Stock Clustering Based on Stock Closing Price Movement: For Investor Reference

Hanik Aulia[1], Bagus Sartono[2], Dedi Budiman Hakim[3]

[1]Business School, Bogor Agricultural University, Indonesia

[2]Department of Statistics, Bogor Agricultural University, Indonesia

[3]Department of Economics, Bogor Agricultural University, Indonesia

Abstract: *Global financial crisis that happened on 2008 in United States triggered a fluctuation in money market that spread to all over the world. The research aims to determine clustering on stocks with similar behaviour based on stock closing price movement, before and after global crisis period, used for investor reference in term of investment decision making in order to reduce the risk came from the fluctuation. Time Series Clustering method was used to determine the optimal number of cluster based on pattern resulted by time series data. The research showed that the series of stock data formed three cluster based stock closing price movement for both period, before and after 2008 crisis. The dendogram test for both period gave a value for entanglement 0.4, which was close to zero. This value showed that the cluster formed by the series of stock data for period before global crisis had a good alignment with the cluster formed by the series of stock data for period after global crisis. We can conclude that global crisis gave almost similar impact to stocks in the same cluster. Market capitalisation and leverage can be used by investor as general guidance for initial selection of stocks.*

Keywords: crisis, stock, cluster, initial selection, time series clustering

1. Introduction

Stocks tend to change and have a certain pattern of movement. It depends on the current economic condition. For example, global financial crisis on 2008 in United States triggered a fluctuation in money market that spread to all over the world. Indonesian government was forced to temporary stop all trading activities (blackout) on 8-10 October 2008. An external disturbance can result different movement for each stocks due to the different way of response. Those make stocks might have a pattern that differ from others eventhough the stocks are in the same sector. Figure 1.1 shows the movement pattern of four stocks in financial sector (BBCA, BBNI, BDMN, BNLI). The figure shows that there is a different movement direction (trend) between BBCA and BDMN. While BBCA tends to increase (positive trend), in other hand BDMN tends to decrease (negative trend). BBNI and BNLI have flatten pattern (no trend).

The fluctuation in money market or stock market can increase the investment risk. Diversification might be a rational way to handle the increase of investment risk. The investor can diversify the investment by combining different kinds of investment instruments or a certain investment instrument that consist of different entities for portfolio construction. The investor can combine stocks with different behaviour. Stocks with different behaviour will respond differently to external shock or disturbance (i.e. crisis). Those behaviour can be seen by doing an observation to the pattern of stocks movement.
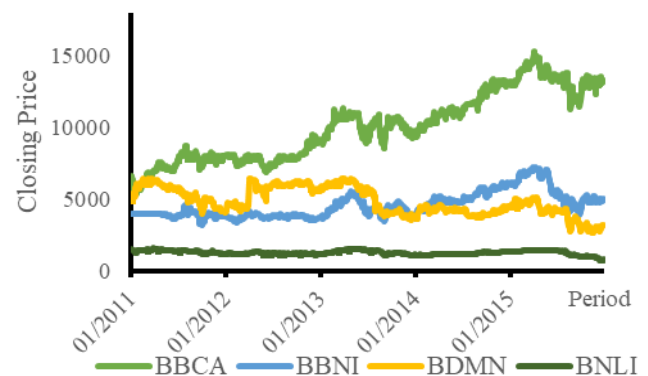

Figure 1: Stocks movements of four companies in financial sector period January 2011 to December 2015.

2. Literature/Theoretical Underpinning

2.1 Stocks, Stocks Movement and Reference Stocks

Bodie et al. (2014) define stocks as securities issued by go public company through initial public offering (IPO) or listed company through seasoned equity offering (SEO). There are two kind of stocks, common stocks and preferred stocks. The differences between two kind of stock lies on the right and responsibility of stock holder. Stocks movement is influenced by internal factor such as dividend payout ratio, price to book value (PBV), earning per share (EPS), return on asset (ROA), debt to equity ratio (DER); and external factor such as kurs, inflation, interest rate, shock (i.e. crisis, political issue, transmission effect, and fed fund rate) (Brigham & Houston 2009; Patar, Darminto, Saifi 2014; Chakrabarti 2011; Liew & Rowland 2016).

There are three variables that often used by investor as consideration, they are company sector, market capitalisation of company, and debt to equity ratio (DER). Company sector

provides information about what type of business the company in, how and when the company will make profit. Market capitalisation show how well company being appraised in the market. It is divide into 3 categories : Small Cap (< Rp 1 T), Mid Cap (Rp 1 T – Rp 10 T) , and Big Cap (>Rp 10 T). DER can reflect the level of leverage of company. So that we will know the structure of company operational financing, weather come from capital or debt or both with certain portion.

