An Analysis of Coronavirus Effects on Global Automakers

Akane Murakami and Yukari Shirota

***Abstract*—In the paper, we report changes of stock prices due to the coronavirus spread which happened in 2020, using the machine learning approach based on the hierarchical clustering. The data we used are the top 72 global automobile manufactures' stock prices from 2019 November to 2020 March which were under the coronavirus's first impact. The involved countries are Germany, Japan, US, China, and Korea. One clear result is that the turmoil gave distinctively different effects on the individual country of automakers. We could identify five different clusters of stock price movements, that are country-based clusters. While we traced the time series changes of the clusters, we found the interesting thing. The country-based clusters had the different movement, but when the turmoil started, it became the same movement and the overall correlation coefficient became positive. In addition, we found that at the beginning of the turmoil, most clearly the country-based clusters appeared. This result is expected to give some insights to the issue of international linkages between the movements of the markets' prices by the coronavirus turmoil.**

***Index Terms*—Coronavirus, hierarchical clustering, automakers, stock prices, Hierarchical Rick Parity.**

I. INTRODUCTION

In the paper, we shall describe the stock price analysis by hierarchical clustering. Our interests are to find a specific company cluster on the stock prices at the coronavirus turmoil. The coronavirus spread which happened in December 2019 has affected severely companies globally. Our target is manufacturing industries and we have conducted the global top 72 automakers' stock price analysis which included Toyota and VW. The interesting findings from the result was that the country-based clusters had the different movement, but when the turmoil started, it became the same movement and the overall correlation coefficient became positive. In addition, we found that at the beginning of the turmoil, most clearly the country-based clusters appeared.

The data we used is stock price data of top automakers which consist of 6 Germany (hereafter DE), 26 Japan (JP), 21 US (US) 11 China (CN), and 8 Korea (KR) companies from November 2019 to March 20, 2020. The company stock price data was retrieved from the data base ORBIS by Bureau Van Dijk. The ETF (Exchange Traded Fund) data that we used was retrieved from Nikkei Financial Quest. For the missing data, we conducted a linear interpolation. The programs for the analyses were written in Mathematica by Wolfram.

Manuscript received March 18, 2021; revised December 12, 2021.
The authors are with Gakushuin Univerity, Japan (e-mail: murakami.akane.1031@gmail.com, yukari.shirota@gakushuin.ac.jp).

II. HIERARCHICAL CLUSTERING APPROACH

We shall explain the hierarchical clustering method and the analysis results.

A. Hierarchical Clustering Method

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 enhance portfolio returns while reducing the risk, by making a well-organized company set. To measure the similarity of the stock price movements of two companies, we use a concept of a distance between them, as companies with close distances have similar stock price fluctuations. In the field of portfolio construction, 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:

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

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 $\boldsymbol{G_{i,j}}$ is N times T. From the matrix $\boldsymbol{G_{i,j}}$, we shall make the correlation coefficient matrix $\boldsymbol{\rho} = \{\boldsymbol{\rho}_{\mathbf{i,j}}\}_{\boldsymbol{i,j=1,\dots,N}}$. In HRP, the distance d is defined as follows:

$$\mathbf{d_{i,j}} = \sqrt{\tfrac{1}{2}(\mathbf{1} - \boldsymbol{\rho_{i,j}})} \tag{2}$$

Then the distance matrix $\mathbf{D} = \{\mathbf{d_{i,j}}\}_{\boldsymbol{i,j=1,\dots,N}}$ is obtained. Next, selecting any two distance columns, we shall calculate the Euclid distance as follows:

$$\tilde{\mathbf{d}}_{\mathbf{i,j}} = \sqrt{\textstyle\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}}(\boldsymbol{d_{n,i}} - \boldsymbol{d_{n,j}})^{\mathbf{2}}} \tag{3}$$

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

doi: 10.18178/ijtef.2022.13.3.724

B. Hierarchical Clustering Sample

In the section, we shall present a small clustering result for your quick comprehension (See Fig. 1). That is a clustering result of the top 10 global automakers such as DAIMLER. A company name is added on the top its country name abbreviation “J:”, “U:”, or “D:”. The period is just two months from 2018 March to April. In the resultant dendrogram in Fig. 1, the horizontal axis shows the cluster 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 [11]. In Fig. 1, we can find three clusters corresponding to an individual country; from the bottom they are Japan, Germany, and US. Being isolated from the three clusters, Audi (Germany) is located. In Fig. 1, we found that the Germany cluster was near to the US cluster than to the Japan cluster.

