

FORECASTING THE VULNERABILITY OF INDUSTRIAL ECONOMIC ACTIVITIES: PREDICTING THE BANKRUPTCY OF COMPANIES

Katsuyuki Tanaka, Kobe University
Takuo Higashide, Chuo University
Takuji Kinkyo, Kobe University
Shigeyuki Hamori, Kobe University

ABSTRACT

This study introduces a framework for forecasting the vulnerability of industrial economic activities based on the financial statements of companies. We consider the task of identifying the bankruptcy of a company as a classification problem and apply the random forest method to build a highly accurate bankruptcy prediction model. Based on the predictions of our model, we summarize the vulnerability of the economic activities of companies in different industries. We also apply the predicted probability of bankruptcy of our model as risk management of industries. Finally, we demonstrate that our bankruptcy model can predict credit default swaps to a significant degree. The high accuracy of the bankruptcy model and usage of more than 500,000 pieces of company data enable us to provide more precise and reliable forecasts of industrial economic activity.

Keywords: Industrial Economic Activities, Bankruptcy.

INTRODUCTION

Assessing the financial vulnerability of an industrial sector is an important aspect of measuring the health of economic activities. In particular, bankruptcy affects a company's subsidiaries, employees and other companies or clients tied closely in business and can damage the economy markedly. Hence, the financial stability of a company is valuable information that decision makers such as the government and analysts require to prevent a devastating chain reaction of bankruptcy and economic loss.

Studies of company bankruptcy date back to the 1960s (Altman, 1968; Beaver, 1966). Altman (1968) was one of the first researchers to predict the likelihood of bankruptcy events based on the financial statements of a company, using Z-scores and applying discriminant analysis and common business ratios such as working capital, retained earnings, earnings before interest and taxes, sales over total assets and the market value of equity over the book value of total liabilities. Another popular approach for predicting the probability of bankruptcy is to use logistic and probit models (Brédart, 2015; Hillegeist, Keating, Cram & Lundstedt, 2004; Lennox, 1999; Ohlson, 1980). Ohlson's (1980) O-score is another popular model for predicting financial distress. Ohlson (1980) employed logistic regression to selected variables such as the ratio of liabilities and income over total assets to predict bankruptcy events.

Some researchers have used a decision tree (Breiman, Friedman, Olshen & Stone, 1984) to predict corporate failures (Messier & Hansen, 1988; Shirata, 1999). A decision tree uses a

sequence of splitting rules to segment the space of explanatory variables to create the decision boundary of failures. Shirata (2003) developed a bankruptcy model by analysing the financial data of 1,426 bankrupt and 3,434 non-bankrupt companies. The author then built a linear model based on the extracted variables of the financial data by tree: Retained earnings to total assets, inventory turnover period, interest expenses to sales and net income before tax to total assets. It was found that this model has a stronger discriminative power than a conventional logistic model (Altman, 1968; Lennox, 1999; Ohlson, 1980).

Despite its long history, research on company bankruptcy has, however, focused on large companies such as those listed on stock markets and has rarely included non-listed companies. In addition, the samples used to build models are typically relatively small (Brédart, 2015; Ohlson, 1980; Shirata, 2003) and may have limitations in terms of generalization. Furthermore, after the global financial crisis of 2008-2009, more studies have examined the stability of the financial sector rather than that of the industrial sector (Frankel & Saravelos, 2012; Tanaka, Kinkyo & Hamori, 2016). However, since the conditions of the industrial sector strongly influence the financial sector and vice versa, studying the health of the industrial sector is important to detect early warnings. Finally, with increased access to big data, there is a much greater scope for adopting economic modelling in this domain (Varian, 2014).

In this study, we introduce a systematic approach to build a model to predict the probability of a firm's bankruptcy by using random forests (Breiman, 2001), a variant of decision trees that significantly improves classification accuracy. Based on the predictions of our model, we also develop a simple but highly useful framework to assess the vulnerability of industrial economic activities. Importantly, our approach differs from those of previous works because we use financial data on both listed and non-listed companies.

By using the model built by a dataset of more than 10,000 Japanese companies, we show that the random forest method outperforms conventional approaches in terms of prediction accuracy. Accordingly, we demonstrate the usefulness of our model by analysing the health of the industrial sector, using the measure of the financial vulnerability of over 500,000 Japanese companies predicted by the proposed random forest model. We also introduce the predicted probability of bankruptcy of our model as a candidate for risk management of industries since it provides high accuracy of such predictions. To the best of our knowledge, this is the first study to employ random forests to build such a model and to analyse the industrial sector using such a broad range of companies. Moreover, we also demonstrate the effectiveness of our bankruptcy model by showing that it can predict the effect on credit default swaps (CDS) to a significant degree.

The remainder of this paper is organized as follows. The next section describes the data and methodology as well as evaluates the performance of the random forest bankruptcy model. The third section forecasts industrial economic activity by assessing the financial vulnerability of Japanese companies. The fourth section introduces risk managements and the fifth section analyzes the extent to which the probability of bankruptcy influences CDS based on our model. The final section concludes.

METHODOLOGY AND EXPERIMENTS

Data

We collect the financial statements of industrial companies from the Orbis database (<http://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>), which lists various types of company information such as financial accounts, status (active or inactive) and M&A activities and covers millions of companies worldwide. The advantage of using this data source is that it provides a broad coverage of standardized data formats across countries.

We use 26 indicators derived from the Global Ratio category, which is classified into three groups: Profitability ratios, operational ratios and structure ratios (thirteen, seven and six variables, respectively). We select Japanese companies with consolidation codes C1 (consolidated accounts only), C2/U2 (both types of accounts) and U1 (unconsolidated accounts only). Our sample thus covers 653,827 companies, consisting of 585,566 active and 68,261 inactive companies. We use the last two annual financial statements of each company: The latest available year and the year before the latest available one.

Random Forest Model

The random forest method is applied in various areas including computer vision and bioinformatics. Random forests are popular because they are simple, flexible and applicable to a range of tasks including classification and regression. Moreover, they improve classification accuracy by building a large number of trees instead of only a single tree. Each decision tree is built by using randomly selected data samples and randomly selected input variables from the original data by choosing the best split of a variable at each node.

A popular algorithm for constructing decision trees is the Gini index (Breiman, 1984). The Gini index is the measurement of the best split criterion based on the impurity of each node. The algorithm aims to select the optimal splitting variable and corresponding threshold value by making each node as pure as possible. Suppose M_n is the number of pieces of information reaching node n and M_n^i is the number of data points belonging to class C_i . The Gini index, GI_n , of node n is obtained from

$$GI_n = 1 - \sum_{i=1}^K (p_n^i)^2, \text{ where } p_n^i = \frac{M_n^i}{M_n} \quad (1)$$

A smaller value of the Gini index of node n represents purity, which implies that the node contains more observations from a single class. Hence, a decreasing Gini index is an important criterion to split a node. After a large number of trees are generated, the prediction is made by voting for the most popular class from all the output of trees.

