

Performance Analysis of Japanese Manufacturing Industry before and after Lehman Shock

Yukari Shirota and Michiya Morita

Abstract—In the paper, we analyze determinants of the operating income to net sales ratio of Japanese manufacturing companies from 2004 to 2010 in-between Lehman crisis was put. The period involves up and down phases. The focus of this analysis is to identify determinants of the profitability in this period involving a swing of business conditions. We tried various regression methods for the analysis to understand those methodologies' characteristics as well. They are Ridge, Lasso and Elastic Net regressions and decision tree-based regressions such as Random Forest and XGBoost. One implication we drew from the analyses suggests those companies that offered high value added product lines sustained relatively high profitability in this period. This implication is plausible because most of companies were at a loss or unable to adapt their supply processes' fixed capacities and operating conditions to drastically changing business conditions such as a sudden crisis. Only high value added offerings could sustain relatively high profitability during the period.

Index Terms—Operating income to net sales ratio, Japanese manufacturing industries, ridge regression, lasso regression, random forest regression, XGBoost regression.

I. INTRODUCTION

In the study, we (shall) conduct analyses on the profitability determinants of Japanese manufacturing companies during the period from 2004-2010 putting Lehman Crisis in-between. Lehman Crisis happened in September, 2008 which severely affected Japanese manufacturing companies. Stock prices of those companies decreased and a recessionary environment prevailed. Our research focus is what determinant factors of the profitability worked during this period. We shall investigate during the recession period which factors affect the performances of companies. The methods we use are various regression methods. When we conducted the machine learning based regression methods under Python frameworks, rather than implementing the algorithms from scratch. Scikit-Learn (<http://scikit-learn.org>) is one of the most widely used frameworks and offering many machine learning algorithms efficiently. We also use Scikit-Learn library for this analysis.

In the paper, we try to use a machine learning based regression to show the potentiality of those methods in management area where traditional regression methods still dominate. References of those methods are [1]-[3] below.

Manuscript received September, 5, 2019; revised December 3, 2019. This work was supported in part by the Gakushuin Computer Centre project Grant in 2019.

Yukari Shirota and Michiya Morita are with Faculty of Economics, Gakushuin University, Japan (e-mail: yukari.shirota@gakushuin.ac.jp, michiya.morita@gakushuin.ac.jp).

II. ANALYSIS GOAL

There are many potential determinants of profitability. In this study, we set the ratio of operating profit over sales as a profitability measure. As determinants of the profitability, we restrict to accounting measures related to personnel, committed physical resources and committed personnel 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 in changing business circumstances. We assume, if there exist certain desirable managerial patterns responding to changeable situations, such patterns could emerge in relationships between those resource commitment measures. Then predictors are prepared as follows:

0. total capital used [million JPY]
1. total operating profit ratio on used capital (ROA) [%]
2. return on equity ratio (ROE) [%]
3. inventory turnover ratio [times]
4. turnover of tangible fixed assets [times]
5. sales per employee [million yen]
6. operating profit per employee [million yen]
7. labor productivity [JPY]
8. value-added ratio [%]
9. labor rate [JPY]
10. sales growth rate [%]
11. total capital (total assets) 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 average values of all involved variables of four years from 2007 to 2010.

The data are all standardized as shown in Fig. 1. Therefore all the figures are dimensionless. We used the data of 222 Japanese manufacturing companies that are listed in the First Section of Tokyo Stock Exchange Market and picked up from three industries, Electronics, Machinery and Vehicle. These industries typically represent the Japanese industry. The sample data size is 222.

We conducted the aforementioned regression analyses to find significant patterns among the 12 variables mentioned above. As the performance measure of prediction, we use the determination of coefficient [4], [5]. The analyses proceed 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.

There are in general two purposes of a regression 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.

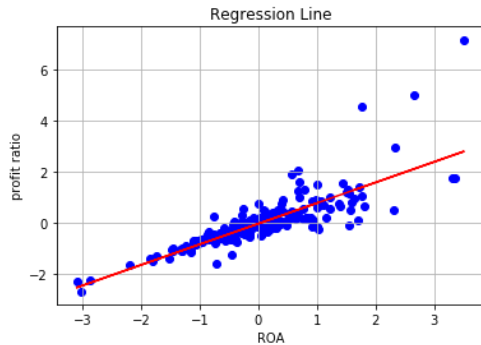


Fig. 1. All the data are standardized.