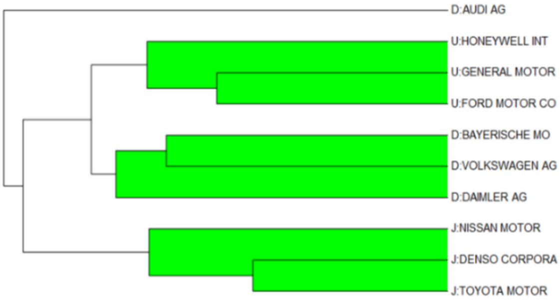

Fig. 1. A hierarchical clustering of the top 10 automakers.

III. Hierarchical Clustering Results

In the section, we shall show the hierarchical clustering results.

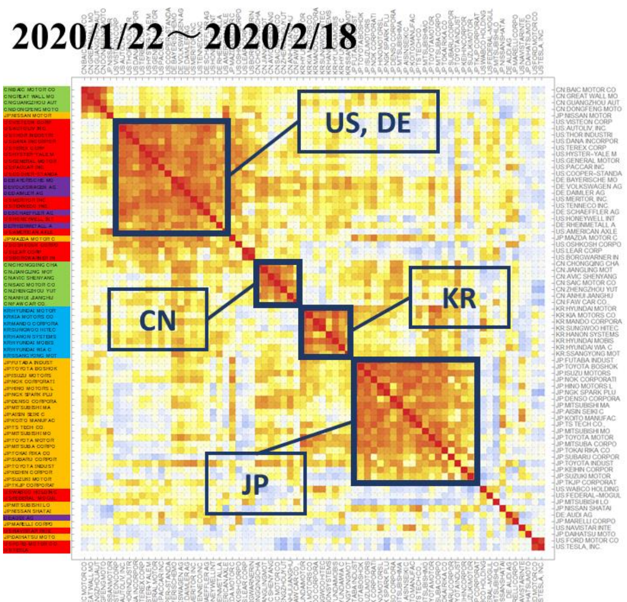

Fig. 2. Coefficient matrix during 1/22 to 2/18.

We present the coefficient matrix after the quasi-diagonalization (See Fig. 2). The number of companies is 72. They include USA (hereafter US), Japan (JP), Germany (DE), Chine (CN), and Korea (KR) companies. The return value is in advance standardized and input. After the hierarchical clustering, we obtain the coefficient matrix (See Fig. 2). The company order was sorted already by the quasi-diagonalization. The correlation coefficient in the matrix is presented in red if the value is 1; The 1s are listed diagonally in the3 matrix. The yellow means that the value is positive values (0 to 1), the white means zero, and the blue means negative values (-1 to 0).

The country-based four representative clusters can be clearly seen which are corresponding to the top US with DE, CN, KR, and JP. The DE cluster is involved in the US cluster. The left-side labels show the company’s country. The colour shows the country; US in red, DE in purple, CN in green, KR in blue, and JP in yellow. In general, we can hardly identify country-based clusters. It may be rare and interesting that the clusters appear clearly as shown in Fig. 2.

Then we shall investigate the time series changes of the coefficient matrixes by every four weeks as follows:

(1) 2019/11/27 to 2019/12/24,
(2) 2019/12/25 to 2020/1/21,
(3) 2020/1/22 to 2020/2/18, and
(4) 2020/2/19 to 2020/3/17.