Compared with single tree modelling, random forests have several desirable features (Breiman, 2001; Kuhn & Johnson, 2013). First, random forests have higher classification accuracy as they build a large number of trees instead of only a single tree. Second, random forests are more able to be generalized and are robust to over-fitting; hence, they may have better out-of-sample accuracy because using a random selection of input variables to split each node and combining the results of multiple trees yield error rates that compare favourably to

alternative methods and are more robust with respect to noise. This also circumvents the over-fitting problem of decision trees. Third, random forests can better handle large datasets because they enable training multiple trees in parallel efficiently. Finally, random forests provide a measure of the relative contribution of each variable to generate a prediction, which helps identify the variables important for distinguishing between active and inactive companies and thus predicting bankruptcy.

We assume that active and inactive companies in the latest available financial statement are financially stable and unstable, respectively. We also assume that the financial conditions influence a company's status, identifying patterns by distinguishing the differences systematically in its financial statements. In contrast to existing studies that are more concerned with identifying the key predictors of bankruptcy, we thus prioritize improving prediction accuracy by using random forests rather than conventional methods such as logistic regressions, which are widely used in the literature.

To avoid model bias created by an imbalanced training set, we even out the sample sizes of active and inactive companies. We randomly sample 50,000 companies from the groups of active and inactive companies as a training dataset to build our model and use the remaining companies as a test dataset (535,566 active and 18,261 inactive companies). In addition, we eliminate a variable if more than 50% of its values are missing (i.e., 50% of the 50,000 active and inactive companies for each variable). As a result, four such variables are eliminated and 19 variables are used for the experiments.

EXPERIMENTAL RESULTS

To evaluate our model's accuracy, we measure the balanced classification accuracy on the test dataset. We evaluate the performance of our model by comparing it with a logic model and a tree model, which are built by using the same experimental setup described above. The baselines of the logic and tree models scored 63.93% and 72.82% in classification accuracy, respectively, while our random forest model scored 78.74%. Hence, our proposed model significantly outperformed the other baseline models, confirming that a random forest model is more reliable for predicting the bankruptcy of companies than baseline models.

The variable importance measurement of random forests helps identify which variables are important for distinguishing between active and inactive companies. Hence, this should provide some clues to understanding the underlying causes of company failure. As shown in Figure 1, the three most important variables found in our model are as follows: (i) Credit period days: $(\text{Creditors}/\text{Operating revenue}) * 360$; (ii) Gross margin: $(\text{Gross profit}/\text{Operating revenue}) * 100$; and (iii) ROA (Return on assets) using net income: $(\text{Net income}/\text{Total assets}) * 100$.

Note that the accuracy of the bankruptcy model is important for analysing the causality of bankruptcy, forecasting various economic activities and managing risk. Naturally, the higher the accuracy of the model, the more reliable these analyses are. In the following section, we introduce the various usages of our random forest model, such as forecasting industrial activities and managing risk.

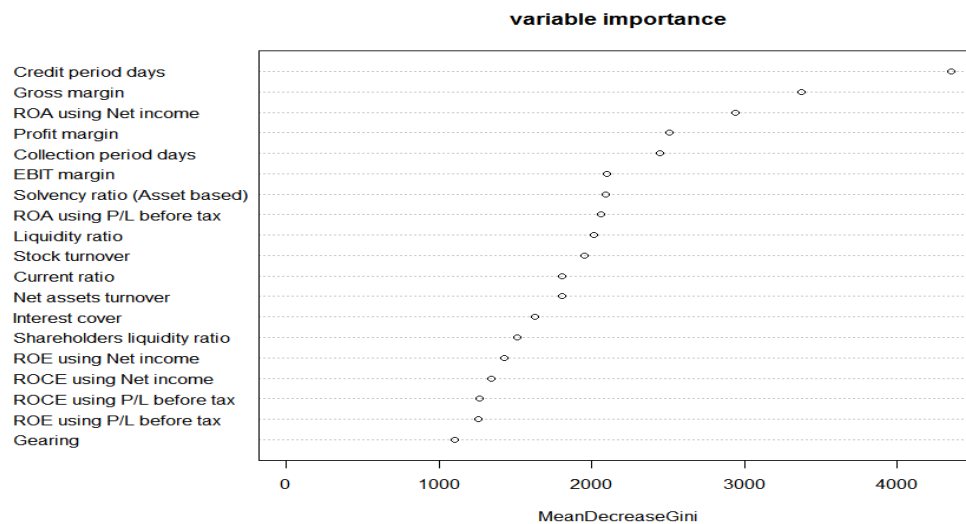


FIGURE 1
VARIABLE IMPORTANCE FOUND BY THE RANDOM FOREST MODEL

Forecasting Industrial Activities with the Random Forest Model

We analyse industrial economic activities based on their classification codes in the European Community (NACE Rev. 2) by using the likelihood of bankruptcy of Japanese companies predicted by our model. We examine these activities for each NACE code in terms of three factors: Operating revenue, number of employees and number of companies. Large losses in revenue and in the number of employees directly and indirectly affect a country's taxes, unemployment levels and public law and order as well as, consequently, its economic health and national strength. Therefore, forecasting industrial activity in these factors is a crucial part of vulnerability analysis on the economic health and stability of industries.

To forecast industrial activity, firstly, we obtained the likelihood of bankruptcy for the 585,566 active Japanese companies by feeding the latest available financial statements into our model. We define a company predicted to have a more than a 50% probability of becoming inactive as in danger of bankruptcy. We analyse the country-level industrial vulnerabilities for each industry based on NACE code with in-danger companies. We aggregate the operating revenue, number of employees and number of companies of in-danger companies and we claim the amount and the ratio of these factors as the economic loss or vulnerability, of each industry. Table 1 shows the total operating revenue, total number of employees and total number of companies for each NACE code in Japan and Figure 2 describes the distribution of the predicted probabilities of failure with a 10-percentage-point interval bases on these three measurements.

Table 1 TOTAL OPERATING REVENUE, TOTAL NUMBER OF EMPLOYEES AND TOTAL NUMBER OF COMPANIES FOR EACH NACE CODE			
NACE Rev. 2 main section	Total Operating revenue (mil in \$US)	Total Number of employees	Total Number of Companies
A-Agriculture, forestry and fishing	17,706,659	55,547	4,764

B-Mining and quarrying	27,463,126	22,907	578
C-Manufacturing	4,155,409,477	12,011,927	46,365
D-Electricity, gas, steam and air conditioning supply	267,750,008	278,530	534
E-Water supply; sewerage, waste management and remediation activities	25,861,398	121,594	4,006
F-Construction	1,034,983,220	3,208,405	320,206
G-Wholesale and retail trade; repair of motor vehicles and motorcycles	3,035,945,561	3,681,700	65,182
H-Transportation and storage	518,364,933	2,089,133	11,149
I-Accommodation and food service activities	88,933,005	300,591	3,971
J-Information and communication	627,494,271	1,578,192	10,539
K-Financial and insurance activities	187,279,527	236,148	2,876
L-Real estate activities	265,338,824	522,424	36,205
M-Professional, scientific and technical activities	220,080,817	800,283	22,873
N-Administrative and support service activities	230,513,557	880,501	9,451
O-Public administration and defence; compulsory social security	215,323	1,129	3
P-Education	17,708,991	96,689	675
Q-Human health and social work activities	187,265,693	1,447,959	38,593
R-Arts, entertainment and recreation	126,427,440	126,243	1,955
S-Other service activities	143,885,743	297,262	5,641

