

# Time-series Clustering of Global Automakers Stock Prices

Akane Murakami <sup>\*</sup>, Yukari Shiota <sup>\*</sup>

## Abstract

In the paper, we describe the stock price data analysis using the hierarchical clustering method named Hierarchical Risk Parity. In the financial engineering, clustering is important for the portfolio development. The data we used are the top 100 global industries' stock prices from 2018 to 2019. The industry field is automobile manufacturing. The countries of the industries are Japan, US and Germany. We analyzed sequentially a bi-month data for the two years. Then we found that when the stock price drastically decreased, there could clearly appear country-based clusters. In the global turmoil period by the US-China trade friction in 2018, we could identify that the Japan cluster and the US cluster appeared with clear boundaries. To verify the hypothesis, we traced the time series changes of the clusters through the two years. As a result, we proved the hypothesis was correct, as far as the period was what we used. In addition, we visualized the resultant country-based clusters of the largest damage period, as a stock price fluctuation plot and the Markowitz's risk-return plot, to see the difference of the three countries' trend.

*Keywords:* Hierarchical Risk Parity, hierarchical clustering, stock price, financial portfolio, automobile manufacturing, Markowitz's risk-return plot.

## 1 Introduction

In the paper, we describe the stock price data analysis using the hierarchical clustering method named Hierarchical Risk Parity. In the financial engineering, clustering is important for the portfolio development. Then many machine-learning based clustering methods have been used in financial field. In the paper, we describe clustering results of stock price data. The industry field is automobile manufacturing. The target data is the global top 100 automakers such as Toyota and Tesla. These global automobile manufacturing companies have been affected by global turmoil. For example, in 2018 and 2019, the companies' stock prices had been severely damaged owing to the US-China trade friction. To illustrate the effects, we plot the Japan and U.S. average stock price movements (See Figure 1). We used there "Nikkei 225 Linked Exchange Traded Fund (ETF)" and "Simple-X NY Dow Jones Indices ETF". Nikkei 225 is the Japanese representative 225 companies. The Dow Jones Industrial Average is a leading stock index of 30 leading companies in the US stock market which comprises various field companies and Dow

---

<sup>\*</sup> Department of Management Faculty of Economics, Gakushuin University, Tokyo, Japan

Jones index is used as a measure of the overall US market. The Simplex is the name of the ETF maker. The two ETFs can move following the movements of Nikkei 225 and Dow Jones. In December 2018, the two ETFs price movements showed a large decrease, as shown in Figure 1.

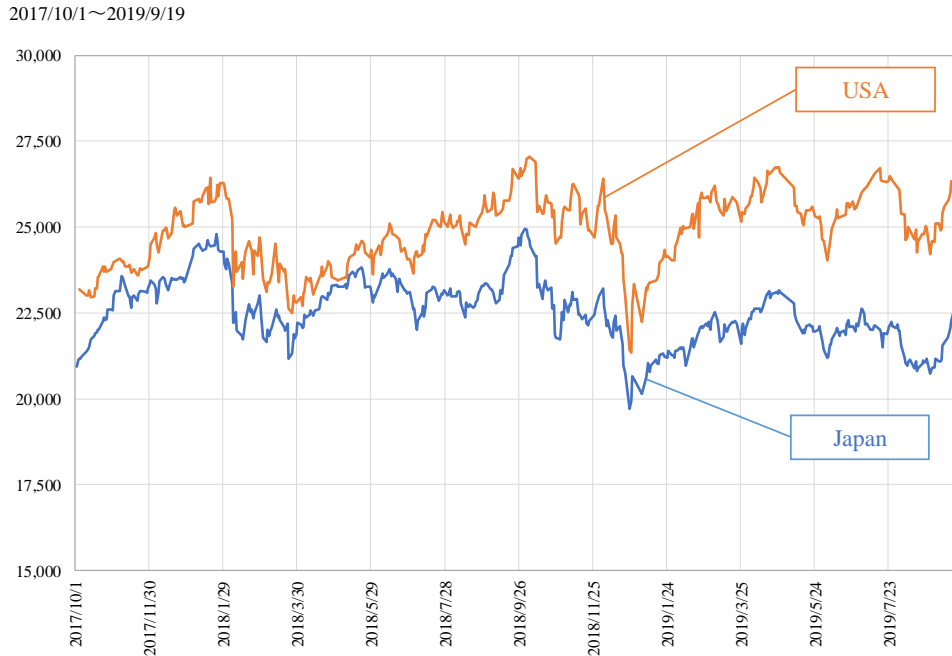


Figure 1: Two ETF price movement concerning US and Japan stock market from 2017 to 2019

