Classification of Japanese Electrical Equipment Manufacturing Industry Recovery Patterns after Disasters: Case Study of August 2019

Kenji Yamaguchi and Yukari Shirota

Abstract—In the paper, we analyze the recovery pattern of Japanese electrical equipment manufacturing companies after the President Trump remark in August 2019. The President's remark made the companies' stock prices decreased severely. The research consists of two parts. In the first part, we conducted Random Matrix Theory to extract representative decline/recovery patterns. Then we tagged A/B/C/D to the companies' recovery types. The class A means a strong recover power. Then as the second part, we conducted machine learning tree-based classification using the tags A/B/C. The predictors are eight variables like ROA, ROE, and VAR. The resultant Decision Tree model provided us with the two different approaches to the class A group. The recovery and repulsion power will be higher in the company with high ROA and in the company that manufactured the product with high VAR. In addition, another class A company group is made and the feature is the high inventory turnover ratio.

Index Terms—Disaster impact on stock prices, Us-China trade friction, random matrix theory, singular value decomposition, disaster impact network.

I. INTRODUCTION

We have been investigated the Japanese industry recovery patterns after disasters, by stock prices. In the paper, we will analyze the President remark's effect on the Japanese electrical equipment manufacturing companies.

President Trump announced on first August, 2019, that the fourth trading sanctions against China would be imposed on almost all Chinese products. In the Tokyo stock market on the second August, 2019, the decline in economic sensitive stocks such as electrical equipment and machinery was noticeable. On 14th August, due to the trade conflict between the United States and China, 1152 stocks, which are more than 50% of the first section of the Tokyo Stock Exchange, have fallen by 1 in PBR (Price Book-value Ratio). They include 84 companies of electrical equipment, 100 companies of chemistry, and 101 companies of machinery. The companies were a company group depending on external demand that is easily affected by the global economy[1].

Our research approach consists of two parts: (1) find the recovery patterns and title them class names like A/B/C/D, and (2) find the dominant management index for the classification. Especially we would like to know the dominant index for a quick recovery pattern which is the class A. The method for the part (1) is **Random Matrix Theory** (abbreviated by **RMT**). The method for the part (2) is machine learning classification methods which are tree-based methods **Random Forest (RF), XGBoosting (XGB)**, and **Decision Tree (DT)**. By the analysis, we finally found that the index ROA and value-added ratio (VAR).are dominant factors for a strong recovery power. In addition, the inventory turnover is also effective.

In the next section, we shall explain the data we used and the recovery pattern finding method. Then in Section III we describe the extracted patterns class A/B/C/D. In Section IV, we shall describe the classification results by the tree-based methods and our evaluation on this will be described. Finally, we shall conclude the paper.

II. DATA AND PATTERN FINDING METHOD

In the section, we shall explain the data and the **RMT** method we used. The data is stock price data of **167 electrical equipment manufacturing companies** in the first section of the Tokyo Stock Exchange. The data period is from **2019/07/25 to 2019/09/24** of which sales days are 42 days.

First we need to select the recovery patterns after the disaster. The method for finding the patterns we selected was RMT. Historically, many physical phenomena have been successfully solved using RMT, and remarkably a great number of applications to finance have arisen during the last 20 years. In 2000, Plerou used RMT to find cross correlations in financial data [2], [3]. Since then in the field of econophysics, RMT has been widely used to make portfolios and to monitor company network [4]-[16].

The RMT math process is **SVD (Singular Value Decomposition)** [17], [18]. In RMT, we conduct the SVD on the standardized return values of stock price data [2], [19], [20].

The return value is a natural logarithm of the ratio between today's price and the previous day's one and defined as follows: $G_{i,j} = ln(S_{i,j}/S_{i,j-1})$ where $S_{i,j}$ is the ith company's stock price on j-th day and $G_{i,j}$ is the return value on j-th day. As that is a natural logarithm of the ratio, if the today's price is greater than the precious one, then the ratio will be greater than one and the logarithm value will be 0 or more. And if the today's price is less than the precious one, then the ratio will be less than one and the logarithm value will be negative. Because different companies stock values

have varying ranges of means and volatilities, the return values should in advance be standardized concerning each company data, so the mean value becomes 0 and the standard deviation becomes 1.