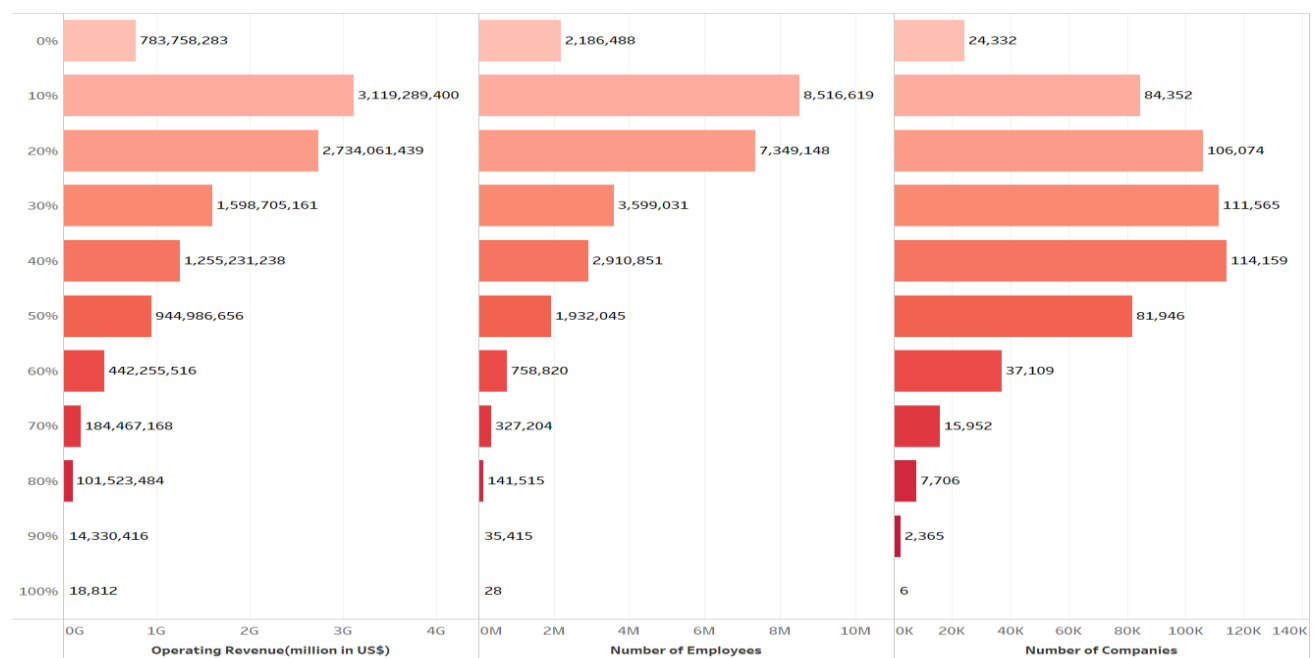


FIGURE 2
DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL BASIS ON OPERATING REVENUE, NUMBER OF EMPLOYEES AND NUMBER OF COMPANIES

Analysis of Industrial Activities based on Operating Revenue

The left column of Figure 2 shows the distribution of the predicted probabilities of failure with a 10-percentage-point interval based on the total operating revenue and Figure 3 describes the total operating revenue of in-danger companies for each NACE code in descending order. Our model predicted that about 15.10% of total operating revenue of Japan is in danger (1,687,582,052 million in US\$). The worst three industries in terms of the absolute loss of operating revenue are “G-Wholesale and retail trade; repair of motor vehicles and motorcycles” (641,897,644), “C-Manufacturing” (323,805,391) and “F-Construction” (136,157,737), which account for 21.14%, 7.79% and 13.16% of the total operating revenue of each industry in danger, respectively.

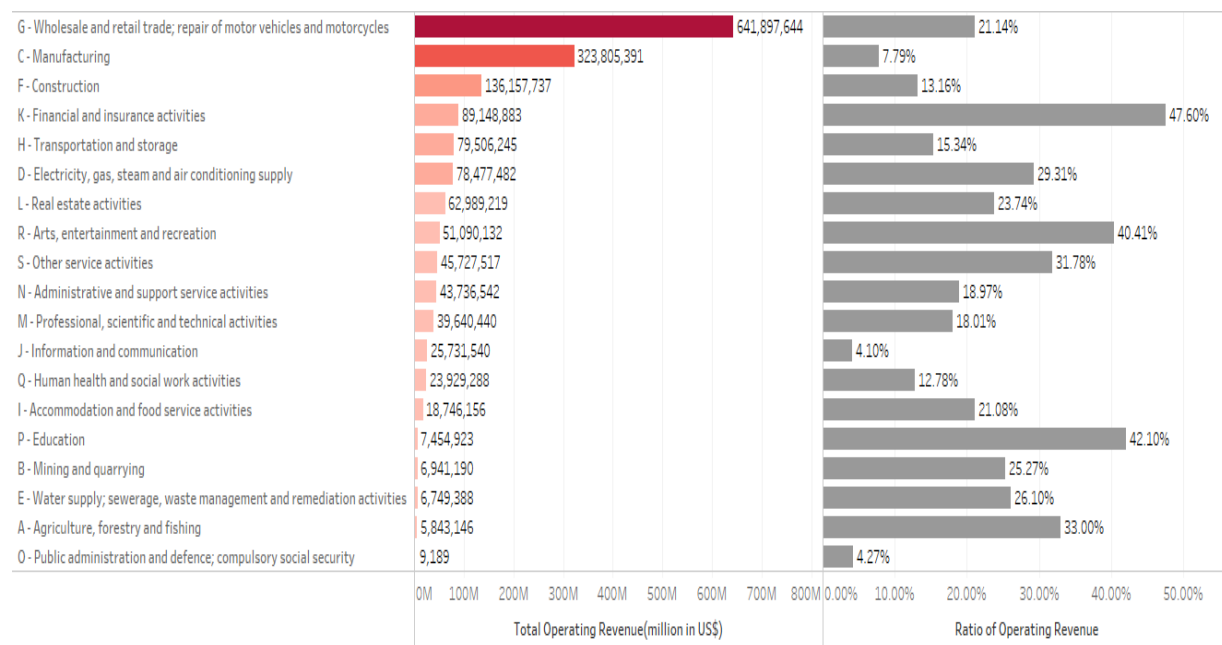


FIGURE 3
TOTAL OPERATING REVENUE (MILLION US\$) OF IN-DANGER COMPANIES FOR EACH NACE CODE IN DECENDING ORDER

We also analyse vulnerability in terms of the ratio of the industry-level operating revenue of in-danger companies. Figure 4 describes the sum of the operating revenue of companies predicted as in danger over the total operating revenue for each NACE code in descending order. According to our model, the most vulnerable industry is “K-Financial and insurance activities” (89,148,883), for which 47.60% of companies are predicted as being in danger, followed by 42.10% for “P-Education” (7,454,923) and 40.41% for “R-Arts, entertainment and recreation” (51,090,132).