We shall utilize the machine learning based clustering methods for the stock price analysis. The clustering methods are used for the portfolio construction in the financial engineering. We conduct the portfolio construction by using the hierarchical clustering. The data we analyzed are the global automakers stock prices of which countries were Japan, US, and Germany(hereafter DE). As the results, we found an interesting relevance between the stock price declines and the clustering results. *When the stock price significantly decreased (which we call turmoil), there appeared the Japan cluster and the US cluster in the result. We could clearly find a country-based cluster.* In other word, at the turmoil, the investment might be determined mainly by the country name, instead of performance and management aspects of an individual company. We make the hypothesis as follows: *When the stock price significantly decreased, a country-based cluster appears.* In the paper, we shall verify the hypothesis.

In the next section we shall explain the hierarchical clustering method we used, and in Section 3, we shall explain related work of time series data clustering. In Section 4, the clustering results will be presented and in Section 5, we shall evaluate the relevance between the stock decline and the clustering. Finally, we shall conclude the paper.

## 2 HRP (Hierarchical Risk Parity) Method

Before beginning this analysis, we shall explain the hierarchical clustering method we used.

In a financial field, for a capital allocation, machine learning approaches have been used actively [1, 2]. Various companies' stock prices have individual return values and the standard deviation values of the returns which are called a risk. Combining the different movement company stocks leads to a risk reducing. The portfolio construction problem is how to increase portfolio

returns while reducing the risk, by making a well-organized company stock set named a portfolio. Then we are interested in similarity among stock price movements. To measure the similarity of the stock price movements of two companies, we use a concept of a distance between them. Companies with short distances have similar stock price fluctuations.

In the field of portfolio construction, recently Prado's proposed HRP (Hierarchical Risk Parity) method has been widely used [3]. One of the main advantage of HRP is an ability in computing a portfolio on an ill-degenerated or even a singular covariance matrix [4]. After HRP, many researches by a hierarchical clustering had been conducted [4-7]. We shall conduct the same clustering approach as HRP.

Let us explain the method. The input data is the matrix of *natural logarithmic return values* (hereafter return) which is defined as follows:

$$G_{i,j} = \ln \left( \frac{S_{i,j}}{S_{i,j-1}} \right)$$

where  $S_{i,j}$  represents the  $i$ -th company's stock price on  $j$ -th day. The number of companies is  $N$  and the number of sales days is  $T$ . The size of the matrix  $G_{i,j}$  is  $N$  times  $T$ . From the matrix  $G_{i,j}$ , we shall make the correlation coefficient matrix  $\rho = \{\rho_{i,j}\}$ . In HRP, the distance  $d$  is defined as follows:

$$d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$$

Then the distance matrix  $D = \{d_{i,j}\}$  is obtained. Next selecting any two distance columns, we shall calculate the Euclid distance as follows:

$$\tilde{d}_{i,j} = \sqrt{\sum_{n=1}^N (d_{n,i} - d_{n,j})^2}$$

Input the matrix  $\{\tilde{d}_{i,j}\}$ , we shall conduct the hierarchical clustering. As the linkage method, HRP uses "single", not "Ward". Therefore we used the single algorithm. The clustering algorithms are explained in the textbooks [8, 9]. Then, using the distances between nodes, we shall sort the distance matrix  $\{\tilde{d}_{i,j}\}$ . Prado calls the matrix seriation "quasi-diagonalization" [10].

### 3 Related Work

In the section, we shall survey related work concerning portfolio construction and clustering methods. Our final objective is to construct a well-organized portfolio. For a long time, the Markowitz's minimum variance portfolio model has been used for the objective [11, 12]. However, given many stocks/assets, the instability of the expected frontiers in the model are likely to become high [10]. Clarke et al. insisted that it was required to incorporate additional constraints in order to achieve its robustness [13]. Prado incorporated the hierarchical structure to the set of stocks/assets in his proposed method called Hierarchical Risk Parity (hereafter HRP)[3,10,14]. The HRP methods uses the inverse variance allocation method which splits a weight in inverse proportion to the subset's variance, because such allocation is theoretically optimal when the covariance matrix is diagonal. As a result, the HRP outperformed the original Markowitz's model [11,12].

Our approach is as follows: (1) make a set of clusters having similar fluctuations, (2) draw the Markowitz's risk and return plot of the representatives/averages of the individual clusters. In the paper, we shall focus on the first step which is a time-series clustering technique and partially the second step. Our objective is to find the companies with similar stock price movement.