The method is conducted as follows: First, we conduct SVD on the time series data matrix X of stock price return values, so that we can obtain, as the output, the matrices U, W, and V^T. Then, we have gotten two kinds of eigenvectors obtained by (U W) and by (S V^T). We call the eigenvectors (1) **Brand-Eigenvector, obtained by (UW)**, and (2) **Dailymotion-Eigenvector, obtained by (SV^T)**. An eigenvector corresponds to a principal component. The principal component means a company group with a similar movement. In the paper, we call the principal component as a **GRP** (group). For example, we can get a bank group, a telephone group, an automaker group, and so forth. A principal component can be expressed in two ways; they are (1)Brand-Eigenvector and (2) Dailymotion-eigenvector. Each Brand-eigenvector consists of element values of all 167 companies (See Fig. 1). In the paper, we represent a Brand-Eigenvector group by "GRP# [+/-] [number]" such as **GRP#+1** and **GRP#-1**, dividing them into the positive and negative parts. On the other hand, each Dailymotion-Eigenvector expresses a stock movement corresponding to the GRP.
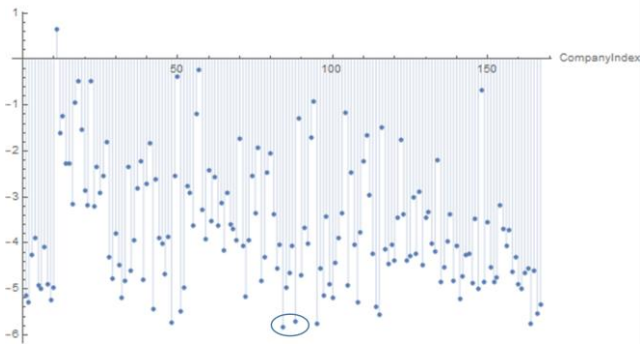

Fig. 1. The Brand-Eigenvector of GRP#-1 and GRP#+1.

We represent **eigenvalues** of the Brand-Eigenvectors and Dailymotion-Eigenvectors as $\{\sqrt{\lambda_i}\}$. The set of eigenvalues are the same in both eigenvectors. On the other hand, we make the matrix $C = \frac{1}{T}XX^T$ which is called the **correlation matrix**. The eigenvectors of the correlation matrix C are $\{\lambda_i\}$ (See Fig. 2). In RMT, $\{\lambda_i\}$ is used to testify the normality of the elements.
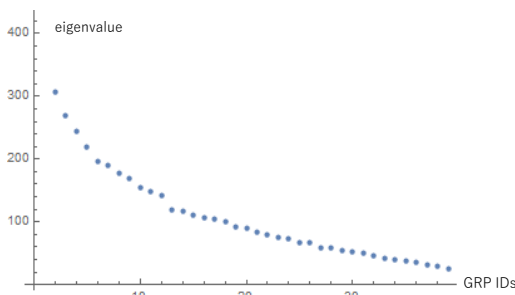

Fig. 2. The eigenvalues of the correlation matrix C.

III. RECOVERY PATTERNS

In the section, we shall describe the recovery patterns selected. Fig. 2 shows the 39 eigenvalues of 39 GRPs. That is $\{\lambda_i\}$, not $\{\sqrt{\lambda_i}\}$. This means that the number of GRPs we obtained is 39. There the first eigenvalue λ_1 was 2452.2; the value is so large and is out of the plot range in Fig. 2. The eigenvalues below #2 are not so different as shown in Fig. 2. From this, the first GRP has the gigantic impact, compared to others. Let us see the Brand-Eigenvector of GRP#1 (See Fig. 1). The x axis shows the company ID. Almost all the companies have affected the large damage. The damage patterns of GRP#-1 are shown in Fig. 3.
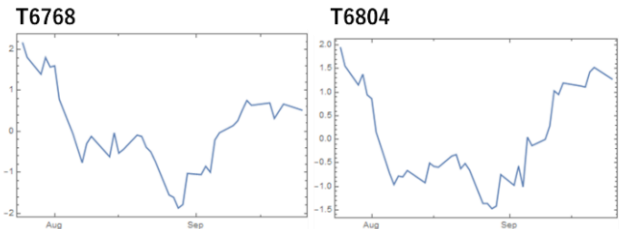