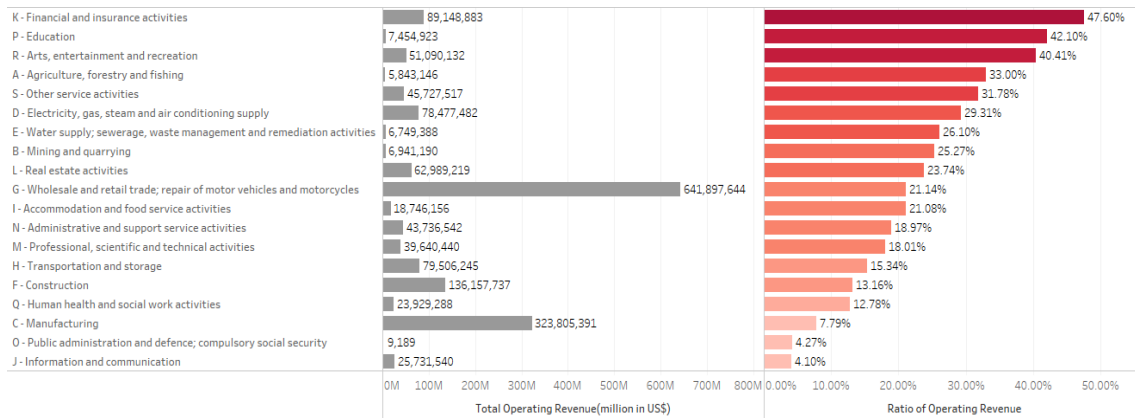


FIGURE 4
RATIO OF OPERATING REVENUE (MILLION US\$) IN DANGER FOR EACH NACE CODE IN DECENDING ORDER

Table 2 summarizes the worst three industries in terms of total operating revenue and the ratio of the operating revenue of in-danger companies. Although the ratios of total operating revenue of these three companies are relatively low since these are the largest industries in Japan, the impact of the absolute loss is huge. Our framework not only enables us to analyse industrial vulnerability but also induces the economic vulnerability of the country.

Table 2 THREE WORST INDUSTRIES IN TERMS OF THE OPERATING REVENUE OF IN-DANGER COMPANIES	
Total operating revenue in-danger companies (million US\$)	Ratio of operating revenue in-danger companies (million US\$)
G-Wholesale and retail trade; repair of motor vehicles and motorcycles 641,897,644 (21.14%)	K-Financial and insurance activities 89,148,883 (47.60%)
C-Manufacturing 323,805,391 (7.79%)	P-Education 7454923 (42.10%)
F-Construction 136,157,737 (13.16%)	R-Arts, entertainment and recreation 51,090,132 (40.41%)

Analysis of Industrial Activities based on the Number of Employees

The middle column of Table 1 describes the total number of employees over the predicted probabilities of failure with a 10-percentage-point interval. According to our model, about 11.51% (3,195,027 in total) of employees in Japan have more than a 50% chance of losing their current job. The three worst industries in terms of the total number of employees working for in-danger companies are “C-Manufacturing” (539,594), “G-Wholesale and retail trade; repair of motor vehicles and motorcycles” (536393) and “F-Construction” (499,810). The ratio with respect to each industry is 4.49%, 14.57% and 15.58%. The total number of employees working for in-danger companies for each NACE code is shown in Figure 5 in descending order.

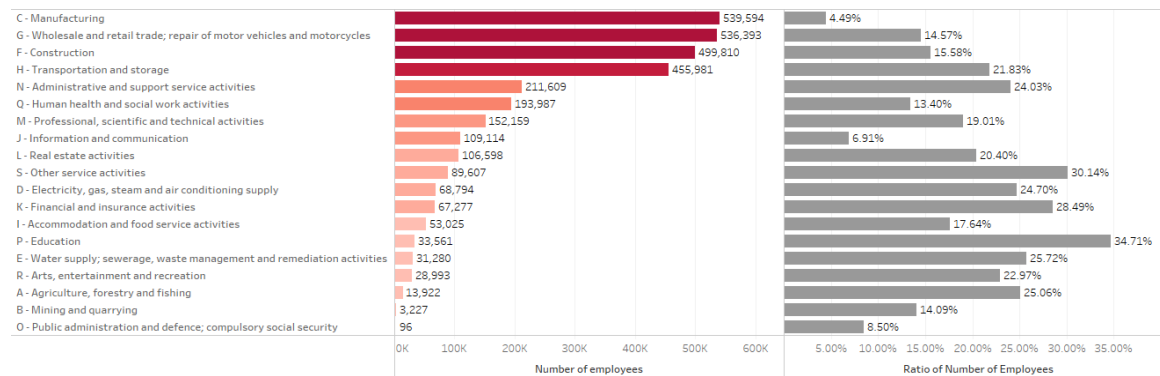


FIGURE 5
TOTAL NUMBER OF EMPLOYEES OF IN-DANGER COMPANIES FOR EACH NACE CODE IN DESCENDING ORDER

Figure 6 presents the ratio of the number of employees of in-danger companies in descending order, obtained from the sum of the number of employees of in-danger companies over the number of employees for each NACE code. Our model predicts that the “P-Education” industry has the highest loss rate with 34.71% (33,561 employees), followed by “S-Other service activities” and “K-Financial and insurance activities” with 30.14% (89,607) and 28.49% (67,277), respectively.

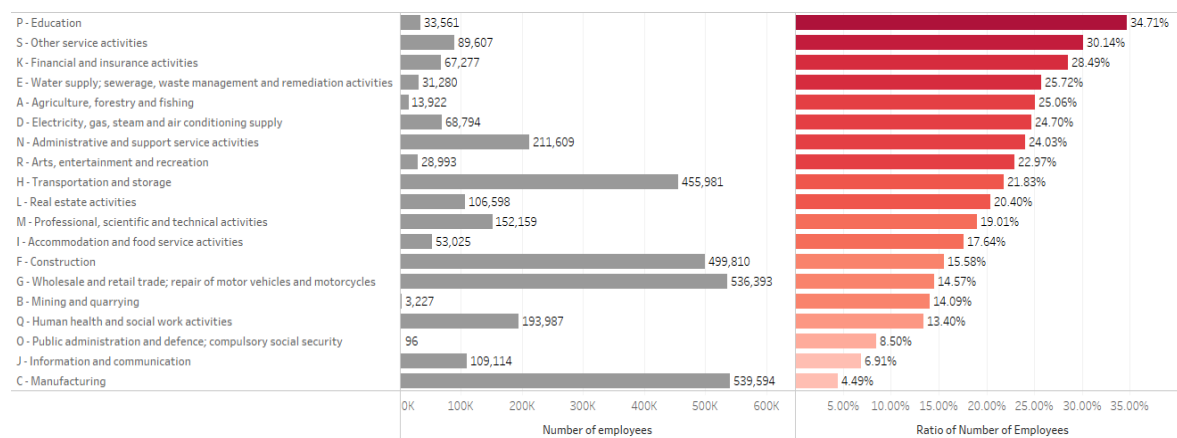


FIGURE 6
RATIO OF THE NUMBER OF EMPLOYEES OF IN-DANGER COMPANIES FOR EACH NACE CODE IN DESCENDING ORDER

Table 3 summarizes the worst three industries based on the total number of employees and ratio of the number of employees of in-danger companies. Again, the impact of the absolute loss of employees in large industries such as “C-Manufacturing” and “G-Wholesale and retail trade; repair of motor vehicles and motorcycles” is very high.