2.2 Clustering for Initial Selection of Stocks

The previous researches suggested to use clustering for portfolio selection such as Da Costa et al. (2005), Nanda et al (2010), dan Wang (2011). The researchers used fundamental variables such as return, risk, earning-price ratio, book-value price ratio, sales-price ratio, sales-number of stock ratio, dividend yield, earnings per share, enterprise value-earnings before interest, taxes, depreciation, and amortization (EBITDA) ratio, net asset value per share, total asset turnover ratio, business growth rate, and liquidity ratio. Clustering was used as guidance constructing the portfolio. Cluster was an initial selection to form stocks combination with optimum risk and return. Da Costa et al. (2005) did something a little bit different from two researchers, they divided the period into two time frame. This aimed to see if cluster resulted from first time frame still eligible to be applied to second time frame with respective composition of risk and return of each cluster.

Marvin (2015) did a modification by dividing the period of observation using global crisis on 2008. Marvin (2015) concluded that clustering method application on period of data before and after crisis more optimum than on crisis period due to a systemic risk that can't be eliminated by diversification.

Liao et al. (2007) and Peachavanish (2016) did clustering using stocks price movements. The research resulted a similar conclusion with others, that clustering can help investor for initial selection of stocks in terms of potfolio construction.

3. Methodology

This research used secondary data of daily stocks closing price from January 2003 to December 2007 (before global crisis period) and January 2009 to December 2016 (after global crisis period). The stocks have been listed in Indonesian Stock Exchange before 1 January 2003 and the status is active according to Jakarta Stock Exchange rules (has trading frequency equal to or more than 75 times in three months).

One of data mining method that can be used to observe the stocks behaviour or the pattern of stock movement is clustering. This technical analysis method can recognize the image pattern of data series (Han *et al*. 2012) by identifying the stocks with similar behaviour or pattern of movement. Clustering process will be applied to time series data of stocks for period January 2003 until December 2016. The

process will also apply a cut off time that is global financial crisis on 2008 in order to see if there is a different stocks behaviour before crisis compared to stocks behaviour after crisis. The information resulted from this process can give the investor such a reference that can be used as consideration in decision making process for investment.

3.1 Time Series Clustering

Clustering is a process of partitioning a set of observation data into subsets so that the objects with high similarity will be in the same subset or cluster, while the objects with high dissimilarity will be in the different subset or cluster (Han et al. 2012). Clustering is one of the important data mining methods used to identify pattern of similar objects within a data set of interest (Kassambara 2017). In general there are four methods for clustering process, they are Partitioning, Hierarchical, Density-based, and Grid-based method. Keogh & Kasetty 2003 and Keogh & Lin 2005 suggested to use hierarchical method for better visualization and the general of process so that we don't need to determine the number of cluster, the algorithm will.

Clustering can be applied to time series data with a certain method of distance measurement for time series data. Dynamic Time Warping (DTW) is one of distance measurement method that often used for time series data (Niennattrakul & Ratanamahatana 2007). DTW use shape-based concept for similarity measurement instead of one-to-one mapping. Let two set data of time series $Q = q_1, q_2, …, q_i, …, q_n$ and $C = c_1, c_2, …, c_i, …, c_n$ with $0 < i \leq n$ and $0 < j \leq m$. Distance matrix formed by the total of distance square between q_i and $c_{j;}$ and minimum total of three surrounding neighbour (i,j).

$$e_{ij} = d_{ij} + \min\left(e_{(i-1)(j-1)}, e_{(i-1)j}, e_{i(j-1)}\right)$$

with $d_{ij} = \left(q_i + c_j\right)^2$ and e_{ij} the element (i,j) from distance matrix. *Optimal path* is a *path* that give minimum kumulatif distance on (n,m). Minimum kumulatif distance is defined as :

$$D_{DTW}(Q, C) = \min_{\forall w \in P} \left\{ \sqrt{\sum_{k=1}^{K} d_{w_k}} \right\}$$

with P is a set of all possible warping path, K is the length of warping path, and w_k is k elemen of (i,j) of warping path (Niennattrakul & Ratanamahatana 2007).

3.2. Measuring the Goodness-of-fit of Cluster

In the end of clustering process, we need to know if the cluster is good enough or not. Evaluation process will show the quality of clustering. One of measurement method to calculate the goodness-of-fit of cluster that often used for time series data is silhouette coefficient. This method combine two concepts of internal evaluation, how close (well-clustered) the objects within cluster (cluster cohesion concept) and how far (well-separated) the objects between cluster (cluster separation concept) (Rousseeuw 1987). Let D a data set of n object divided into k cluster, $C_1, …, C_k$. Calculate $a(o)$ for each $o \epsilon D$, the average of distance between o to other objects in the same cluster. Calculate

$b(o)$, the minimum average of distance between o to other objects in different cluster. Let $C_i (1 \leq i \leq k)$, then

$$a(o) = \frac{\sum_{o' \in C_i, o \neq o'} dist(o, o')}{|C_i| - 1}$$

and

$$b(o) = \min_{C_j : 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} dist(o, o')}{|C_j|} \right\}.$$

The silhouette coefficient of o can be defined as :

$$s(o) = \frac{b(o) - a(o)}{max\{a(o), b(o)\}}.$$

The value of silhouette coefficient should be $-1 \leq s(o) \leq 1$. The cluster is well-clustered when the value of silhouette coefficient is close to 1 (Han *et al*. 2012; Rousseeuw 1987).