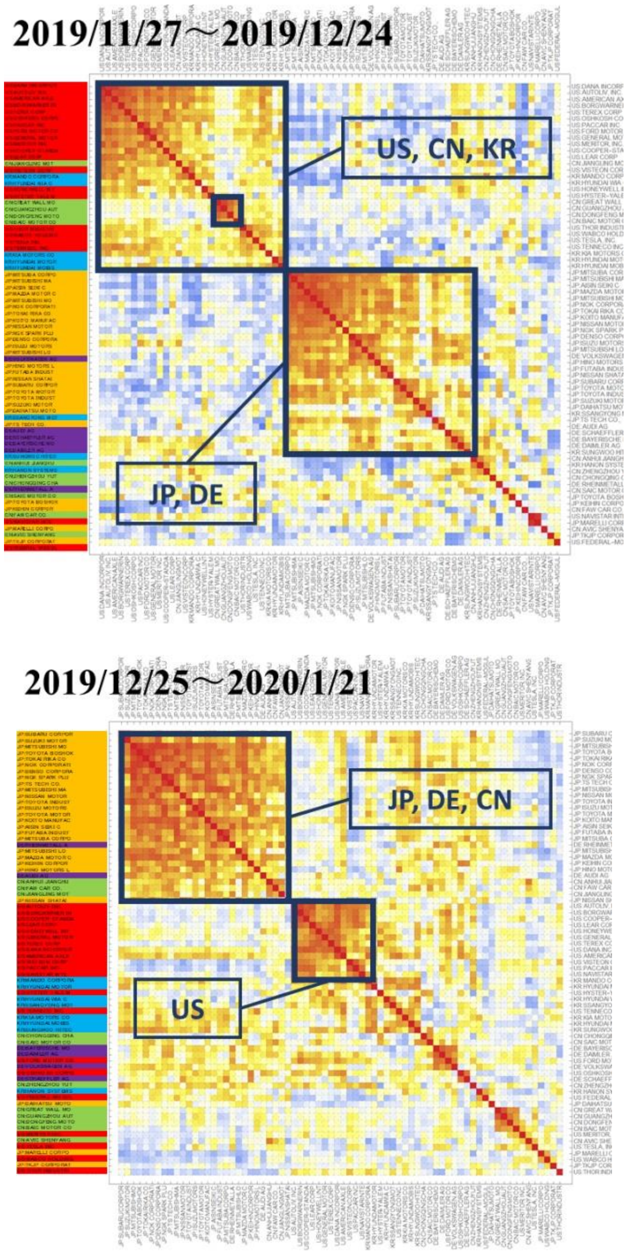

Fig. 3. Covariance matrixes #1 (from 11/27) and #2 (from 12/25).

We call them the four periods called #1 to #4. The resultant four coefficient matrixes are illustrated in Fig. 3 and Fig. 4. In general, they think that the new coronavirus occurred in Wuhan City in December 2019 and spread from Wuhan to the world. Then the period #1 coefficient matrix does not yet show the virus effect (See Fig. 3).