Table 3 THREE WORST INDUSTRIES IN TERMS OF THE NUMBER OF EMPLOYEES OF IN-DANGER COMPANIES	
Total number of employees of in-danger companies	Ratio of the number of employees of in-danger companies
C–Manufacturing 539,594 (4.49%)	P–Education 33,561 (34.71%)
G–Wholesale and retail trade; repair of motor vehicles and motorcycles 536,393 (14.57%)	S–Other service activities 89,607 (30.14%)
F–Construction 499,810 (15.58%)	K–Financial and insurance activities 67,277 (28.49%)

Analysis of Industrial Activities based on the Number of Companies

The analysis of vulnerability based on the number of companies is an important indicator of recovering from loss. As stated, the bankruptcy of one company may favour a competitor in the same industry. However, while loss may be compensated by competitors, if many companies in the same industry are at risk of bankruptcy, the recovery may slow; in the worst case, this can destroy the entire industry. Furthermore, for employees of in-danger companies, the bankruptcy of many companies increases the possibility of not being able to find a new job in the same industry. In other words, countries face a high unemployment risk. The right column of Table 4 shows the total number of companies over the predicted probabilities of failure with a 10-percentage-point interval. Figure 7 describes the total number of in-danger companies for each NACE code and Figure 8 is the ratio of the number of in-danger companies for each NACE code.

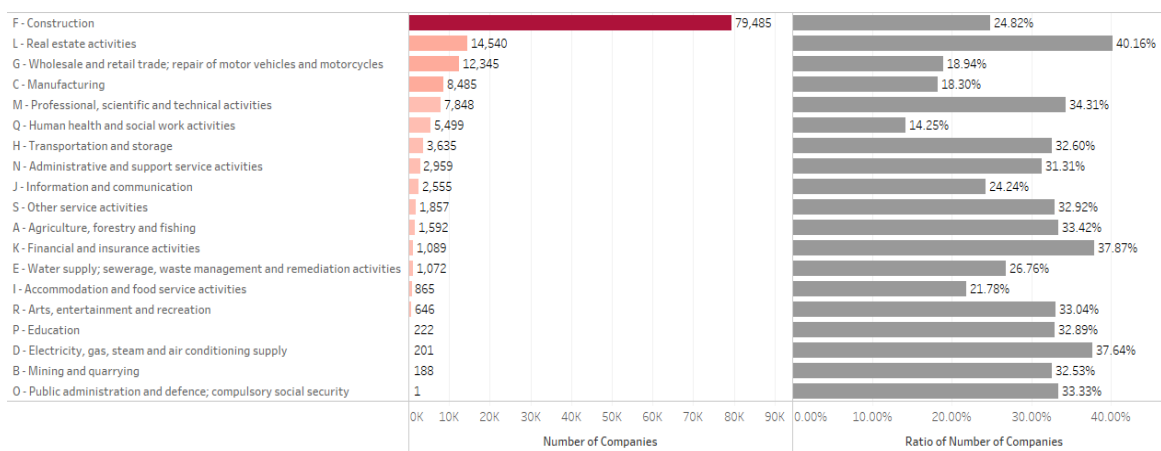


FIGURE 7
TOTAL NUMBER OF IN-DANGER COMPANIES FOR EACH NACE CODE IN DESCENDING ORDER

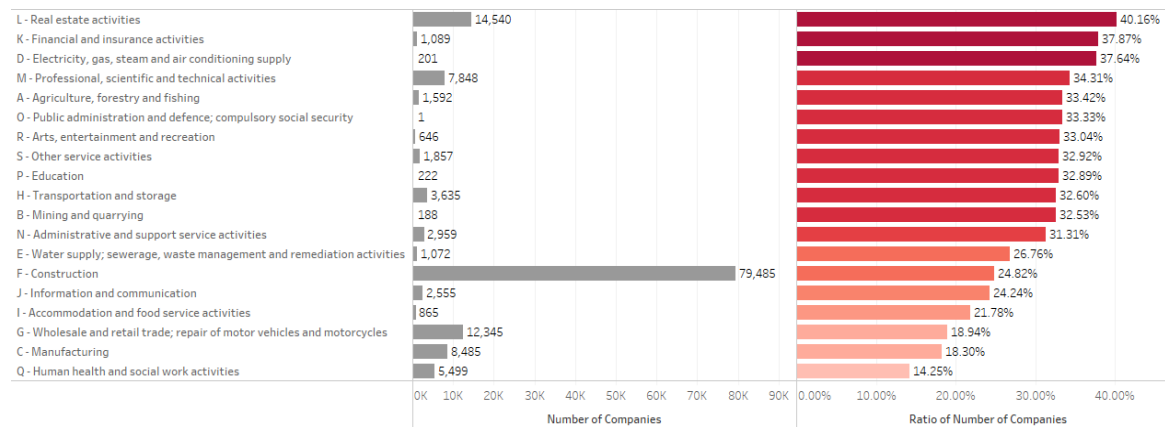


FIGURE 8
RATIO OF THE NUMBER OF IN-DANGER COMPANIES FOR EACH NACE CODE
IN DESCENDING ORDER

The results show that about 27.09% of companies in Japan (145,084) are at risk of bankruptcy with more than a 50% probability. The results of our model show that the most number of companies in-danger is in “F-Construction” (79,485), followed by “L-Real estate activities” (14,540) and “G-Wholesale and retail trade; repair of motor vehicles and motorcycles” (12,345) (the ratios are 24.82%, 40.16% and 18.94%, respectively). For the ratios, “L-Real estate activities” “K-Financial and insurance activities” and “D-Electricity, gas, steam and air conditioning supply” are the three worst industries, with ratios of 40.16%, 37.87% and 37.64%, respectively.

Table 4 THREE WORST INDUSTRIES IN TERMS OF THE NUMBER OF IN-DANGER COMPANIES	
Total number of in-danger companies	Ratio of the number of in-danger companies
F-Construction 79,485 (24.82%)	L-Real estate activities 14,540 (40.16%)
L-Real estate activities 14,540 (40.16%)	K-Financial and insurance activities 1,089 (37.87%)
G-Wholesale and retail trade; repair of motor vehicles and motorcycles 12,345 (18.94%)	D-Electricity, gas, steam and air conditioning supply 201 (37.64%)

Overall Vulnerability of Industries

The overall vulnerability of industries can be analysed by combining the aforementioned three indicators. The scatterplot in Figure 9 visualizes the correlation of the ratios of operating revenue, the number of employees and the number of companies of in-danger companies. Vulnerability increases as a larger circle located in the upper right corner of the coordinate. This shows that the “P-Education” and “K-Financial and insurance activities” industries are highly vulnerable, as are the “D-Electricity, gas, steam and air conditioning supply” and “A-Agriculture, forestry and fishing” industries. Our model predicts the risk of the “P-Education”

and “K-Financial and insurance activities” industries with the following ratios: Loss of operating revenue as 42.10% and 47.60%, the number of employees as 34.71% and 28.49% and the number of companies as 32.89% and 37.87%, respectively.