4. Results/Findings

4.1 Stocks Clustering before Global Crisis

The determination of the number of cluster use two methods, Elbow (Figure 2) and Silhouette method (Figure 3), as comparation. Both of methods give the same result. Elbow chart shows that sum square of error decreasing and then start to become stabil at k = 3. This means the optimum number of cluster based on Elbow method is 3. Silhouette chart shows that the maximum value of silhouette coefficient is at k = 3. This means the optimum number of cluster based on silhouette method is 3. The silhouette coefficient also will be used for clustering evaluation. As seen on Figure 2, the maximum value of silhouette coefficient is 0.48, which is close to 1, so that we can conclude that the series of data before global crisis is well-clustered. The clustering process forms three cluster that consist of 77 stocks for cluster 1, 29 stocks for cluster 2, and 27 stocks for cluster 3.

Table 1 shows the sector composition of each cluster for period of data before global crisis. None of cluster dominated by certain sector. The maximum number of cluster 1 is 16 from Trade, Services and Investment sector, it's only 21% compare to total member of respective cluster. The maximum number of cluster 2 is 8 from Trade, Services and Investment sector also, it's about 28% compare to total member of respective cluster. It happens also in cluster 3, the maximum number is 7 from Basic Industry and Chemicals sector, which is about 26% compare to total member of respective cluster. All of clusters are constructed by almost all of sectors except cluster 2, it does not has Infrastructure, Utilities and Transportation sector as a constructor.

Table 1 also shows the composition of each cluster for period of data before global crisis based on the category of market capitalisation. As seen on the Table, the portion of Mid Cap stocks in cluster 1, which is 40% (31 stocks), is almost similar to the portion of Big Cap stocks, which is 43% (33 stocks). This means cluster 1 mostly constructed by companies which the market capitalisation lies on middle-up level. In contrary, cluster 2 and 3 mostly constructed by Small Cap companies, 55% (16 stocks) for cluster 2 and 59% (16 stocks) for cluster3. This means that cluster 2 and 3 mostly constructed by companies which the market capitalisation lies on middle-low level.

As seen on Table 1, cluster 1 and 2 dominated by stock of companies which have low level of leverage. The thing to be noted is the composition gap between high and low level of leverage is not quite different. This means both of cluster are constructed by companies in all level of leverage with a little bit domination of low level of leverage. Cluster 3 has different result, it's dominated by companies which have high level of leverage. Cluster constructors are mostly companies with higher debt for operational financing.
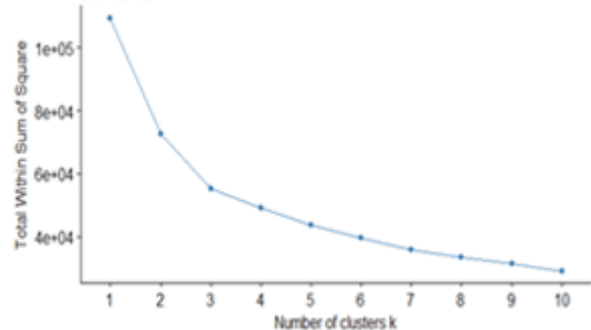

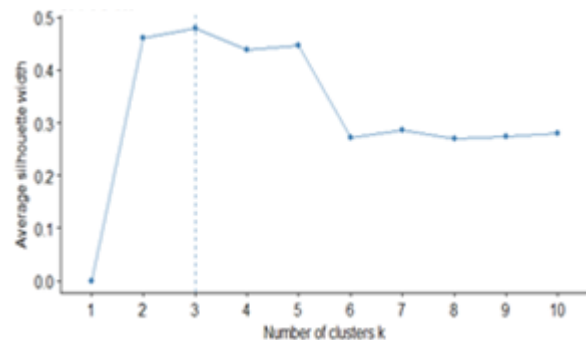**Figure 2:** Number of k determination with Elbow method, before global crisis.

Figure 3: Number of k determination with Silhouette method, before global crisis.