There are many kinds of time-series data clustering. Time-series data clustering consists of (1) Whole time-series clustering, (2) Subsequence time-series clustering, and (3) Time point clustering [16]. Among them, our target is (1) Whole time-series clustering which is as clustering of a set of individual time-series with respect to their similarity. The whole time-series clustering has significant two attributes which are (A) *Distance measurement* such as Dynamic Time Warping (DTW), Pearson's correlation coefficient and related distances, and Euclidean distance and (B) *Clustering algorithm* such as hierarchical clustering and partitioning clustering.

Concerning the distance measurement, DTW [17] constructs, regardless of time points, clusters of time-series data with similar patterns of change. For example, to cluster stock price related to different companies which have a common pattern in their stock independent on its occurrence in time-series [18]. The advantage of DTW is its capability to align one point in a time series to multiple points in another one [19, 20]. WDTW (Weighted Dynamic Time Warping) is a weighted version of DTW[21, 22]. There are many researches which used DTW as the distance measurement [23-25]. In HRP that we used, the distance measurement based the correlation coefficient is used.

The clustering method we used is the Hierarchical clustering[26] which constructs a hierarchy of clusters using agglomerative or divisive algorithm. That is a bottom-up method. On the other hand, k-Means [27] which is a representative partitioning method minimizes the total distance between all objects in a cluster from their cluster center.

There are many approaches for portfolio selection using various clustering methods. Massahi et al. compared the original Markowitz's method and cluster-based portfolio optimization methods such as WDTW, autocorrelation coefficient (ACC), and Pearson's correlation coefficient (PCC) [28]. The data they used is 500 stocks in New York Stock Exchange. As a result, the cluster-based portfolio optimization model often (in most cases) outperform the original Markowitz's minimum variance portfolio model. Especially in high-volatility periods, the ACC and WDTW based models mostly lead to superior results, compared to the PCC model. Massahi et al. said that the reason laid in that the cross-correlations were then very unstable measure of similarity. Another comparison of distance measurement of a cluster-based model is conducted by Puspita et al. [29] in which Euclidean distance and DTW are evaluated. The data they used are stocks in Indonesia Sharia Stock Index and Jakarta Islamic Index. They estimated Silhouette index to estimate the clustering quality. Their conclusion was that there was no significant difference on the Silhouette index among the distance measurement methods as far as the data was the Indonesia data.

Although we used the HRP approach in this analysis, a comparison of many clustering methods would be required for our hypothesis verification.

## 4 Automakers Clustering Results

In the section, we shall present the top 100 global automakers' clustering result.

The countries we selected are Japan, US, and Germany (DE). The number of companies in each country is 56(JP), 35(US), and 9(DE). The data is stock prices from 2017/10/1 to 2019/9/19. The two year data is divided by every 2 months; we finally obtained 12 data sets. The stock data is taken from ORBIS database by Bureau Van Dijk. The markets of US, Japan and Germany have

different sale days. Therefore, we have to interpolate the missing data. For the interpolation, we used a linear interpolation.

Let us present a tiny demonstration (See Figure 2). That is a clustering result of the top 10 global automakers such as DAIMLER. The period is just two months from 2018 March to April. In the resultant dendrogram in Figure 2, the horizontal axis shows the linkage distance. For a dendrogram, we can swap some node's left children and right children and it would then still be the same dendrogram [30]. In Figure 2, we can find three clusters corresponding to an individual country; from the bottom they are Japan, Germany, and US. The company name is added on the top a country name abbreviation "J:", "U:", or "D:". Being isolated from the three clusters, Audi (Germany) is located. In Figure 2, we found that the Germany cluster was near to the US cluster than to the Japan cluster.

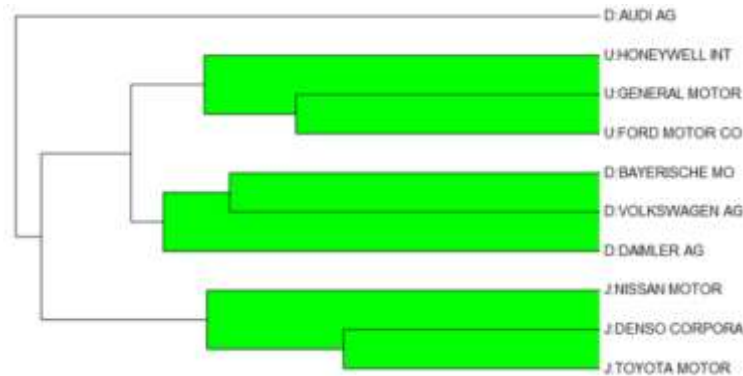


Figure 2: A hierarchical clustering of the top 10 automakers.