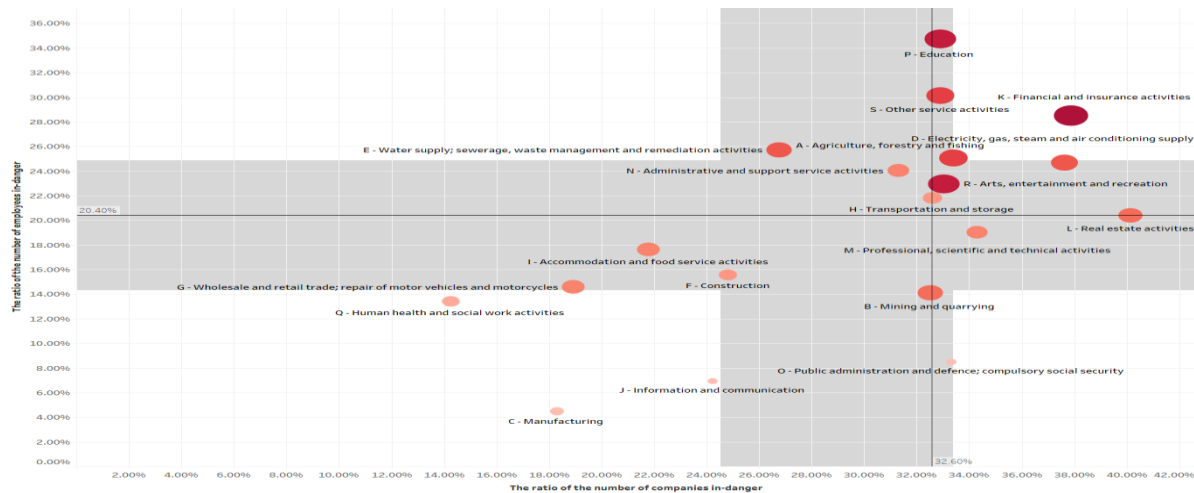


FIGURE 9
CORRELATION OF THE RATIOS OF OPERATING REVENUE, THE NUMBER OF EMPLOYEES AND THE NUMBER OF IN-DANGER COMPANIES FOR EACH NACE CODE. THE SIZE OF THE CIRCLE DESCRIBES THE RATIO OF THE OPERATING REVENUE OF IN-DANGER COMPANIES. THE SHADOWS INDICATE THE QUANTILES

Vulnerability Analysis at the Prefecture Level

Analysing the prefecture level of vulnerability based on industry is also informative for decision makers. As one example, we introduce the vulnerability analysis based on the prefectures in Japan by defining the weighted sum of the ratios of operating revenue, the number of employees and the number of companies of in-danger companies:

$$vulnerability_{p=prefecture} = \alpha * Ratio\ of\ operating\ revenue_p + \beta * Ratio\ of\ number\ of\ employees_p + \gamma * Ratio\ of\ number\ of\ companies_p \dots\dots\dots (2)$$

where $\alpha + \beta + \gamma = 1$. The output value of $vulnerability_p$ indicate the overall decline in the risk of industrial activities in prefecture p .

Figure 10a shows the vulnerability of prefectures based on the accumulated sum of the industries in each prefecture. The darker the colour of the prefecture, the more vulnerable it is. The three most vulnerable prefectures are Hokkaido, Chiba and Aomori and it is predicted that 25.32%, 24.29% and 23.15% of industry will decline in these prefectures. Figure 10b describes the heat map of the entire industry in each prefecture to provide a more detailed analysis based on the vulnerable industries found in the previous sections, namely “A-Agriculture, forestry and fishing,” “C-Manufacturing,” “D-Electricity, gas, steam and air conditioning supply,” “G-Wholesale and retail trade; repair of motor vehicles and motorcycles,” “K-Financial and insurance activities,” “L-Real estate activities,” “P-Education” and “S-Other service activities.”