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 [4]. 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 regression model has hyperparameters to be set.

To find the best hyperparameter combination, we have to repeatedly conduct the training. To do that, Scikit-Learn's GridSearchCV is convenient which will evaluate all the possible combinations of hyperparameter values, using Cross-Validation. Finally the GridSearchCV 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.

III. LINEAR REGRESSION ANALYSIS

In the section, first we shall show the result by a traditional linear regression which minimizes the mean squared error [3]. Then we shall conduct the Ridge regression and the Lasso regression as the regularized models [4]. Finally the result of the mixed type of the Ridge and Lasso named Elastic Net regression will be shown. The cost functions of them are as follows:

$$\text{Ridge regression} \quad J(\theta) = \text{MSE}(\theta) + \lambda \sum_{i=1}^n (\theta_i)^2$$

$$\text{Lasso regression} \quad J(\theta) = \text{MSE}(\theta) + \alpha \sum_{i=1}^n |\theta_i|$$

Elastic Net regression

$$J(\theta) = \text{MSE}(\theta) + r\alpha \sum_{i=1}^n |\theta_i| + (1-r)\alpha^2 \sum_{i=1}^n (\theta_i)^2$$

θ is the model's parameter vector, containing the bias term and the predictor weights. Mean Square Error (MSE) is defined as $SE(X, h) = \sum (h(X^{(i)}) - y^{(i)})^2$.

When we use the Elastic Net regression, we set r to be 0.5. First we should conduct Ridge regression and if there are a few dominant predictors then we should try Lasso. In general, Elastic Net is preferred over Lasso since Lasso may behave erratically when the number of predictors is greater than the number of training instances or when several predictors are strongly correlated [6].

In Fig. 2 to Fig. 4, the results of coefficient magnitudes learned by Ridge, Lasso and Elastic regressions are shown. The three results show the same predictors as important ones. They are #6 (operating profit per employee POSITIVE), #7 (labor productivity NEGATIVE), #8 (value-added ratio POSITIVE), and #2 (ROA POSITIVE). The three regression method and the simple linear regression show the same top four predictors. Therefore, the result is reliable.

Then let us compare the R squared scores with different values of α . We split the data to training data and test data by the ratio 8:2.

The Ridge result was

RIDGE $\alpha=1$ Training set score: 0.95

RIDGE $\alpha=1$ Test set score: 0.91

RIDGE $\alpha=10$ Training set score: 0.92

RIDGE $\alpha=10$ Test set score: 0.89

RIDGE $\alpha=0.1$ Training set score: 0.96

RIDGE $\alpha=0.1$ Test set score: 0.91 .

In the Ridge regression, regression with $\alpha=0.1$ offers the best score set (0.96, 0.91).

The Lasso result was

LASSO $\alpha=1$ Training set score: 0.00

LASSO $\alpha=1$ Test set score: -0.31

LASSO $\alpha=1$ Number of features used: 0

LASSO $\alpha=0.01$ Training set score: 0.95

LASSO $\alpha=0.01$ Test set score: 0.91

LASSO $\alpha=0.01$ Number of features used: 9

LASSO $\alpha=0.001$ Training set score: 0.96

LASSO $\alpha=0.001$ Test set score: 0.91

LASSO $\alpha=0.001$ Number of features used: 12

In the Ridge regression, regression with $\alpha=0.001$ offers the best score set (0.96, 0.91).

The Elastic Net result was

ELASTIC $\alpha=0.1$ Training set score: 0.88

ELASTIC $\alpha=0.1$ Test set score: 0.87

ELASTIC $\alpha=0.1$ Number of features used: 6

ELASTIC $\alpha=0.01$ Training set score: 0.95

ELASTIC $\alpha=0.01$ Test set score: 0.91

ELASTIC $\alpha=0.01$ Number of features used: 10

ELASTIC $\alpha=0.001$ Training set score: 0.96

ELASTIC $\alpha=0.001$ Test set score: 0.91

ELASTIC $\alpha=0.001$ Number of features used: 12

In the Elastic Net regression, regression with $\alpha=0.001$ offers the best score set (0.96, 0.91).