The determination of the number of cluster use two methods, Elbow and Silhouette method (Figure 2), as comparation. Both of methods give the same result. Elbow chart shows that sum square of error decreasing and then start to become stabil at k = 3. This means the optimum number of cluster based on Elbow method is 3. Silhouette chart shows that the maximum value of silhouette coefficient is at k = 3. This means the optimum number of cluster based on silhouette method is 3. The silhouette coefficient also will be used for clustering evaluation. As seen on Figure 2, the maximum value of silhouette coefficient is 0.48, which is close to 1, so that we can conclude that the series of data before global crisis is well-clustered. The clustering process forms three cluster that consist of 77 stocks for cluster 1, 29 stocks for cluster 2, and 27 stocks for cluster 3.

Table 1 shows the sector composition of each cluster for period of data before global crisis. None of cluster dominated by certain sector. The maximum number of cluster 1 is 16 from Trade, Services and Investment sector, it's only 21% compare to total member of respective cluster. The maximum number of cluster 2 is 8 from Trade, Services and Investment sector also, it's about 28% compare to total member of respective cluster. It happens also in cluster 3, the maximum

number is 7 from Basic Industry and Chemicals sector, which is about 26% compare to total member of respective cluster. All of clusters are constructed by almost all of sectors except cluster 2, it does not has Infrastructure, Utilities and Transportation sector as a constructor.

Table 1 also shows the composition of each cluster for period of data before global crisis based on the category of market capitalisation. As seen on the Table, the portion of Mid Cap stocks in cluster 1, which is 40% (31 stocks), is almost similar to the portion of Big Cap stocks, which is 43% (33 stocks). This means cluster 1 mostly constructed by companies which the market capitalisation lies on middle-up level. In contrary, cluster 2 and 3 mostly constructed by Small Cap companies, 55% (16 stocks) for cluster 2 and 59% (16 stocks) for cluster3. This means that cluster 2 and 3 mostly constructed by companies which the market capitalisation lies on middle-low level.

As seen on Table 1, cluster 1 and 2 dominated by stock of companies which have low level of leverage. The thing to be noted is the composition gap between high and low level of leverage is not quite different. This means both of cluster are constructed by companies in all level of leverage with a little bit domination of low level of leverage. Cluster 3 has different result, it's dominated by companies which have high level of leverage. Cluster constructors are mostly companies with higher debt for operational financing.

Table 1: The member of cluster before global crisis based on Company Sector, Category of Market Capitalisation, and level of Leverage

Company Differentiation	Cluster		
	1	2	3
By Company Sector:			
Agriculture	3	1	1
Basic Industry And Chemicals	14	5	7
Consumer Goods Industry	11	5	1
Finance	12	5	4
Infrastructure, Utilities And Transportation	3		2
Mining	3	2	6
Miscellaneous Industry	5	2	2
Property, Real Estate & Building Construction	10	1	3
Trade, Services & Investment	16	8	1
By Category of Market Capitalisation:			
Big Cap	33	2	6
Mid Cap	31	11	5
Small Cap	13	16	16
By Level of Leverage:			
Tinggi	34	14	17
Rendah	43	15	10

4.2 Stocks Clustering after Global Crisis

The determination of the number of cluster use two methods also, Elbow (Figure 4) and Silhouette method (Figure 5), as comparation. Both of methods give the same result with the previous period of data. Elbow chart shows that sum square of error decreasing and then start to become stabil at k = 3. This means the optimum number of cluster based on Elbow method is 3. Silhouette chart shows that the maximum value of silhouette coefficient is at k = 3. This means the optimum number of cluster based on silhouette method is 3. The

silhouette coefficient also will be used for clustering evaluation. As seen on Figure 5, the maximum value of silhouette coefficient is 0.28, which is close to 1, so that we can conclude that the series of data after global crisis is well-clustered. The clustering process forms three cluster that consist of 52 stocks for cluster 1, 49 stocks for cluster 2, and 32 stocks for cluster 3.
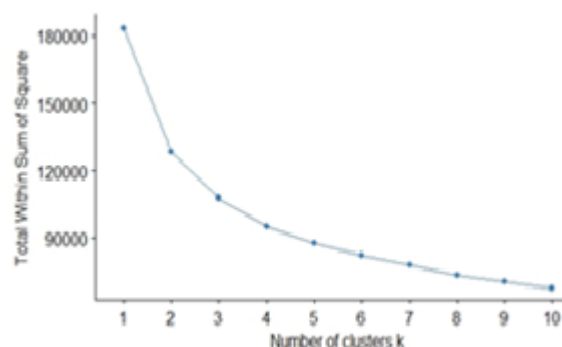

Figure 4: Number of k determination with Elbow method, after global crisis.
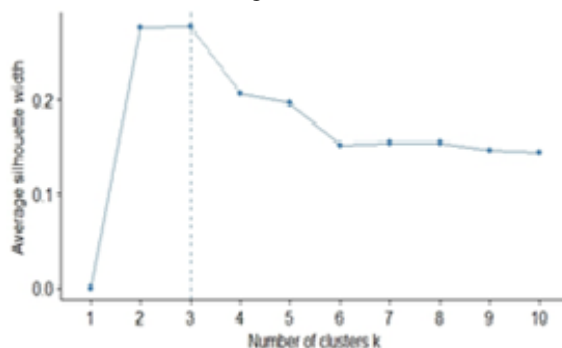

Figure 5: Number of k determination with Silhouette method, after global crisis.