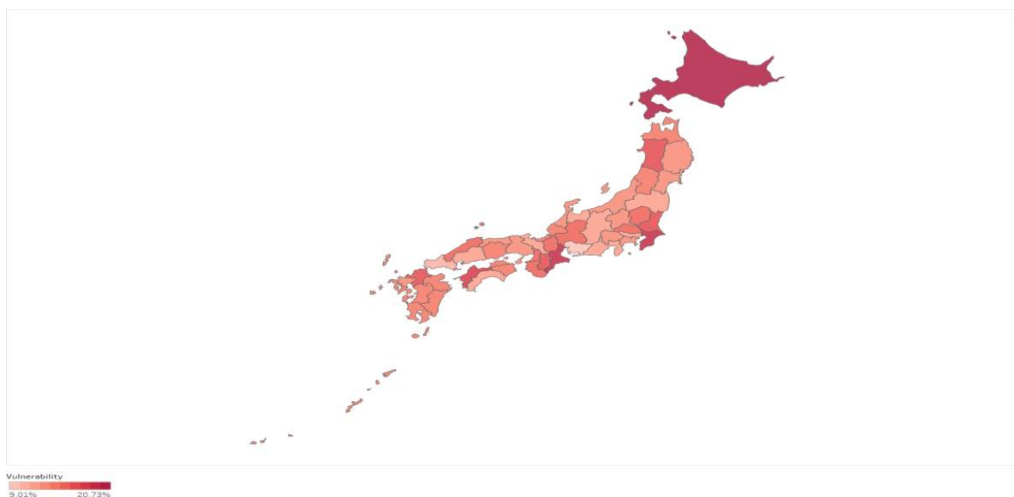


FIGURE 10a
VULNERABILITY OF PREFECTURES

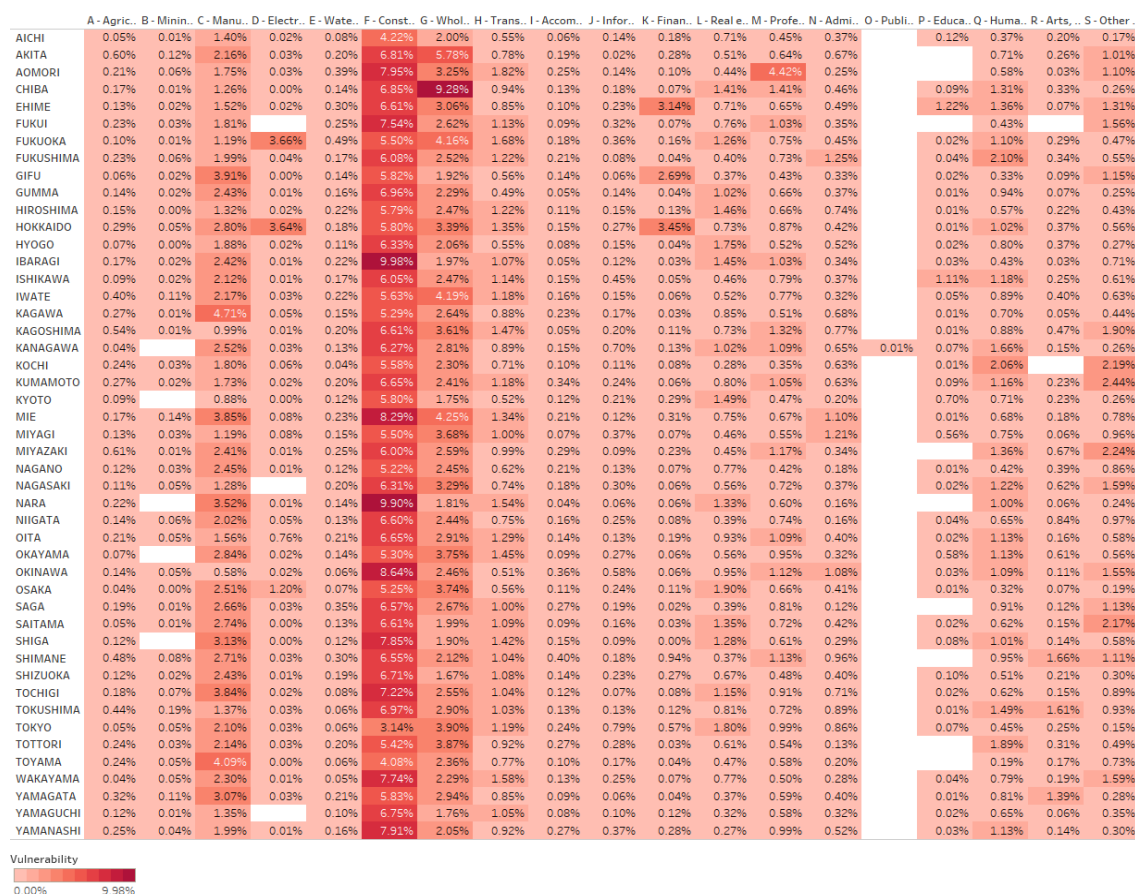


FIGURE 10b
HEATMAP OF THE ENTIRE INDUSTRY IN EACH PREFECTURE

Risk Management with the Random Forest Model

The financial health and stability of a company are important since they directly influence its subsidiaries, employees and other companies and clients closely tied in business. This information is also valuable for decision makers such as the government and analysts to prevent a chain reaction of bankruptcy and the risk of economic loss. The second wave of bankruptcy of companies may also hit the financial sector since it may indirectly affect financial institutions in the unrecoverable collection of claims. In the worst case, this might trigger a bankruptcy domino and consequently lead to financial crisis. Hence, the risk management of a company or industry is a critical issue to avoid various risks.

The predicted probability of bankruptcy of our model is a good candidate for such risk management since it provides highly accurate predictions. Various factors, as introduced in the third section, can be linked to the probability of bankruptcy for the vulnerability analysis. Naturally, the bankruptcy of a company with many subsidiaries, employees and/or clients poses a great risk not only to its industry but also to its country. Furthermore, the bankruptcy of smaller companies is also not negligible, since they are the foundation of certain industries. For some industries, the bankruptcy of many small companies damages its functioning.

Another advantage of the random forest model is that the importance variables can be used as vulnerability indicators for risk management. We further investigate the relationship between the value of the importance variables found by our model and the probability of bankruptcy. To analyse the characteristics of the three importance variables found in our model, namely “Credit period days,” “Gross margin,” and “ROA using net income,” we create a heat map based on the predicted probabilities of failure with a 10-percentage-point interval for each industry. Table 5 indicates the summary statistics for these variables per NACE code and Figures 11, 12 and 13 describe the heat maps for each variable for each NACE code.

Table 5
MEAN, MEDIAN, STANDARD DEVIATION AND VARIANCE OF EACH NACE CODE BASED ON “CREDIT PERIOD DAYS,” “GROSS MARGIN” AND “ROA USING NET INCOME”

	Credit period days				Gross margin				ROA using Net income			
	Mean	Median	Sdv	Var	Mean	Median	Sdv	Var	Mean	Median	Sdv	Var
A-Agriculture, forestry and fishing	18	3	47	2243	34	33	29	828	5	2	20	384
B-Mining and quarrying	33	21	54	2956	27	23	22	490	1	1	10	99
C-Manufacturing	38	29	37	1345	28	24	19	377	2	2	11	119
D-Electricity, gas, steam and air conditioning supply	17	9	29	842	45	35	35	1246	0	1	13	161
E-Water supply; sewerage, waste management and remediation activities	15	5	34	1144	47	39	31	937	3	2	11	119
F-Construction	23	13	36	1310	30	26	19	349	6	3	22	469
G-Wholesale and retail trade; repair of	40	28	43	1838	28	23	19	344	1	1	11	120

motor vehicles and motorcycles												
H-Transportation and storage	18	10	31	990	33	18	33	1119	2	2	10	110
I-Accommodation and food service activities	11	9	20	413	66	68	19	366	0	1	14	206
J-Information and communication	16	7	31	947	55	52	31	948	3	3	16	265
K-Financial and insurance activities	12	0	62	3817	81	100	31	951	3	2	15	236
L-Real estate activities	12	0	60	3595	69	83	34	1163	-1	1	19	348
M-Professional, scientific and technical activities	16	1	37	1401	51	44	31	980	3	2	19	365
N-Administrative and support service activities	19	5	43	1879	47	36	32	1048	2	2	14	196
0-Public administration and defence; compulsory social security	7	0	12	133	49	63	44	1935	-10	0	18	325
P-Education	3	0	7	50	75	92	30	919	-1	1	18	336
O-Human health and social work activities	2	0	7	54	10	3	28	788	2	2	13	164
R-Arts, entertainment and recreation	7	2	28	803	56	61	34	1157	0	1	15	213
S-Other service activities	22	7	55	2976	45	36	34	1175	1	1	12	135