In conclusion of the linear regression analysis, the same coefficients were selected and the amplitudes were almost the same when the parameter α was best arranged in each regression model. Then, to clarify whether the result is

correct, another regression method by Random Forest will be conducted in the next section.

IV. RANDOM FOREST AND XGBOOST REGRESSION ANALYSIS

A Random Forest is an ensemble of decision trees [6]. XGBoost stands for extreme gradient boosting and belongs to Tree ensemble methods [7]. The XGBoost algorithm is described in [4]. In the section, the results by the two methods are shown. We call them decision tree-based models.

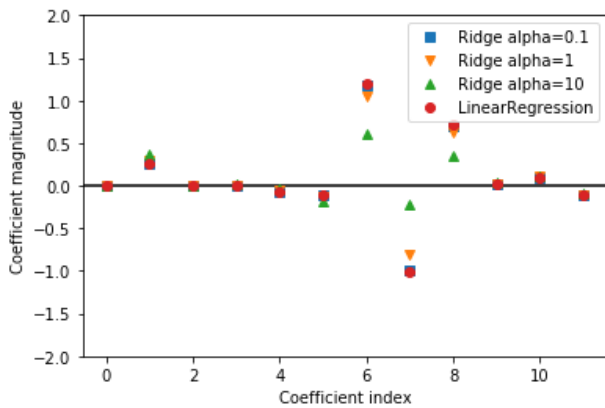


Fig. 2. Coefficients learned by Ridge regression for different values of α .

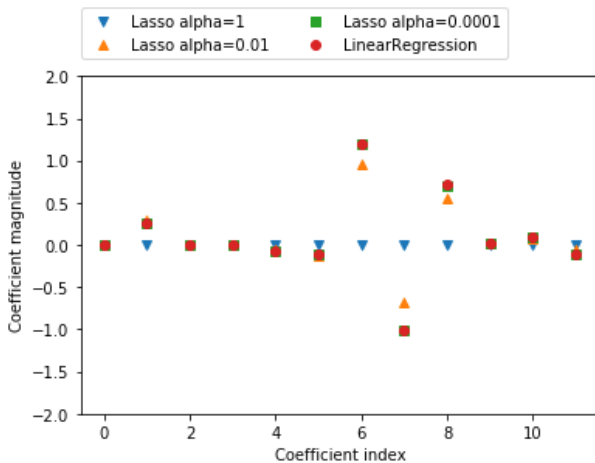


Fig. 3. Coefficients learned by Lasso regression for different values of α .

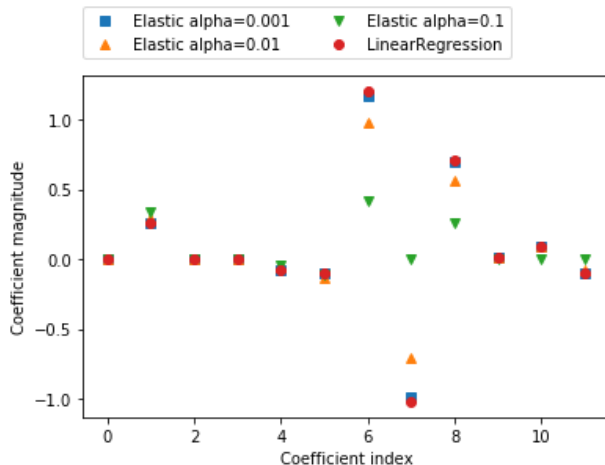


Fig. 4. Coefficients learned by Elastic regression for different values of α .

When we use decision tree-based models for a regression,

the difference compared to linear regressions such as Ridge and Lasso is that the decision tree-based model does not return positive/negative of each predictor's weight (coefficient). 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 [6]. The relative importance value is just positive or zero. To identify if the effect of the predictor on the target variable is positive or negative, we should make a scatter graph of them as shown in Fig. 1.