Table 2 shows the sector composition of each cluster for period of data after global crisis. It has similar result with the previous period, none of cluster dominated by certain sector. The maximum number of cluster 1 is 11 from Consumer Goods Industry sector, it's only 21% compare to total member of respective cluster. The maximum number of cluster 2 is 14 from Trade, Services and Investment sector, it's about 29% compare to total member of respective cluster. It happens also in cluster 3, the maximum number is 8 from Basic Industry and Chemicals sector, which is only about 25% compare to total member of respective cluster. All of clusters are constructed by almost all of sectors except cluster 1, the cluster does not has Mining sector as a constructor.

Table 2 also shows the composition of each cluster for period of data after global crisis based on the category of market capitalisation. The portion of Mid Cap stocks in cluster 1 is as same as the portion of Big Cap stocks, which is 42% (22 stocks). This means cluster 1 mostly constructed by companies which the market capitalisation lies on middle-up level. For cluster 2, all level of market capitalisation have almost the same portion, no significant differences. In contrary, cluster 3 dominated by Small Cap companies, which is 56% (18 stocks). This means that cluster 3 mostly constructed by companies which the market capitalisation lies on low level.

As seen on Table 2, Cluster 1 dominated by stock of companies which have low level of leverage. Cluster 2 and 3 have different result, both of clusters dominated by companies which have high level of leverage. The thing to be noted is the composition gap between high and low level of leverage is not quite different. This means that the companies from all level of leverage can be constructor of all clusters formed.

Table 2: The member of cluster after global crisis based on Company Sector, Category of Market Capitalisation, and level of Leverage

Company Differentiation	Cluster		
	1	2	3
By Company Sector:			
Agriculture	2	2	1
Basic Industry And Chemicals	8	10	8
Consumer Goods Industry	11	3	3
Finance	9	7	5
Infrastructure, Utilities And Transportation	1	3	1
Mining		5	6
Miscellaneous Industry	4	2	3
Property, Real Estate & Building Construction	9	3	2
Trade, Services & Investment	8	14	3
By Category of Market Capitalisation:			
Big Cap	22	14	5
Mid Cap	22	16	9
Small Cap	8	19	18
By Level of Leverage:			
Tinggi	21	25	19
Rendah	31	24	13

4.3 Cluster Alignment

Clustering on both periods, before and after global crisis, result the same number of cluster (k=3). In order to know if the cluster change due to global crisis, we calculate entanglement value of both dendogram (before and after global crisis). The entanglement value (0.4, close to zero) let us know that the cluster resulted from the process have a good alignment. This means that global crisis happened in 2008 gave almost similar impact to stocks which in the same cluster. So that the behaviour of stocks have just a little bit change which is not significant to change the whole structure of cluster.

4.4. Behaviour Changing of Stocks along Global Crisis

Table 3: Cross tabulation of stocks position, before and after global crisis

Before Crisis	After Crisis			Total
	1	2	3	
1	41	33	3	77
2	10	11	8	29
3	1	5	21	27
Total	52	49	32	133

Table 3 shows the position of stocks before global crisis compare to after global crisis. As seen on table, 73 stocks or 55% are remain stay together in the same cluster. The behaviour changing can be seen on Figure 6 until Figure 11. We can the trend changing on cluster 3 before global crisis compare to after global crisis, from positive trend turn into negative trend. The stocks which are remain stay together will have that changing pattern. Those 21 stocks might have serious impact of global crisis. For further analysis, we spot two different patterns of stocks behaviour. Let A be a set of stocks which are remain stay together in the same cluster and have positive trend regardless the period (41 stocks of cluster 1 before global crisis and remain stay in the same cluster after global crisis). Let B be a set of stocks which are remain stay together in the same cluster with positive trend before global crisis that turn into negative trend after global crisis (21 stocks of cluster 3).
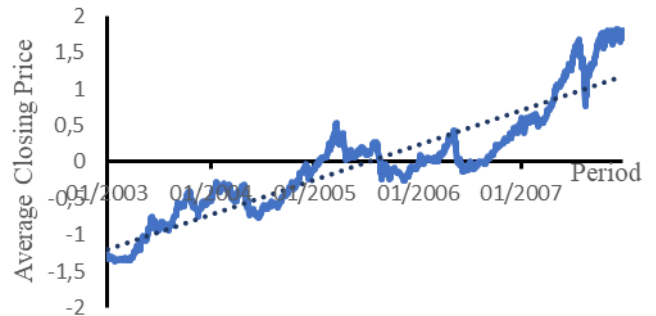