Fig. 3. The standardized stock price of representative companies of GRP#-1. The element values of them are -5.7 both.
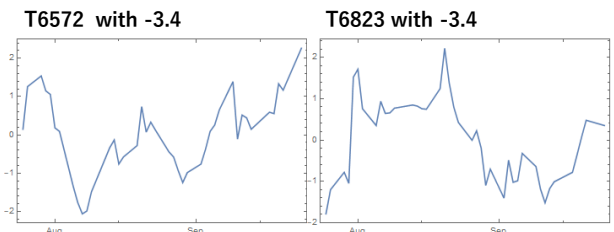

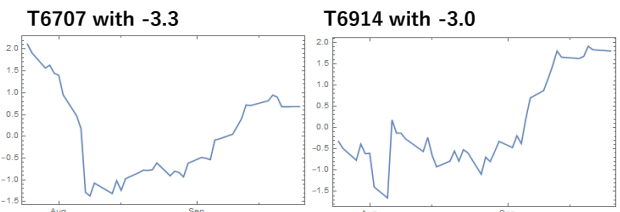Fig. 4. The standardized stock price of representative companies of GRP#-2.

Fig. 5. The standardized stock price of representative companies of GRP#-2.
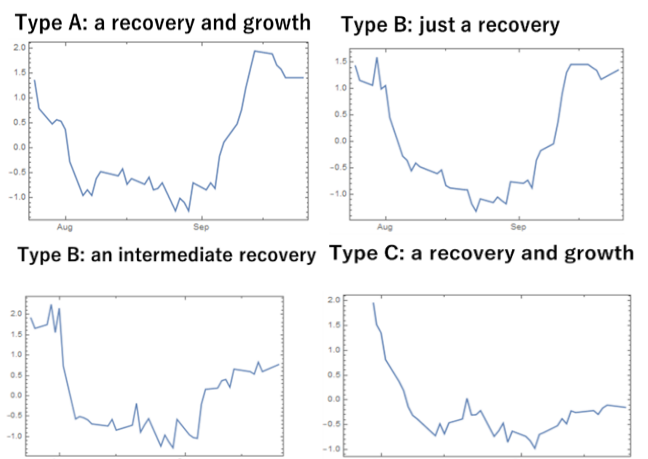

Fig. 6. Examples of a type A/B/C.

The two companies in Fig. 3 are ones marked in a circle in Fig. 2. The movement pattern is a decline and recovery. Before we conducted the RMT analysis, we expected that GRP#-2 and GRP#-3 had their specific movement patterns. We wanted to use GRP#-2 and GRP#-3 patterns as well as GRP#-1 one in order to express each company's movement.

However, as shown in Fig. 4 and Fig. 5, we cannot interpret the specific patterns from them. The reason why we cannot find the pattern from them is that the first GRP impact is gigantic compared to others; the others movements are weak and vulnerable, compared to the first impact. Since we cannot extract/interpret patterns of GRP#2, GRP#3, and GRP#4, we tagged A/B/C/D for each company, manually. The label A/B/C/D means the followings:

A) A recover with strong repulsion and growth
B) A mediocre recovery or just a recovery to the previous value
C) A decline without a repulsion
D) Others (complicated patterns)

In Fig. 6, each type example is shown. Next, we shall conduct the machine learning classification using the A/B/C classes tags.

IV. ANALYSIS GOAL

In the second part analysis, we will classify the electric companies by using the recovery patterns A/B/C that we obtained in the previous section. We finally obtained the data of 122 Japanese electrical equipment manufacturing companies from them. The number of each class is A:51, B:54, and C:17. These industries typically represent the Japanese manufacturing industry.

As determinants of the classification, we restrict to accounting measures related to committed physical resources because we pay attention to managerial properties related to resource commitment rather than environmental determinant factors such as business and competitive ones. In other words, our analysis focuses on whether any common managerial desirable properties exist for the recovery after a disaster. Then predictors are prepared as follows:

0. operating income to net sales ratio [%]
1. inventory turnover ratio [times]
2. Book-value Per Share (BPS)[JPY]
3. total operating profit ratio on used capital (ROA) [%]
4. return on equity ratio (ROE) [%]
5. turnover of tangible fixed assets [times]
6. value-added ratio [%]
7. sales growth rate [%].