A. Random Forest Regression

In the left graph of Fig. 5, the importance values of 12 predictors by Random Forest are shown. The importance value is a relative one and the total of them becomes one. The important predictors are #1 (ROA), and #8 (value-added ratio). The left graph is one by XGBoost in Fig. 5. The both show #8 (value-added ratio) is an important predictor. In the Elastic Net regression result, #6 (operating profit per employee) and #7 (labor productivity) have large coefficient values which are about 1 and -1 and one of #1 (ROA) is about 0.25 and small. However, the results by tree decision-based model the most important ones are #8 and #1.

Let's see the parameter comparison (See Fig. 6). Changing three parameters, we looked for the best parameter set. The parameter **n_estimators** is the number of trees in the forest. The parameter **max_features** is the number of selected features when the algorithm randomly selects a subset of features. In general, it is a good rule of thumb to use **max_feature=n_features** (in this case 12) for regression [6]. The best parameter set selected by GridSearchCV was as follows:

```
RandomForestRegressor(bootstrap=True, criterion='mse',
max_depth=7, max_features=8, max_leaf_nodes=None,
min_impurity_split=1e-07, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=50, n_jobs=1, oob_score=False,
random_state=None, verbose=0, warm_start=False).
```

In the resultant best parameter set, **max_features=8** was selected. To avoid overfitting, we set **max_depth** which is a tree depth constraint to be small which are 5, 6 and 7. The **n_estimators** (the number of trees) was calculated to be 50.

B. XGBoost Regression

In the right graph of Fig. 5, the importance values of 12 predictors by XGBoost are already shown. Let us explain here the parameter calculated. We arranged three parameters which are **n_estimators** (the number of trees to be combined), **max_depth** of each tree, and **learning_rate**.

The parameter **learning_rate** controls how strongly each tree tries to correct the mistakes of the previous trees. To reduce overfitting, we could either apply stronger pre-pruning by limiting the **max_depth** or lower the **learning_rate** [6].

Using the best parameter set, we calculated the scores:

Train score R2: 0.97 test score R2: 0.81

Fig. 6 shows the parameter set comparison. Finally, the best parameters were selected by GridSearchCV with CV=3 as follows:

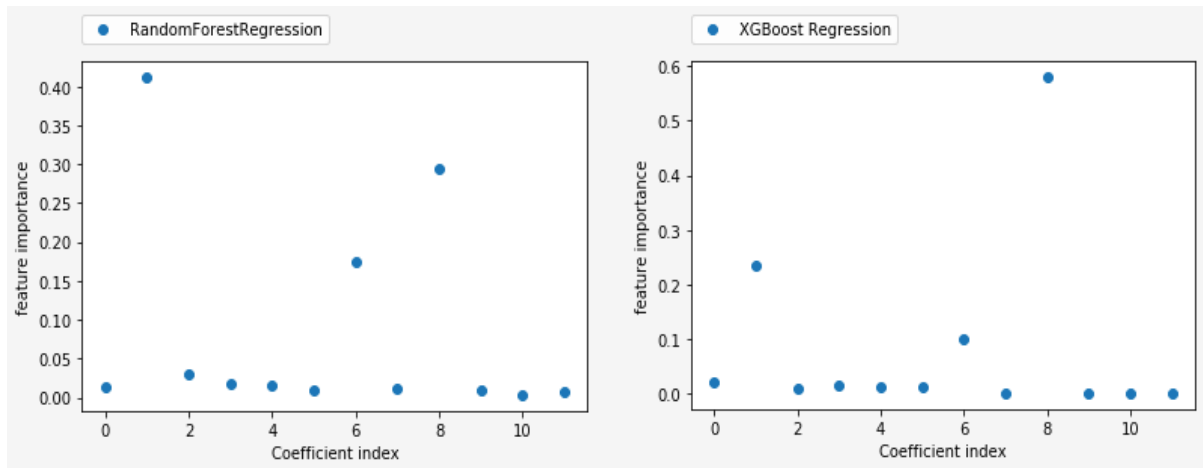


Fig. 5. Relative importance of 12 predictors.