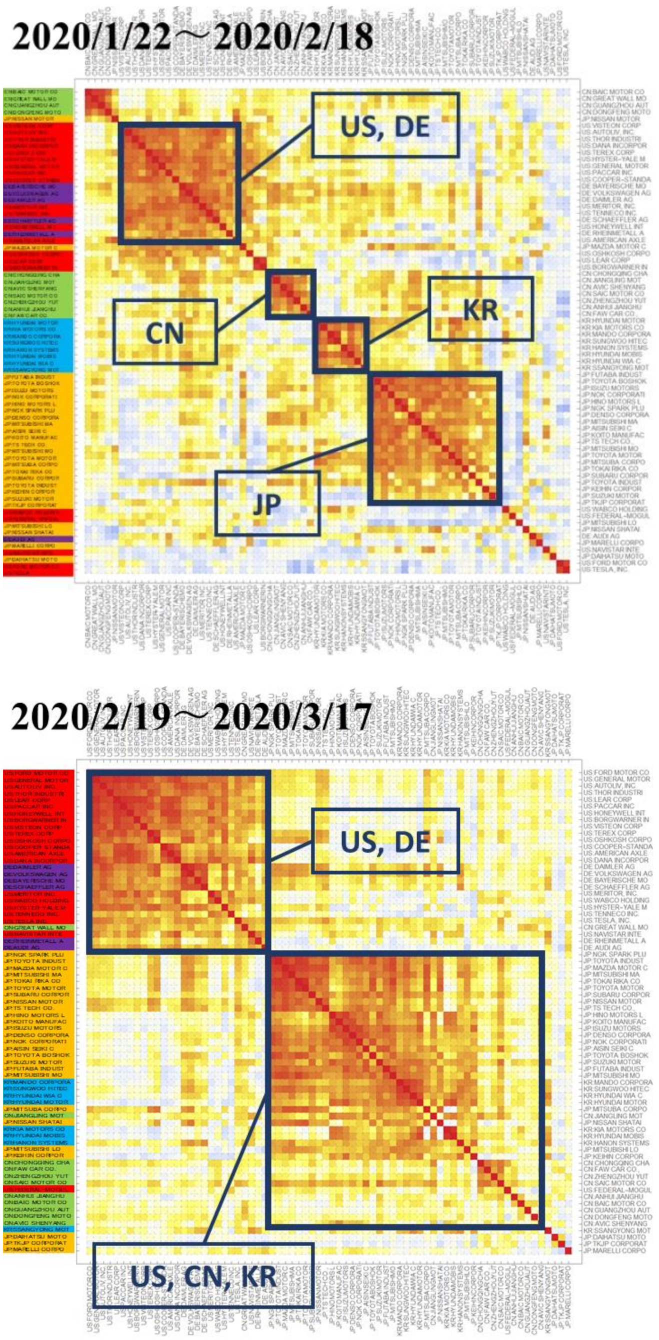

Fig. 4. Covariance matrixes #3 (from 1/22) and #4 (from 2/19).

In the matrix #1, there are two large clusters which are a US cluster with CN and KR companies, and a JP cluster with DE companies. The two large clusters' movements are not similar because the margin of the clusters is in blue, which means the correlation coefficient is negative. We can say the two clusters show the opposing movements.

Then we see the matrix #2, gradually the blue coloured negative coefficients between the two clusters decrease. In the matrix #2, there also exist two country-based clusters which are a JP cluster and a US cluster. There, we can see the increase of similarity between the two clusters, because the number of blue cells decreases.

In addition, the correlation within the cluster increased. The density of the Japan cluster and the US cluster increased and the square size of the clusters became smaller.

Then, in the matrix #3, we can see a clear set of country-based clusters. The identified clusters are, a US cluster, a CN cluster, a KR cluster, and a JP cluster. Such an identification of the country-base clusters is interesting. We think that the coronavirus spread is the reason of the country-based cluster appearance.

In the #3 matrix, the correlation coefficients among clusters present warm coloured, compared with the matrix #2. The warm colour such as yellow or orange means positive values of correlation coefficients. In other words, we can say that the similarity of the movement was spreading globally. In the matrix #3, when we see the relationship between the US cluster and the JP cluster, the coefficient is partially negative. When we see the relationship between the CN cluster and the KR cluster, the coefficient is partially negative. As a result, in the matrix #3, the whole trend came to show more similarity but partially negative relationships are left.

Finally, let's see the period #4. There the similarity overall increased, which must be derived from the whole declining trend. In the period #4, as a whole, the automakers' stock prices globally decreased.

IV. Evaluation

In the section, we shall evaluate the above-mentioned coefficient matrix change, compare to the stock price indices.

To evaluate the effects of the coronavirus, let's see the individual country average stock price movements (See Fig. 5). We used Exchange Traded Fund (ETF) prices sold in Tokyo Stock Market. To see the US market movement, we used the "Simple-X NY Dow Jones Indices ETF" which was an ETF of Dow Jones Industrial Average, a leading stock index of 30 leading companies in the US stock market. Dow Jones is used as a measure of the overall US market. In addition, ETF US (S&P) is also presented in Fig. 5 to see the US stock price movement. The similarity between "Simple-X NY Dow Jones Indices ETF" and "ETF US (S&P) is high, because both move, based on the US market movement.

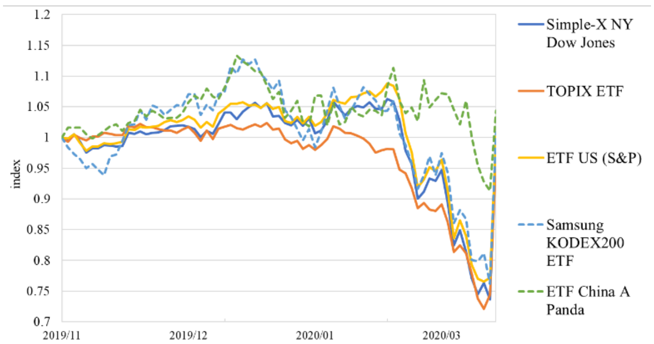

Fig. 5. World stock price movement.