The data we used in this study are the annual data from the financial information database of Japanese domestic companies titled EOL by PRONEXUS Inc. (https://www.pronexus.co.jp/english/). We use the fourth quarter report values in 2018. The data are in advance all standardized. Therefore all the data are dimensionless.

The machine learning classification proceeded as follows: (1) data standardization, (2) split data to training data and test data, (3) Cross-Validation, (4) Grid-search. For generalization, it is needed to split training data and test data. We set the test_data ratio to be 20% in the analysis. The machine learning library we used is Scikit-Learn in Python[21]. A great advantage to use Scikit-Learn is that we can easily conduct (3) Cross-Validation, and (4) Grid-search. In the Cross-Validation, the training data is randomly split into 3 to 5 distinct subsets[17]. Leaving one subset for evaluation, training is conducted on the other (*n*-1) subsets and finally evaluate the result model by the rest subset.

Repeat the training and evaluation in n times. Then the result is an array containing the n evaluation scores. The classification models have hyper-parameters to be set. To find the best hyper-parameter combination, we have to repeatedly conduct the training. To do that, Scikit-Learn's GridSerachCV is convenient which will evaluate all the possible combinations of hyper-parameter values, using Cross-Validation. Finally the GridSerachCV offers the best parameter set. With the best parameter set, we evaluate the model using the left test data. Usually we get and compare a score by the train data and a score by the test data on the best parameter set.

There are in general two purposes of a classification analysis which are (1) making a good predictive model, and (2) identifying some of the most important predictors. These are closely related but different tasks. Our purpose here is identifying some of the most important predictors.

V. RANDOM FOREST AND XGBOOSTING CLASSIFICATION

We used the three methods of the machine learning classification which are a Random Forest (RF), a XGBoosting (XGB), and a decision tree(DT). The RF is an ensemble of decision trees [21]. XGB stands for extreme gradient boosting and belongs to tree ensemble methods [22]. The XGB algorithm is described in [17].

The decision tree-based model returns only relative importance values of predictors. Scikit-Learn measures a predictor's importance by looking at how much the tree nodes that use the predictor reduce impurity on average across all the trees in the forest [21]. The relative importance value is 0 or more. In the followings, we shall describe results of the three methods.
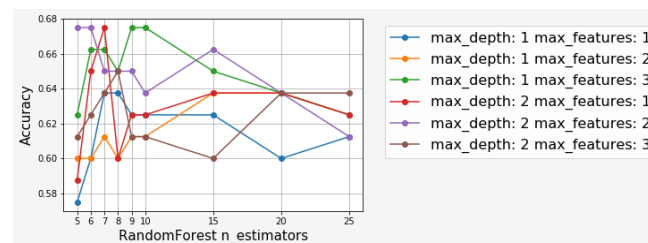

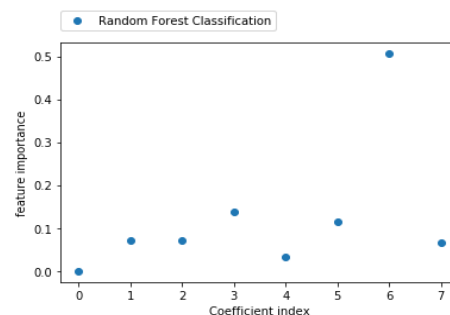Fig. 7. Three parameter value comparison in the random forest.

Fig. 8. Relative importance of 8 predictors in RF.

In Fig. 7, the result by RF is shown. The best accuracy is about 0.675. In Fig. 8, the importance values of 8 predictors by RF are shown. The importance value is a relative one and the total of them becomes one. The most important predictor is #6 value-added ratio (VAR) there. The result by XGB is shown in Fig. 9. The best accuracy is about 0.65. The most

important predictors by XGB are #3 ROA and #6 VAR. Compared to the result by RF, the variance of relative importance is large. The result by DT is shown in Fig. 11. The best accuracy is about 0.625. The most important predictor is #3 ROA.
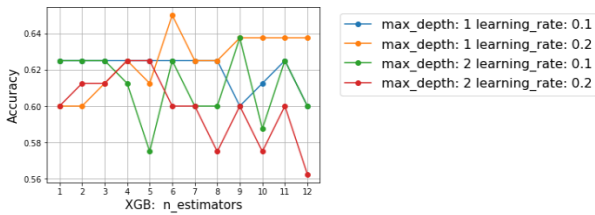