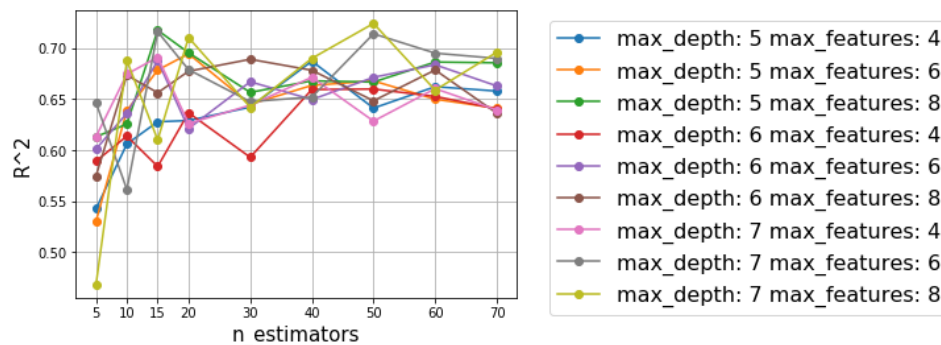


Fig.6. Three parameter value comparison in the Random Forest.

XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bytree=1, gamma=0, importance_type='gain', **learning_rate=0.2**, max_delta_step=0, **max_depth=3**, min_child_weight=1, missing=None, **n_estimators=100**, n_jobs=1, nthread=None, objective='reg:linear', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=0, silent=True, subsample=1)

Using the best parameter set, we calculated the scores:

Train score R2: 0.99 test score R2: 0.81

V. EVALUATION

Seeing the results from linear regressions and decision tree-based regressions, we shall evaluate which predictor is an important index.

The target variable is operating income to net sales ratio [%]. As candidates, four predictors were selected: they are ROA, operating profit per employee, labor productivity, and value-added ratio. 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 operating profit per employee is the figure which is operating income from the income statement is divided by the number of employees needed to produce that revenue. The index labor productivity is the amount of goods and services that workers produce in a given amount of time. The index value-added ratio is a percentage of value added to sales. If the value added ratio is high, it can be said that the ratio of value newly created by a company is large.

Let's calculate correlation coefficient between the target

variable and each predictor. The correlation coefficients are {-0.0338546, 0.823017, 0.438363, 0.0656191, -0.00453574, 0.127255, 0.801499, 0.759537, 0.825566, 0.136, 0.302369, 0.474519}.

The selected four have high values greater than 0.75. The greatest one is one by value-added ratio and the second one is one by ROA. We think that the most important predictor is value-added ratio (VAR). The correlation coefficient is a simple correlation between a target variable and a predictor variable. If we would like to express a cross effect of two predictors on the target variable, what should we do? The decision tree-based model can, in general, express the interactive effect by using a tree with a depth n ($n > 1$), because we use the n dimensional world. Therefore, compared to the linear regression such as Lasso, a decision tree-based model has more descriptive power. Then, we think that in this case we should use the decision tree-based results.

Then which result should be used, one by Random Forest or one by XGBoost? This time the XGBoost result showed more robust than one by Random Forest, because we could control the performance by adjusting the number of estimators (See Fig. 7). Then we determined to use the XGBoost result and then we selected VAR as the most important factor.

Let us explain our interpretation. We suppose that many Japanese companies were likely to keep the assets and the workers as they are, even in a recession, which may be a virtue of a Japanese company. If the assets and the workers are kept as they are in a recessionary situation, the operating income to net sales ratio will be higher in the company that manufactured the product with high VAR.

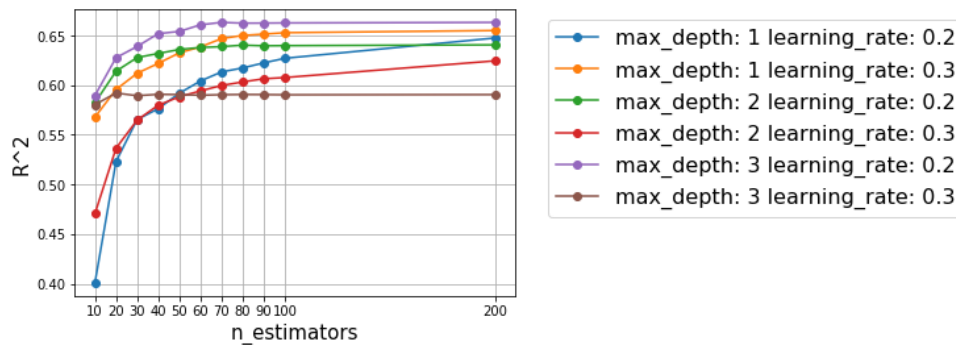


Fig. 7. Three parameter value comparison in the XGBoost.