Figure 6: The chart of average stocks closing price for cluster 1 before global crisis.
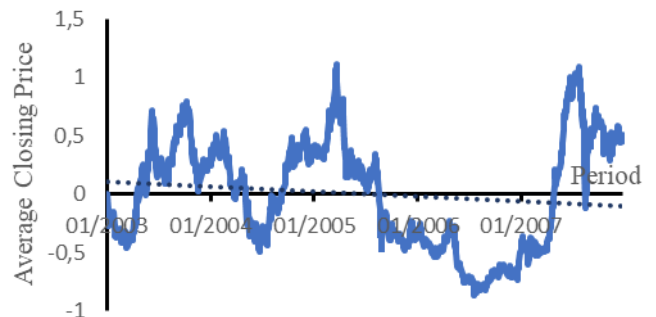

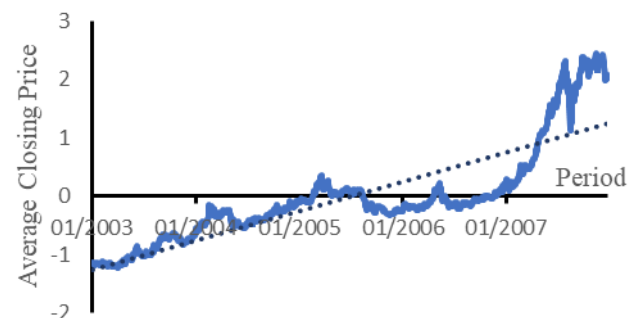**Figure 7:** The chart of average stocks closing price for cluster 2 before global crisis.

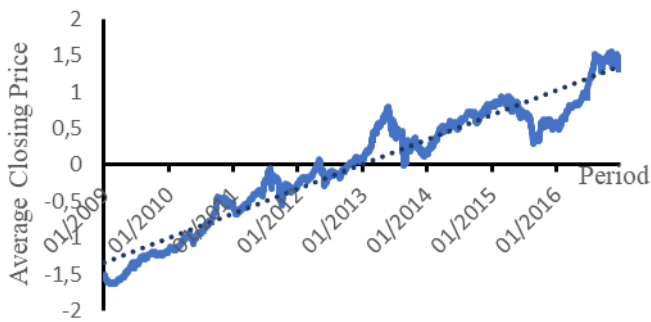**Figure 8:** The chart of average stocks closing price for cluster 3 before global crisis.

Figure 9: The chart of average stocks closing price for cluster 1 after global crisis.
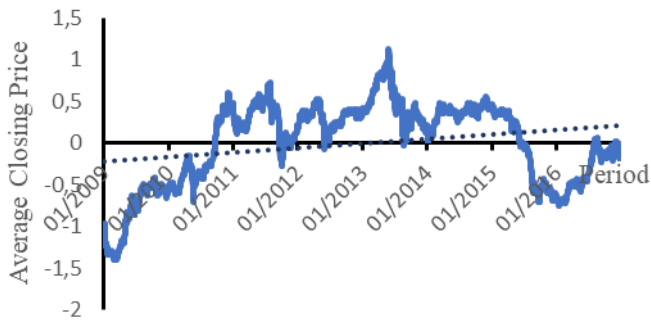

Figure 10: The chart of average stocks closing price for cluster 2 after global crisis.
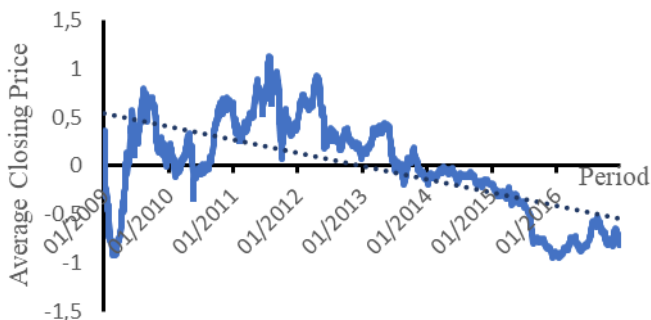

Figure 11: The chart of average stocks closing price for cluster 2 after global crisis.

Figure 12 shows that A have a positive trend since before crisis happened until the crisis ended. This means the A team has less sensitivity to the global crisis. On the contrary, B team has a changing on the behaviour, from positive trend turn into negative trend (Figure 13). The global crisis might give a shock to the stocks on B. We do test on slope of each stocks to see if the trend changing is significant. The test result probability 0.015 which is less than alpha 0.05, then we can conclude that the trend changing between period before and after global crisis is significantly different.