Next, we shall now analyze the top 100 global automakers' data. We conducted the clustering sequentially from 2017/10/1 to 2019/9/18 by every 2 months. In Figure 3, the stock market movement and heatmaps of the correlation coefficient matrix are shown. The above-mentioned "Dow Jones Indices ETF" and "Nikkei 225 Linked ETF" are illustrated as the stock market movement there. Among the two, the upper one is one of Dow Jones. In the correlation coefficient matrix, the value "1"s are listed diagonally, because the correlation coefficient to itself is one; In the Figure 3, the value "1" is colored red. Among the 12 results, we selected six periods with the most remarkable appearances which are named "A" to "F".

The most clear clustering by countries was one of 2018/11/25 to 2019/1/23 which was marked "A". The marked "A" period from 2018/11/25 to 2019/1/23 corresponds to the worst decline in the US financial market. The President Trump tweeted on December 25th, 2018 as follows: "The only problem our economy has is the Fed. They don't have a feel for the Market, they don't understand necessary Trade Wars or Strong Dollars or even Democrat Shutdowns over Borders." After his tweet, the US financial market decline started in the morning. The decline was so severe and "npr" reported that as "Dow Suffers Record-Breaking Christmas Eve Losses" (<https://www.npr.org/2018/12/25/680010415/>, Dec. 25th, 2018). And, as terrible as the decline was reported as "Markets see worst Christmas Eve in history, worst December since Great Depression" by The Mercury News on Dec. 24th, 2018 (<https://www.mercurynews.com/2018/12/24/>). The clustering result, we could see clearly the Japan large cluster, the US cluster, and the DE small cluster. The clear cluster appearance may correspond to the significant declines.