VI. CONCLUSION

In the paper, we analyze operating income to net sales ratio of Japanese manufacturing industries before and after the Lehman shock. The period was a recessionary status for the companies. As a result, we consider that if companies keep the assets and workers as they are, instead of cut-backs, the operating income to net sales ratio will be higher in the company that manufactured the product with the high value-added ratio (VAR). We used various regression methods for the analysis. They are Ridge, Lasso and Elastic Net regressions and decision tree-based regressions such as Random Forest and XGBoost. When we conduct regressions, important things are (1) data standardization, (2) split training data and test data, (3) Cross-Validation, (4) Grid-search. The regression environment has been drastically changed before and after Scikit-Learn. Especially, regression libraries themselves, and Cross-Validation and Grid-search packages offered by Scikit-Learn are very helpful to data analysts. However, the reliability of the result would be changed by the parameter tuning. It is difficult to select the best parameter set. Via Scikit-Learn, we have studied the machine learning based regressions. To control and harness the machine learning approaches, we will continue to conduct a lot of industry data analyses.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENT

This work was supported in part by the Gakushuin Computer Centre project Grant in 2019.

REFERENCES

- [1] S. Raschka, *Python Machine Learning*, Packt Publishing Ltd., 2015.
- [2] G. Hackeling, *Mastering Machine Learning with Scikit-Learn*, Packt Publishing Ltd., 2017.
- [3] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.
- [4] J. Friedman, R. Tibshirani, and T. Hastie, "Data mining, inference, and prediction: With 200 full-color illustrations," in *The Elements of Statistical Learning*, Springer, 2013.

- [5] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, O'Reilly Media, 2016.
- [6] J. Anderson, *Hands on Machine Learning with Python*, Create Space Independent Publishing Platform, 2018, p. 223.
- [7] N. Sharma, *XGBoost. The Extreme Gradient Boosting for Mining Applications*, Norderstedt, Germany: GRIN Verlag, 2017.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Yukari Shirota DSc. is a professor of Faculty of Economics, Gakushuin University. Her research fields are visualization of 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 ICAS 2019 to provide the discussion place on the newest analysis methods.

Various visual teaching material sites of her have been published on the webs which are freely available as follows:

<http://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/>
<http://www.gakushuin.ac.jp/univ/eco/english/teacher/sirota.html>
<http://www-cc.gakushuin.ac.jp/~20010570/VDStat/>
<http://www-cc.gakushuin.ac.jp/~20010570/mathABC/ABC/>



Michiya Morita is an emeritus professor of Gakushuin University. His research fields include strategic management, manufacturing management and supply chain management and quantitative analyses. His recent research works are presented at international conferences such as EurOMA, CSCP Annual Conference and the World Conference on POM (2000, 2004, 2008, 2012 and 2016) and published in academic journals such as International Journal of Production Economics, International Journal of Manufacturing Technology Management and CSCP Supply Chain Quarterly. He took roles for the World Conference on POM such as conference chair, steering committee and program committee. He is one of the founders of Japanese Operations Management and Strategy Association. He is the first president of this organization. His recent research focus on the mechanism of integrating product development and supply chain processes. Part of this research includes how to drive Industry 4.0 or leverage emerging information technologies for strengthening value creation processes of companies. Applications of emerging information technologies include AI analyses to improve environments of decision makings such as enriching information contents, timeliness, relevancy and coverage. He is a member of Japanese Operations Management and Strategy Association (JOMSA), European Operations Management Association (EurOMA), Production and Operations Management Society (POMs) and International System Dynamics Society.