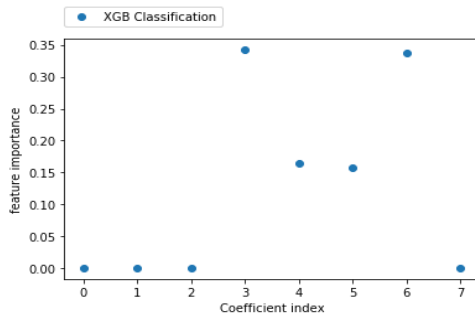Fig. 9. Three parameter value comparison in the XGB.

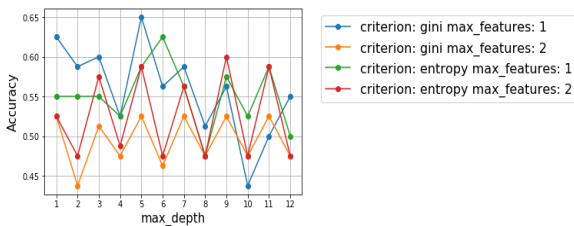Fig. 10. Relative importance of 8 predictors in the XGBoost.

Fig. 11. Three parameter value comparison in the DT.

Seeing the results from tree-based classifications, we shall evaluate which predictor is an important index for the recovery type A. As candidates, two predictors were selected: ROA, and VAR. The index ROA (Return on Asset) is an indicator of how profitable a company is relative to its total assets. ROA gives us an idea as to how efficient a company's management is at using its assets to generate earnings. The index VAR is a percentage of value added to sales. If the VAR is high, it can be said that the ratio of value newly created by a company is large.
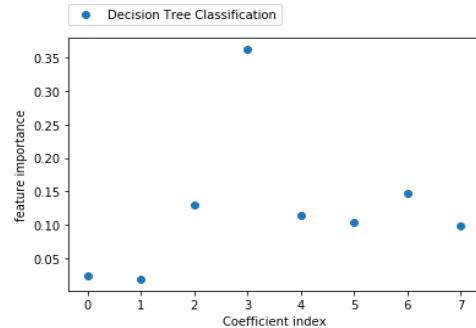

Fig. 12. Relative importance of 8 predictors in DT.

Then which result should be used, one by RF or one by XGB or one by DT? The accuracy level is around 0.65 which is not reliable.

Then let's see the generated DT by the DT classification in Fig. 13. We can see the DT visually, although we cannot see this kind of visual material by RF or by XGB. In Fig. 13, we can find the important two division nodes which use ROA and inventory turnover. The two nodes are marked by bold fonts in Fig. 13. The right group divided by the ROA node makes a class A group which is marked by a triangle. The value there shows the number of each classes like [A, B, C]. The value of the ROA node was [26,39,2]. The [26,39,2] was divided to the right side value[14,2,1] and the left side value [12,37,1]. The ratio of class A in the right side becomes 14/(14+2+1)=82%. On the other hand, the ratio of class A in the left side becomes 12/(12+37+1)=24%. The ROA node reduced the impurity level concerning class A. We can say that ROA is dominant for class A selection.
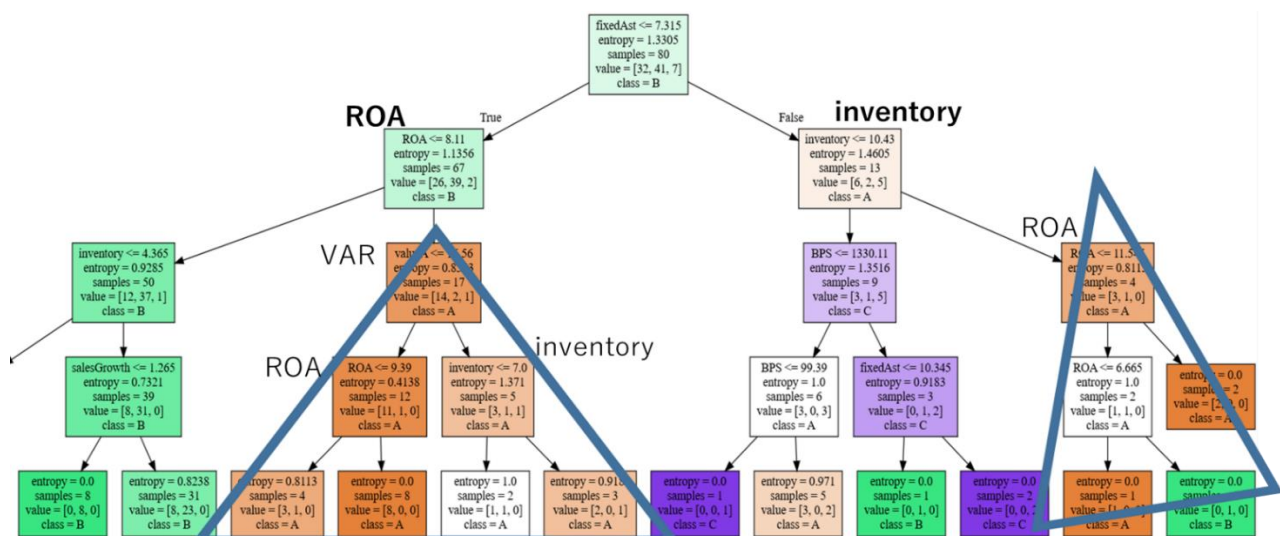