Let us see the profile of A and B. As seen on Table 4, we can't find the domination of a certain sector in A or B set. It implies that we can't use company sector to predict the behaviour and the changing behaviour of stocks if crisis occurred in terms of portfolio construction. Table 4 also shows the composition of A and B set based on the category of market capitalisation. The portion of Mid Cap stocks (41% - 17 stocks) in A set is close to the portion of Big Cap stocks

(49% - 20 stocks). This means A set mostly constructed by companies which the market capitalisation lies on middle-up level. In contrary, B set dominated by Small Cap companies, which is 57% (12 stocks). This means that B set mostly constructed by companies which the market capitalisation lies on low level. Table 4 gives the information of cluster composition on A and B set based on level of leverage of each company. A set dominated by stock of companies which have low level of leverage (59% - 24 stocks). B set has different result, it is dominated by companies which have high level of leverage (71% - 15 stocks). Cluster constructors of B set are mostly companies with higher debt for operational financing.
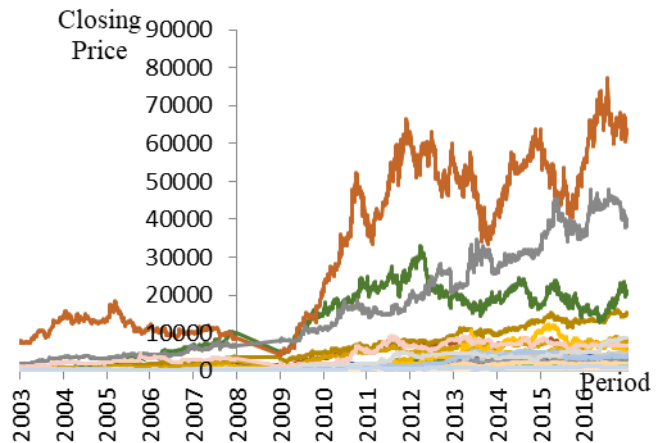

Figure 12: The chart of each stocks on A before and after global crisis
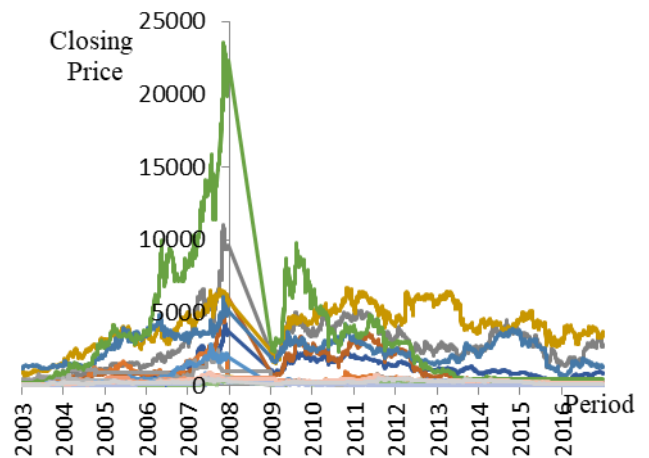

Figure 13: The chart of each stocks on B before and after global crisis

More from Table 4, we will find example of the changing of stocks behaviour regarding to the global crisis, mostly from Basic Industry and Chemicals sector and Mining sector. From Basic Industry and Chemicals sector we find 6 companies have a changing in their behaviour, from positive trend before global crisis period then turn into negative trend after global crisis period. They are ETWA, JPRS, ALMI, SRSN, SPMA, and SULI. From Mining sector we find also 6 companies that have a changing in their behaviour regarding to global crisis. They are MITI, CKRA, MEDC, BUMI, ANTM, and INCO. Most of them have a profile that suit to stocks mapping profile of B cluster, they have Middle-Low category of market capitalisation and High level of leverage.

But a different condition happen to ANTM and INCO. Both of companies are Big cap company with low level of leverage.

Both ANTM and INCO got direct effect of global krisis on 2008. The crisis made international nickel price drop at the lowest level. ANTM that has nickel as one of potential profit source and INCO that has nickel as main product and potential profit source should experience a big loss. Later on that made both of companies lost a trust from the investors.

Table 4: The member of A and B cluster based on Company Sector, Category of Market Capitalisation, and level of Leverage

Company Differentiation	Cluster	
	A	B
By Company Sector:		
Agriculture	1	1
Basic Industry And Chemicals	8	6
Consumer Goods Industry	8	1
Finance	6	2
Infrastructure, Utilities And Transportation	1	1
Mining		6
Miscellaneous Industry	3	2
Property, Real Estate & Building Construction	8	1
Trade, Services & Investment	6	1
By Category of Market Capitalisation:		
Big Cap	20	4
Mid Cap	17	5
Small Cap	4	12
By Level of Leverage:		
Tinggi	17	15
Rendah	24	6

5. Discussion

Clustering method was applied to two time frame or period of data, before and after global crises. Clustering was applied to 133 stocks listed in Indonesia Stocks Exchange and still active until respective period of research. The research also excluded period of crisis itself. This refered to marvin (2005), said that diversification on the period of crisis will not give optimum result due to systemic risk can't be eliminated through diversification. Clustering on both time frame let us know the cohesion of stocks. It also gives information about the general characteristic of stocks which have similar behaviour (pattern). The result showed that a disturbance can change the structure of cluster regardless the scale of change. Some stocks might be insensitive or less sensitive than others in facing a shock or disturbance such as crisis.