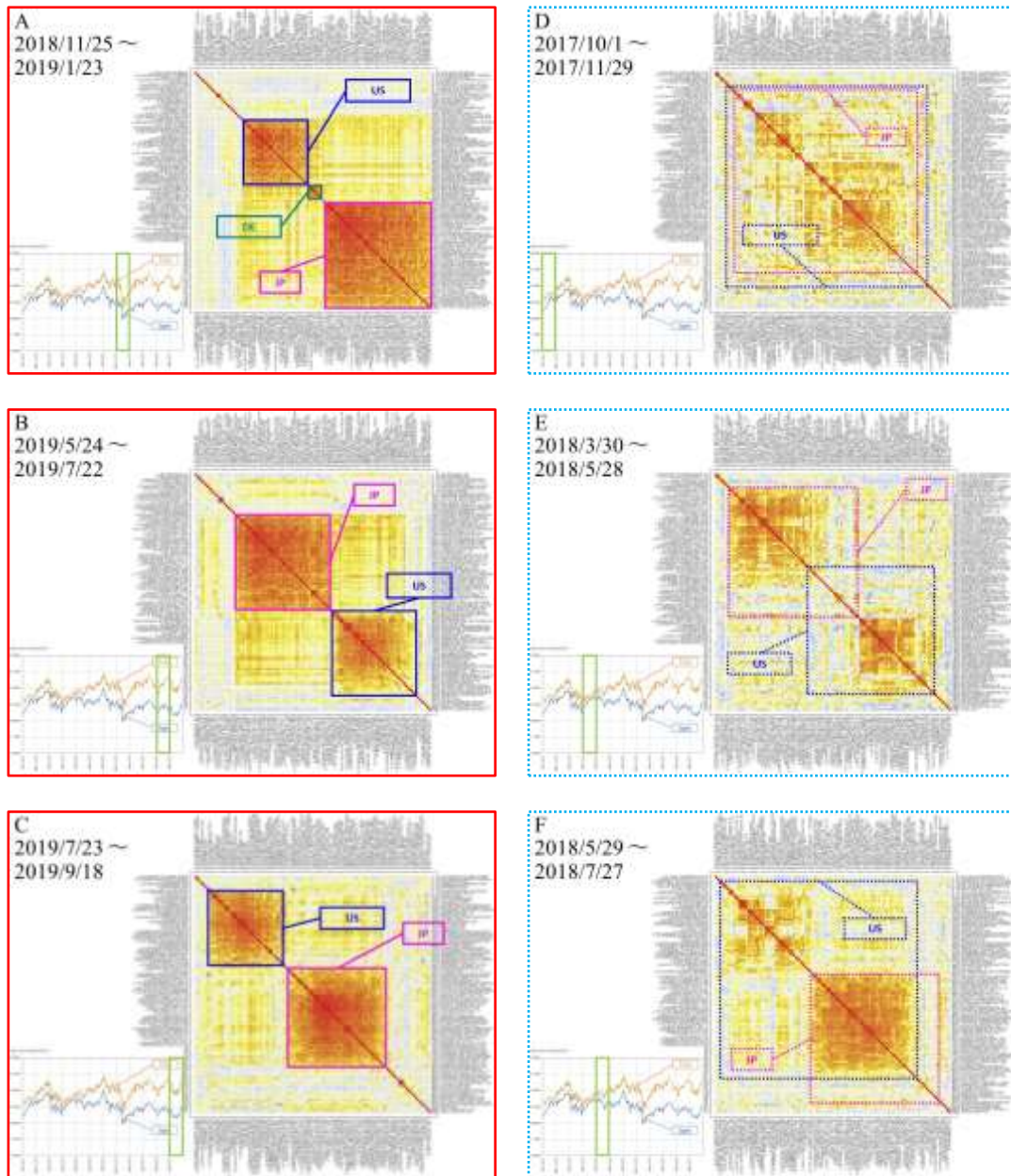


Figure 3: Selected 6 resultant correlation matrix heatmaps of the top 100 global automakers.

Then let's see the result marked "B" in which we can also see the clear Japan cluster. However, we cannot there spot the German cluster so well. We think the reason of this is that the number of German companies is much smaller than others. The "B" period surely indicated a decrease as the effects by the US-Chine trade friction. However, the latter half of the period, the stock prices have recovered. The period "A" was just a decline but the period "C" showed a recovery pattern. Compared to the decrease periods "A", "B", and "C", the periods "D" and "E" showed a growth of stock prices. In these increase periods, the clusters diffused and the boundaries became blurred. The set of Japan companies in the marked "A" Japan cluster diffused as shown in "D" or "E". In the heatmaps, the unclear cluster is framed in a dotted line. Therefore, we formulate the following hypothesis: *When the stock price significantly decreased, there could clearly*

appear a cluster by a country. The evaluation of the hypothesis will be described in the next section.

## 5 Evaluation

In the section, we shall evaluate our hypothesis by our defined objective measurements. In addition, we shall illustrate the clusters' stock price fluctuation and the Markowitz's risk-return plot in order to see the difference among the cluster's stock price patterns. When we conduct clustering, we shall have to see the insight of the resultant clusters.

### 5.1 Stability rate

We shall focus on fluctuations of the Japanese cluster size. The Japan's cluster changes the size time to time. We focus on the side length of the square in the heatmap which means how long the Japan cluster extends. The Japan cluster boundary sometimes expands and sometimes shrinks. When the Japanese cluster density is high with Japanese companies, the Japanese cluster size is small. On the other hands, when the Japanese cluster density is low including many US and DE companies in the cluster, the cluster size increases.

For the measurement, we define and use the following two indices:

- (1) *Japanese cluster's size* : This is measured from the above-mentioned heatmap. The size is measured by the number of companies in the cluster square. As the base of size, we use that of period "A" which consists of 45 Japanese companies. This index shows how extent/shrink the 45 Japanese cluster does.
- (2) *Stock price min value divided by the initial value*: This is an indicator to show whether the decline happened or not in the period, and to show how much severe the decline was. If stock prices always increased, this index becomes 1. When the period includes the decrease, the value becomes less than 1. Smaller the index is, more severe the decrease is. Therefore, the index indicates a stability rate.

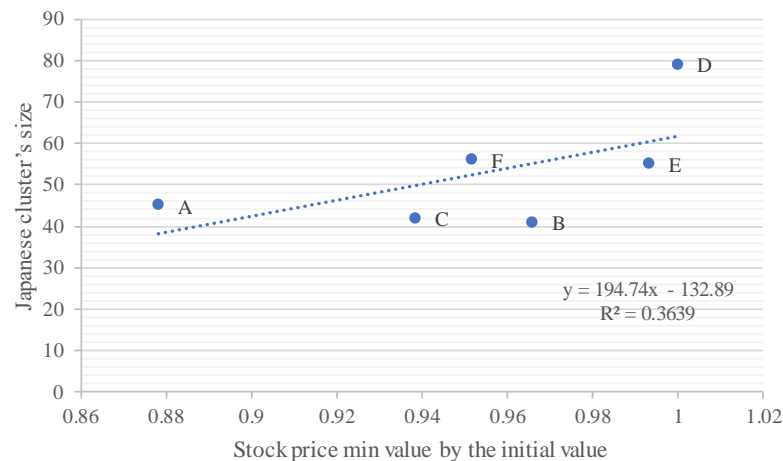


Figure 4: The plot of the min value by the initial value and the cluster size.

