appl. Math. Sci. 9 4781–9

[19] Devianto D 2018 J. Phys. Conf. Ser. 1097

[20] Devianto D, Sarah, Yozza H, Yanuar F and Maiyastri 2019 J. Phys. Conf. Ser. 1397

[21] Devianto D, Maiyastri M and Azmi K 2020 Proc. 5Th Int. Symp. Curr. Prog. Math. Sci. 2242 030003

[22] Aghabozorgi S, Seyed Shirkhorshidi A and Ying Wah T 2015 Inf. Syst. 53 16–38

[23] Aggarwal C C and Reddy C K 2014 DATA CLUSTERING Algorithms and Applications (New York: CRC Press is an imprint of Taylor & Francis Group, an Informa business)

[24] IDX 2017 Idx lq45