As seen on Tabel 4, there is A set that consist of stocks with positive trend even after crisis occurred. On the contrary, there is B set that facing trend changing, from positive trend turn into negative trend. Those we can say that A set is less sensitive than B set to crisis, specifically global crisis on 2008.

Identifying the behaviour changing through A and B set, we get a hint that company sector can't be used as guidance for initial selection because the distribution of sector in each set almost uniform. There is no significant differences in term of constructor portion. Market capitalisation can give guidance by take a look at the size. The small cap companies will respond a shock or disturbance (i.e. global crisis) differently compare to the middle cap or big cap companies. Similar logic to level of leverage, the company with low level of leverage will have different movement compare to the company with high level of leverage in responding a shock or disturbance (i.e. global crisis). The company with high level of leverage will be more sensitive when crisis occurred. High level of leverage means the operational financing of company mostly come from debt, so that it will be very sensitive to external factor. On the contrary, low level of leverage means the operational financing mostly come from equity, it's more rigid to external factor.

6. Implication to Research and Practice

The investor can use two main variables, market capitalisation and leverage, as a guidance for initial selection. The guidance is still in general term, but it still useful to set preliminary list of stocks to be included in portfolio. The investor can create the combination of stocks which have different behaviour (pattern) or decide to select the stocks with similar behaviour to construct the portfolio, eventhough it will not be suggested. In stocks selection, it's also depend on the risk appetite of investor.

7. Conclusion

The research explore one of method to do initial selection to construct portfolio. Clustering method is suggested to be used. The main reason is clustering method can do image pattern recognition of stocks well. Based on entanglement value, we know that the global crisis give just a little bit change to structure of cluster, still have a good alignment.

The investors can use market capitalisation and leverage as guidance for initial selection of stocks. One thing to be noted is the guidance is still in general form, so it can be used only to set preliminary list of stocks. The investor still need a further analysis for portfolio construction.

8. Future Research

This result may be useful for the investor in term of portfolio construction in order to minimize the risk and maximize the return. For specific refference such as determining the leading stocks for others and in which time the stocks will be followed by others, it will need a further analysis with more advanced method.

References

[1] Bodie Z, Kane A, Marcus AJ, Jain R. 2014. Investments, Asia Global Edition. New York : McGraw-Hill.
[2] Brigham EF & Houston JF. 2009. Fundamentals of Financial Management 12th edition. Mason : South-Western Cengage Learning.

[3] Chakrabarti G. 2011. Financial Crisis and the Changing Nature of Volatility Contagion in the Asia-Pacific Region. Journal of Asset Management, vol.12, 172-178.

[4] Da Costa Jr N, Cunha J, Da Silva S. 2005. Stock selection based on cluster analysis. Economics Bulletin, 13(1), pp.1-9.

[5] Han J, Kamber M, Pei J. 2012. *Data Mining : Concepts and Techniques, 3th Edition*. Massachusetts : Morgan Kaufmann.

[6] Kassambara A. 2017. *Practical Guide to Cluster Analysis in R*. Marseille : STHDA.

[7] Keogh E, Kasetty S. 2003. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and Knowledge Discovery, vol.7, pp 349–371.

[8] Keogh E, Lin J. 2005. Clustering of time-series subsequences is meaningless: implications for previous and future research. Knowledge and Information Systems, vol.8, pp 154–177.

[9] Liew VKS & Rowland R. 2016. The Effect of Malaysia General Election on Stock Market Returns. Springer Plus, vol.5, 1975.

[10] Nanda SR, Mahanty B, Tiwari MK. 2010. Clustering Indian Stock Market Data for Portfolio Management. Expert Systems with Applications, vol. 37, 8793-8798.

[11] Niennattrakul V & Ratanamahatana C.A.. (2007), On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping. International Conference Multimedia and Ubiquitous Engineering, 07, hal. 733-738.

[12] Patar A, Darminto, Saifi M. 2014. Faktor Internal dan Eksternal yang Mempengaruhi Pergerakan Harga Saham (Studi pada Saham – Saham Indeks LQ45 Periode 2009 – 2013). Jurnal Administrasi Agribisnis, Vol.11, No.1.

[13] Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, vol.20, 53-56.