In Figure 4, the plot of the above-mentioned the stability level and the Japanese cluster's size. The horizontal axis means the stability level. The linear regression result is also shown there. We found a weak positive relationship between the cluster size and the stability level, because the R

squared value is 0.3639. The reason why the R squared values is small is derived from the definition of the second indicator *Stock price min value divided by the initial value*. Because the definition is so simple to express the complicated pattern of stock price movement, the resultant indicator values were roughly evaluated. However, the tendency that the harder the decline is, the smaller the Japan's cluster size becomes were found. This result supports our hypothesis.



Figure 5: Standardized stock prices of three country-based clusters

## 5.2 Stock price fluctuations by clusters

In the section, we shall illustrate the fluctuation of stock prices in each cluster, because it is important to see the details of the resultant clusters and find the features of the clusters. We would



like to compare the country-based clusters' movement in period A with the worst decline. When we conduct the clustering, the input data to the clustering method is the standardized **return** values, not the standardized **stock price** values, which is a normal method in a financial analysis. However, it is difficult for us to see the return value movement. Then we shall plot the stock values in the following.

We shall compare the three countries stock price movement in period A. Figure 5 is the fluctuation of the standardized stock prices of companies in the US, GE, and JP clusters for the period A (2018/11/25 to 2019/1/23). The vertical axis shows stock price fluctuations, and the horizontal axis shows the date. The stock price data was standardized for the comparison. The red dotted line is the period during which the Christmas crash of 2018 (12/24 to 12/25) occurred.

The whole trend is a decline and recovery type movement of a v-shaped one. The bottom of the movement corresponds to the Christmas Crash. From Figure 5, we can find the following things:

- (1) After the crash, US companies recovered quickly, and their stock prices returned to the previous levels.
- (2) The stock prices of Japanese companies have been affected seriously and crashed to a greater extent than those of American companies. In addition, their stock prices could not return to the previous levels after the crash.
- (3) The stock prices of German companies were hardly affected by the Christmas crash of 2018 in the United States, because the decline at the crash is smaller, compared with the other fluctuation before/after the crash, which means the damage by the crash would be small relatively.

From the above, we can see that the stock prices of US, German, and Japanese companies moved differently during the period A. The damage by the crash was the largest for Japanese companies, compared with the damage on the US companies. When we conduct clustering, it is required to identify the features of individual clusters, plotting the fluctuations. From the resultant clusters, we could surely identify the features of the clusters.

### 5.3 Risk and average-return plot

In a financial analysis, clustering is mainly conducted to make a suitable portfolio. Modern portfolio theory based on Markowitz's mean-variance portfolio model has been widely used for making a portfolio which attempts to minimize portfolio risk for a given level of return or maximizes the return for a given level of risk. Since investors think about a portfolio by using this scatter plot, it is useful to show the results of clustering in this plot when explaining the machine learning clustering results. There the risk is a standard deviation of return values during the given period such as 2 months. The scatter plot of the risk and the average return by the Markowitz's model has been widely used in the portfolio theory. We shall plot the risk and average return values of the three country-based clusters for the comparison. In Figure 6, the returns and risks of US, German, and Japanese companies for period A are plotted. The vertical axis is the average of the return values for this period, and the horizontal axis is the risk values. German companies are represented by blue triangles, Japanese companies by orange squares, and US companies by gray circles.