Fig. 13. The DT from the DT classification.

Then let's see the inventory node of which value is [6,2,5]. After the division, the value in the right side becomes [3,1,0] and the ratio of class A becomes 3/(3+1)=75%. The value in the left side becomes [3,1,5] and the ratio of class A becomes 3/(3+1+5)=33%. The inventory turnover node reduced the impurity level concerning class A, too. We can say that the inventory turnover is also dominant for class A selection, although the importance level was low in Fig. 8, Fig. 10, and

Fig. 12. The XGB and RF results showed that the importance level of VAR was high in Fig. 8 and Fig. 10. However, seeing the DT in Fig. 13, the VAR division is just after the ROA division and there is not so important to classify the class A companies.

In conclusion, we think that strong recovery power is featured as the high ROA, high VAR and perhaps high inventory turnover ratio. The recovery and repulsion power will be higher in the company with high ROA and in the company that manufactured the product with high VAR. In addition, another class A company group is made and the feature is the high inventory turnover ratio.

VI. CONCLUSION

In the paper, we analyze the recovery pattern of Japanese electrical equipment manufacturing companies after the President Trump remark in August 2019. The President's remark made the companies' stock prices decreased severely. In a sense, his remark is a kind of disasters like an earthquake and a flooding. We have been analyzing the stock price movement just after the disaster. In the paper we analyzed the potential recovery power of companies just after the disaster.

The research consists of two parts. In the first part, we conducted RMT to extract representative decline/recovery patterns. Some companies could recover quickly and others might take longer days. To find the quick recovery patterns, we used RMT and found which principal component was corresponding to the quick recovery pattern.

Contrary to our expectations, the principal component extracted by RMT which we can interpret is just one; the second and third principal components patterns are complicated and so we cannot interpret the pattern. Then we tagged A/B/C/D to the companies, manually. The class A means its strong recover power. The automatic classification A/B/C/D is our future work.

As the second part of the analysis, our research focus is what determinant factors of the recovery power worked for the recovery. The methods we used were machine-learning based classification methods, using the above-mentioned tags A/B/C. The methods are Random Forest, XGBoosting, and Decision Tree models. The predictors are eight variables like ROA, ROE, and VAR. Among them, as dominant factors, ROA and VAR were selected. The index ROA (Return on Asset) is an indicator of how profitable a company is relative to its total assets. ROA gives us an idea as to how efficient a company's management is important for the recovery. The index VAR (value-added ratio) is a percentage of value added to sales. If the VAR is high, it can be said that the ratio of value newly created by a company is large. VAR gives us an idea as to a company with high VAR has strong recovery power. From the Decision Tree method, we could extract two different approaches to class A groups. The recovery and repulsion power will be higher in the company with high ROA and in the company that manufactured the product with high VAR. In addition, there is another approach to become a class A company. In the approach, the inventory turnover ratio was dominant as well as ROA. We will continue to conduct a lot of industry data analyses, concerning the recovery patterns.