To see the Japanese stock price movement, we use the "TOPIX Linked ETF". To see the Chinese stock price movement, we use the "China A panda" ETF. The "China A panda" ETF is of CSI300 which consists of the top 200 stocks listed "China A Stock" market across the Shanghai and Shenzhen Stock Exchanges. To see the Korea stock price movement, we use the "Samsung KODEX200" ETF. The

"Samsung KODEX200" ETF is linking the value of KOSPI200 (Korea General Stock Index 200) which is the Korea's leading stock index.

In the graph of the five ETFs in Fig. 6 are index values, being set the first value on 2019/11/27 as the base value1. As shown here, the damage by the coronavirus was so hard and from 2/20 the prices drastically decreased.

We would like to represent the turmoil level of a stock market. For the measurement of the stock price stability, we define and use the indices "stability rate" by (the stock price minimum value divided by the initial value in the period). 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 in the period increased, this index becomes 1. When the period includes the decline phase, the "stability rate" value becomes less than 1. Smaller the index "stability rate" value is, more severe the decrease is.

In Fig. 6, the stability rate values are illustrated. The periods are from #1 to #4 (See the horizontal axis). As shown in Fig. 6, in the period #1 and #2, the movements were stable but in the period #3 the decline happened, and then we identified the KR decline was larger than others; the KR stability rate is about 0.92. And then much larger decline happened in the period #4 on all of the global markets. That is the effect by the coronavirus.

In the period #4, for example, the Dow Jones ETF showed the decrease to be about 70 %, compared to the initial value of the period (2020/2/19). Compared to the other countries' companies, the China A Panda's decline was not large; the stability rate in the period #4 was 0.85.

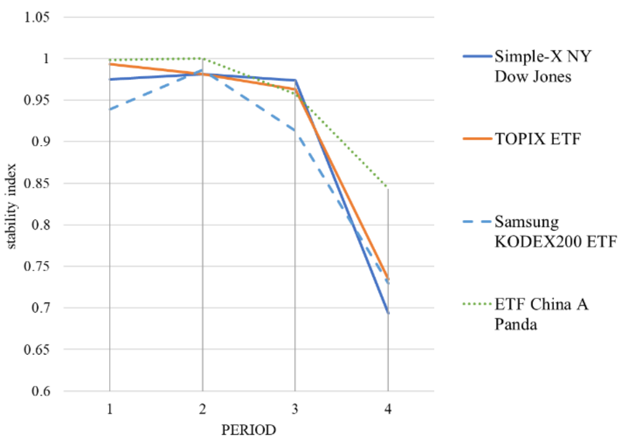

Fig. 6. Stock market stability rates of the four periods.

Compared with the stock market stability rate change, how are the coefficient matrix changes? The following three things we could say:

1) Contrary to our expectations, we could find country-based clusters here and there. That was an unexpected thing. Even if it is in the period #1, we could see the US and JP clusters more clearly.
2) It was the beginning period of the drastic decline by the coronavirus that the country-based clusters most clearly appeared. That corresponds to the period #3.
3) In the turmoil periods from #2 to #4, the similarity of the movement among the country-based clusters gradually increased; the correlation coefficients became from negative to positive and finally came to show the same decline movement as a worldwide turmoil.

While there are not disasters such as earthquakes or political conflicts, there is no country-based clusters in the global automakers' stock prices. However, in 2019 December, the coronavirus had been spread and the coronavirus damage made the global automakers' clusters in the stock price clustering. The cluster was made as a country-based cluster such as a Japan automaker cluster and a US automaker cluster.

At this stage, we cannot mention whether the things we found can apply to a general disaster or not. Maybe the target is limited to be automakers and the reason is limited to be the coronavirus. We cannot say whether the above three things are applicable in other industry fields or so. Further analyses are required to say that.

V. Conclusions

In the paper, we presented the stock price analysis by the hierarchical clustering. In February 2020, the global top 72 automakers' stock prices were declined owing to the coronavirus. The companies' countries we analyzed are Japan, US, China, Korea, and Germany.

We found the country-based clusters in the results more frequently than we expected. To see the effect of the coronavirus, we analyzed the stock data changes, splitting the data every four weeks for 2019/11/27 to 2020/3/17. The number of periods is four. We conducted a hierarchical clustering sequentially these four sets of four-weeks data in order to investigate the clusters' time series changes. Using the resultant correlation coefficient matrixes.