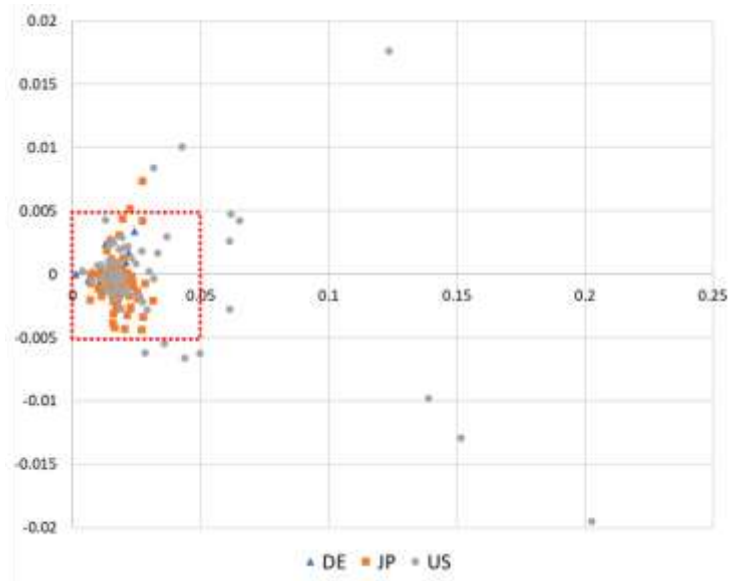


Figure 6-1: Risk and return-average graph of 100 companies in the period A

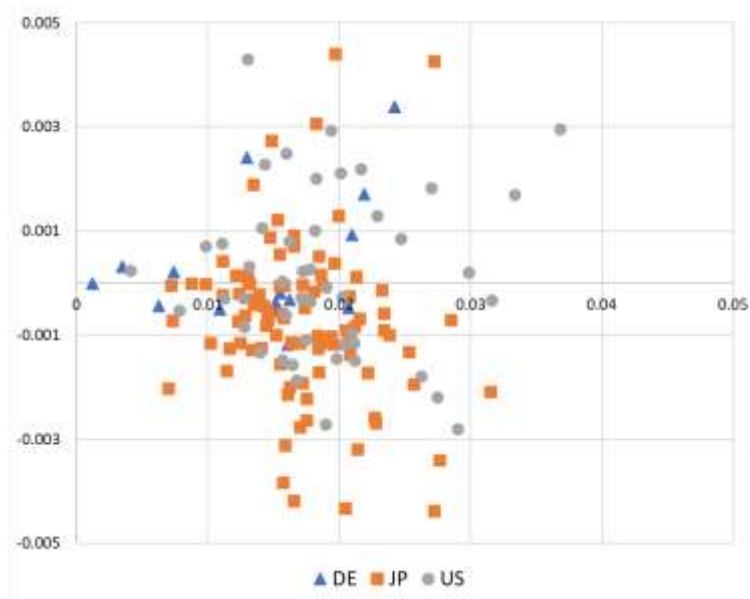


Figure 6-2: Risk and return-average graph of 100 companies in the period A  
(enlarged the red dotted line area.)

Figure 6-2 is an enlarged view of the red line area in Figure 6-1. When we see the plots from the viewpoint of investors, from Figure 6-2, we found the following features:

- (1) US companies stocks' risk range and return range are wider than the others. They include some high-risk and high-return ones and high-risk and low-return ones. Concerning the return values, US shows totally higher return than Japanese ones.
- (2) German companies stocks are the most stable with the small risk range and are relatively gathered at the higher return levels. Because the risk is small and the return is high as a whole, German ones are the most stable.

- (3) Japanese companies stocks are gathered at lower return levels, compared to the others. Many companies mean-returns are negative values. The risk range of the Japan's cluster is also larger than German ones.

These features are the same as ones from the stock price fluctuation analysis result of Section 5.2. From both results of Section 5.2 and 5.3, we could verify that the resultant clusters in period A surely had different fluctuation patterns and different risk-return relationships. To explain the results by machine learning clustering methods to investors, it is required to illustrate the results in their used approach. For data analysts also, it is required to verify how the clustering algorithm identify the differences among clusters.

## 6 Conclusion

We conducted the clustering on stock prices. The target industry field is automobile manufacturing. The number of the global companies is 100. The countries were Japan, US, and Germany. The method of the analysis was machine learning based hierarchical clustering. The data was the two years' stock prices from 2018 to 2019. The two years included the worst decrease of the stock market which happened in December, 2018.

We found that in such kind of a drastic decline which we call a turmoil, there existed clear clusters by countries such as a Japan cluster and a US cluster. We formulate the hypothesis as follows: *When the stock price significantly decreased, there could clearly appear a country-based cluster.* To evaluate our hypothesis, we analyzed the time series changes of the Japan cluster, measuring the cluster size and the stability rate in each period. As a result, there is a positive relationship between the stability level and the cluster size.

In addition, we verified the quality of the clustering results whether the country-based cluster has a different pattern of stock price fluctuation or not. In the largest crash period which period A, we visualized the stock price fluctuations, and the Markowitz's risk and return plots for the three country-based clusters. From the two results, we could verify that the country-based cluster had a different fluctuation pattern. Those results support our hypothesis "at the turmoil, the investment might be determined mainly by the country name, instead of performance and management features of each company".

The future work of this analysis is to verify the hypothesis in various cases.