REFERENCES

[1] *Nikkei Newspaper*, p. 17, 2019.
[2] 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.
[3] 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, June 27, 2002.
[4] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, 1999.
[5] M. Potters, J.-P. Bouchaud, and L. Laloux, "Financial applications of random matrix theory: Old laces and new pieces," arXiv preprint physics/0507111, 2005.
[6] M. C. Münnix *et al.*, "Identifying states of a financial market," Scientific Reports, vol. 2, p. 644, 2012.
[7] S. Maslov, "Measures of globalization based on cross-correlations of world financial indices," *Physica A: Statistical Mechanics and its Applications*, vol. 301, no. 1-4, pp. 397-406, 2001.
[8] D. Wang, B. Podobnik, D. Horvatić, and H. E. Stanley, "Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices," *Physical Review E*, vol. 83, no. 4, p. 046121, 2011.
[9] L. S. Junior and I. D. P. Franca, "Correlation of financial markets in times of crisis," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 1-2, pp. 187-208, 2012.
[10] S. Kumar and N. Deo, "Correlation and network analysis of global financial indices," *Physical Review E,* vol. 86, no. 2, p. 026101, 2012.
[11] A. Nobi, S. E. Maeng, G. G. Ha, and J. W. Lee, "Random matrix theory and cross-correlations in global financial indices and local stock market indices," *Journal of the Korean Physical Society*, vol. 62, no. 4, pp. 569-574, February 1, 2013.
[12] T. Kido, X. Yang, M. Tanaka-Yamawaki, and T. Takaishi, "Historical study of major sectors in japanese and american stock markets by means of RMT-based principal component analysis," *TOM*, vol. 4, no. 4, pp. 104-110, 2011.
[13] E. N. Suparno, S. K. Jo, K. Lim, A. Purqon, and S. Y. Kim, "Group identification in Indonesian stock market," *Journal of Physics: Conference Series*, vol. 739, no. 1, p. 012037, 2016.
[14] P. Bhadola and N. Deo, "Extreme eigenvector analysis of global financial correlation matrices," *Econophysics and Sociophysics: Recent Progress and Future Directions*, Springer, pp. 59-69, 2017.
[15] M. J. Bommarito and Duran, "Spectral analysis of time-dependent market-adjusted return correlation matrix," *Physica A: Statistical Mechanics Its Applications,* vol. 503, pp. 273-282, 2018.
[16] C. Kuyyamudi, A. S. Chakrabarti, and S. Sinha, "Emergence of frustration signals systemic risk," *Physical Review E*, vol. 99, no. 5, p. 052306, 2019.
[17] J. Friedman, R. Tibshirani, and T. Hastie, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2013.
[18] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[19] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices (Cambridge Studies in Advanced Mathematics)*. Cambridge University Press, 2009.

[20] J.-P. Bouchaud, M. Potters, G. Akemann, J. Baik, and P. D. Francesco, *Financial Applications of Random Matrix Theory* (The Oxford Handbook of Random Matrix Theory). Oxford University Press, 2011.

[21] J. Anderson, *Hands on Machine Learning with Python*, CreateSpace Independent Publishing Platform, 2018.

[22] N. Sharma, *XGBoost. The Extreme Gradient Boosting for Mining* Applications, Norderstedt, Germany, GRIN Verlag, 2017.

Kenji Yamaguchi was specially appointed as a lecturer in Science & Education Center (SEC), Ochanomizu University. His research fields are in stock price data, educational technology, and programming education. He has worked in the following workplaces: Computer Centre, Gakushuin University from 2010 to 2014; IT Center, Ochanomizu University from 2014 to 2017; Ochanomizu University Senior High School from 2017.

He is a member of Japanese Operations Management and Strategy Association (JOMSA), Japan Society of Business Mathematics, and Information Processing Society of Japan.

Yukari Shirota is a Prof. of Faculty of Economics, Gakushuin University. Her research fields are in data on the web, data visualization, social media analysis, and visual education methods for business mathematics. For over 20 years, she has developed visual teaching materials for business mathematics and statistics. She was invited and talked in many tutorials such as one titled "Visually Do Statistical Shape Analysis!" in DSAA 2017 (Data Science and Advanced Analytics). She organized the special session titled "Awareness computing for economics and social science analysis" in IEEE iCAST 2019 to provide the discussion place on the newest analysis methods.