From the hierarchical clustering analysis, we found (1) Contrary to our expectations, we can find country-based clusters here and there, (2) It was the beginning period of the drastic decline by the coronavirus that the country-based clusters most clearly appeared, and (3) In the turmoil periods from #2 to #4, the similarity of the movement among the country-based clusters gradually increased.

Although we had understood that the stocks of the global automakers have fallen by the influence of the coronavirus, it was meaningful to visualize the appearance of the clusters in detail. It is also interesting that the cluster of each country appeared most clearly at the time when it started falling just before the drastic decline, though the companies showed the same decline movement trend after the stock price collapsed globally. We shall continue to analyze the global automakers' stock prices using the hierarchical clustering.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Murakami and Shirota conducted the research. Murakami analyzed the data. Shirota mainly evaluated the data. Murakami and Shirota wrote the paper. All authors had approved the final version.

ACKNOWLEDGMENT

The research was partly supported by the fund of Gakushuin University Computer Centre project 2020 and by JSPS KAKENHI Grant Number 20H01537 as Kaken B.

REFERENCES

[1] D. Prado and M. Lopez, *Advances in Financial Machine Learning*, John Wiley & Sons, 2018.

[2] D. Prado and M. Lopez, "Building diversified portfolios that outperform out of sample," *The Journal of Portfolio Management*, vol. 42, no. 4, pp. 59-69, 2016.

[3] K. Koutroumbas and S. Theodoridis, *Pattern Recognition*, 4th ed., Elsevier, 2009.

[4] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, O'Reilly Media, 2016.

[5] J. Friedman, R. Tibshirani, and T. Hastie, "Data mining, inference, and prediction," *The Elements of Statistical Learning*, Springer, 2013.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[7] H. Aoyama, Y. Fujiwara, Y. Ikeda, H. Iyetomi, W. Souma, and H. J. C. B. Yoshikawa, *Macro-econophysics*, Cambridge University Press, 2017.

[8] J. Miśkiewicz, A. Tadla, and Z. Trela, "Does the monetary policy influenced cross-correlations on the main world stocks markets? Power Law Classification Scheme analysis," *J Physica A: Statistical Mechanics Its Application*, vol. 519, pp. 72-81, 2019.

[9] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Physical Review E*, vol. 65, no. 6, p. 066126, 2002.

[10] Y. Shirota and B. Chakraborty, "Visual explanation of eigenvalues and math process in latent semantic analysis," *Information Engineering Express, Information Engineering Express*, vol. 2, no. 1, pp. 87-96, 2016.

[11] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "A random matrix theory approach to financial cross-correlations," *Physica A: Statistical Mechanics and Its Applications*, vol. 287, no. 3–4, pp. 374-382, 2000.

[12] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices (Cambridge Studies in Advanced Mathematics)*, Cambridge University Press, 2009.

[13] J.-P. Bouchaud and M. Potters, *Financial Applications of Random Matrix Theory*, Oxford University Press, 2011.

Akane Murakami graduated with a degree in philosophy in 2015 at Gakushuin University in Tokyo, Japan. Now she is a graduate student in the Graduate School of Business Administration at Gakushuin University. She is interested in economics data analysis. Her research theme is stock price/ETF price analyses by using machine learning based methods. Especially she uses hierarchical clustering methods. She writes the programs in Mathematica and python. She is interested in creation of visual teaching materials in which both her hand-writing illustrations and the Mathematica written visual graphics are integrated.

Yukari Shirota received Dr.Sc. from the University of Tokyo (Faculty of Science, Department of Information Science), Tokyo, Japan in 1989.

From 2002 to present, she is a professor at the Faculty of Economics, Gakushuin University, Tokyo. Currently, and She is the director of Gakushuin University for Research Institute for Economics and Management. From June 2020, she is a fellow of Information Processing Society of Japan.

She currently teaches the following courses: "Business Mathematics", "Data Analysis for Management", and "Seminar class (Machine-Learning for Management)" for undergraduate students, and "AI in Finance" for graduate students.

Her research interests are AI approaches for companies' performance analysis, and Indonesia gender empowerment analysis (See https://www-cc.gakushuin.ac.jp/~20010570/Indonesia/).

She is interested in development of visual teaching materials in Mathematica or so. The target subjects are math in AI and statistics.