## Acknowledgement

The research was partly supported by the fund of Gakushuin University Computer Centre project 2021 and by JSPS KAKENHI Grant Number 20H01537.

## References

- [1] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications: A survey," *arXiv preprint arXiv:2005.05786*, 2020.
- [2] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," *arXiv preprint arXiv:2001.01859*, 2020.

- [3] M. L. de Prado, "Building diversified portfolios that outperform out of sample," *The Journal of Portfolio Management*, vol. 42, no. 4, pp. 59-69, 2016.
- [4] T. Raffinot, "Hierarchical clustering-based asset allocation," *The Journal of Portfolio Management*, vol. 44, no. 2, pp. 89-99, 2017.
- [5] N. Hautsch and S. Voigt, "Large-scale portfolio allocation under transaction costs and model uncertainty," *Journal of Econometrics*, vol. 212, no. 1, pp. 221-240, 2019.
- [6] J. Pfitzinger and N. Katzke, "A constrained hierarchical risk parity algorithm with cluster-based capital allocation," Stellenbosch University, Department of Economics, 2019.
- [7] G. Konstantinov, A. Chorus, and J. Rebmann, "A Network and Machine Learning Approach to Factor, Asset, and Blended Allocation," *The Journal of Portfolio Management*, no. Multi-Asset Special Issue pp. 1-18, 2020.
- [8] K. Koutroumbas and S. Theodoridis, "Pattern Recognition (4th edition)," ed: Elsevier, 2009.
- [9] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists* O'Reilly Media, 2016.
- [10] M. L. De Prado, *Advances in financial machine learning*. John Wiley & Sons, 2018.
- [11] H. Markowitz, "Portfolio Selection," *Journal of Finance*, pp. 77-91, 1952.
- [12] H. Markowitz, "Portfolio selection," *Investment under Uncertainty*, 1959.
- [13] R. Clarke, H. De Silva, and S. Thorley, "Portfolio constraints and the fundamental law of active management," *Financial Analysts Journal*, vol. 58, no. 5, pp. 48-66, 2002.
- [14] M. L. de Prado, *Machine Learning for Asset Managers*. Cambridge University Press, 2020.
- [15] M. Kolanovic and R. T. Krishnamachari, *Big Data and AI Strategies*. J.P. Morgan, 2017.
- [16] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering—a decade review," *Information Systems*, vol. 53, pp. 16-38, 2015.
- [17] S. Chu, E. Keogh, D. Hart, and M. Pazzani, "Iterative deepening dynamic time warping for time series," in *Proceedings of the 2002 SIAM International Conference on Data Mining*, 2002: SIAM, pp. 195-212.
- [18] A. Bagnall and G. Janacek, "Clustering time series with clipped data," *Machine Learning*, vol. 58, no. 2-3, pp. 151-178, 2005.
- [19] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *KDD workshop*, 1994, vol. 10, no. 16: Seattle, WA, USA:, pp. 359-370.

- [20] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 1033-1040.
- [21] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern recognition*, vol. 44, no. 9, pp. 2231-2240, 2011.
- [22] Y.-S. Jeong and R. Jayaraman, "Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification," *Knowledge-based systems*, vol. 75, pp. 184-191, 2015.
- [23] H. Y. Sigaki, M. Perc, and H. V. Ribeiro, "Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market," *Scientific reports*, vol. 9, no. 1, pp. 1-9, 2019.
- [24] P. D'Urso, L. De Giovanni, and R. Massari, "Trimmed fuzzy clustering of financial time series based on dynamic time warping," *Annals of Operations Research*, pp. 1-17, 2019.
- [25] S. Majumdar and A. K. Laha, "Clustering and classification of time series using topological data analysis with applications to finance," *Expert Systems with Applications*, vol. 162, p. 113868, 2020.
- [26] P. J. Rousseeuw and L. Kaufman, "Finding groups in data," *Hoboken: Wiley Online Library*, vol. 1, 1990.
- [27] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, vol. 1, no. 14: Oakland, CA, USA, pp. 281-297.
- [28] M. Massahi, M. Mahootchi, and A. A. Khamseh, "Development of an efficient cluster-based portfolio optimization model under realistic market conditions," *Empirical Economics*, pp. 1-20, 2020.
- [29] P. E. Puspita, "A Practical Evaluation of Dynamic Time Warping in Financial Time Series Clustering," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2020: IEEE, pp. 61-68.
- [30] J. R. Shewchuk. "The Singular Value Decomposition: Clustering." <https://people.eecs.berkeley.edu/~jrs/189s17/lec/21.pdf> (accessed 2020/04/01).