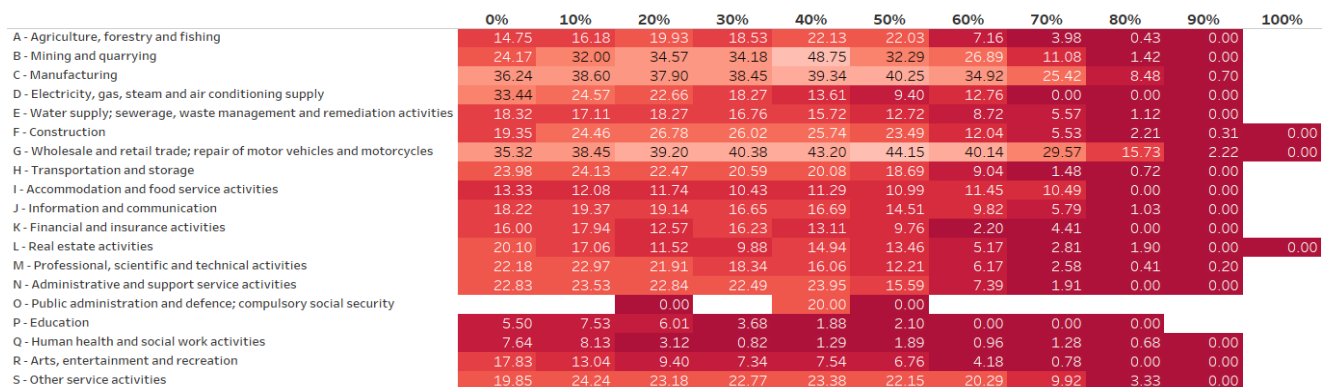


FIGURE 11
HEATMAP OF THE AVERAGE “CREDIT PERIOD DAYS” BASED ON THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL FOR EACH NACE CODE. “CREDIT PERIOD DAYS” IS THE MOST IMPORTANT VARIABLE FOUND TO DISCRIMINATE THE BANKRUPTCY OF A COMPANY



FIGURE 12
HEATMAP OF THE AVERAGE “GROSS MARGIN” BASED ON THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL FOR EACH NACE CODE. “GROSS MARGIN” IS THE SECOND MOST IMPORTANT VARIABLE FOUND TO DISCRIMINATE THE BANKRUPTCY OF A COMPANY

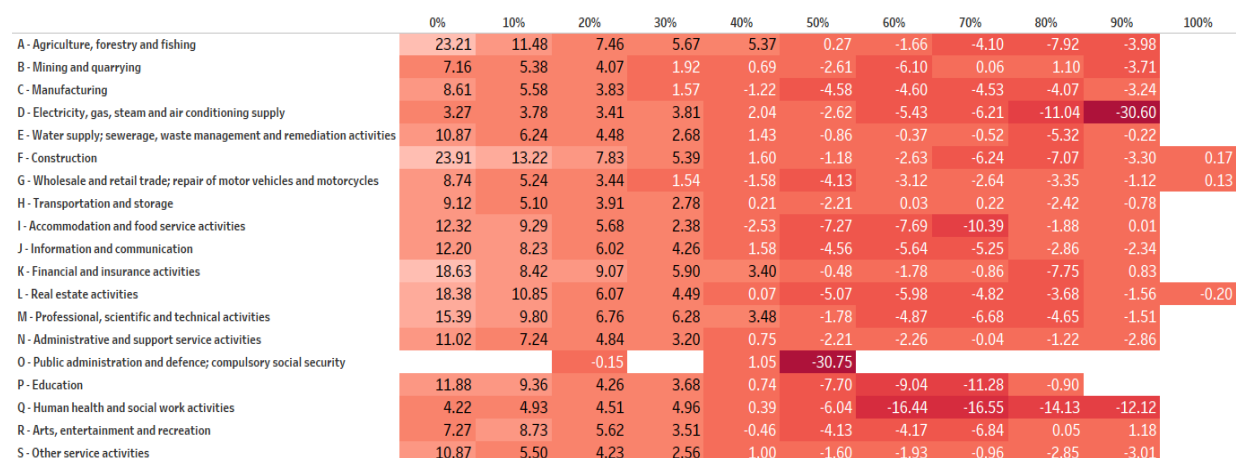


FIGURE 13
HEATMAP OF AVERAGE “ROA USING NET INCOME” BASED ON THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL FOR EACH NACE CODE. “ROA USING NET INCOME” IS THE SECOND MOST IMPORTANT VARIABLE FOUND TO DISCRIMINATE THE BANKRUPTCY OF A COMPANY

Since different industries have different financing structures, the threshold of bankruptcy differs. However, although performance is not comparable across industries, the values tend to decrease as the probability of bankruptcy increases for all industries. It sharply drops after 60% to 70% for both “Credit period days” and “Gross margin” and this is especially prominent for

“ROA using net income” since the value becomes negative after 50%. Because different industries have different financing structures, building a bankruptcy model for each NACE may allow us to capture such behavioral differences, but this is left to future work.

We also investigate the relationship between the predicted probability of bankruptcy of our model and the liquidity ratios since this is often used to evaluate the financial health of a company in practice. Liquidity ratios measure the ability of a company to pay off its debt obligations. The liquidity ratio is critical for the cash management of a company, as it indicates cash flow positioning. In general, a high liquidity ratio indicates a higher margin of safety for the company. A company with a lower liquidity ratio will have difficulty meeting its debt obligations. We obtain the average liquidity of companies based on the predicted probabilities of failure with a 10-percentage-point interval. As shown in Figure 14, generally, the average liquidity ratio of in-danger companies with more than 70% probabilities is comparatively lower than those of safer companies for all industries. One exception is “D–Electricity, gas, steam and air conditioning supply,” which shows the reverse result, namely lower liquidity for a lower probability of bankruptcy.

Although the question of the liquidity threshold in each industry remains, we found the predicted probability of failure to be consistent with the liquidity ratio. Our random forest modelling framework found that the liquidity ratio is only the ninth most important variable to identify the bankruptcy of a company, which indicates that more important variables for vulnerability management exist.

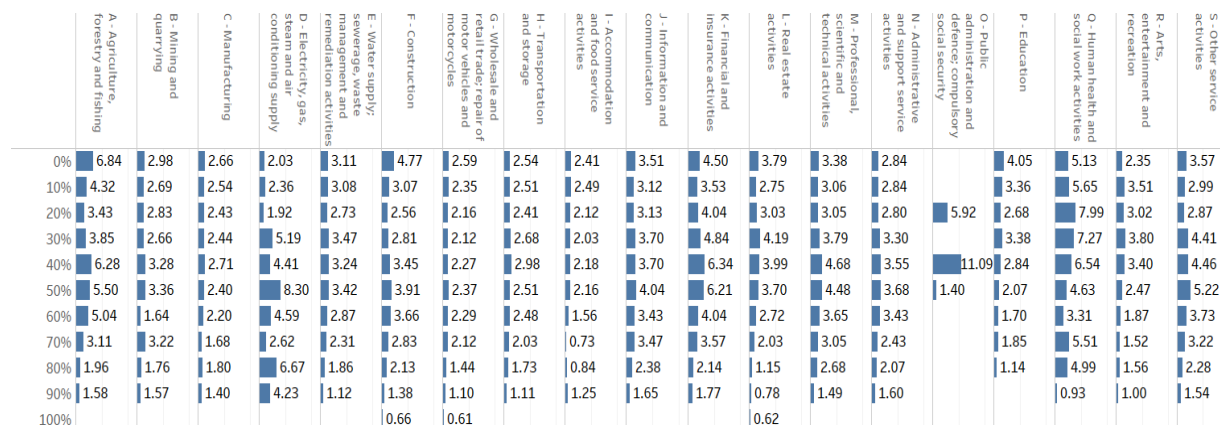


FIGURE 14
THE AVERAGE LIQUIDITY OF COMPANIES FOR EACH NACE CODE BASED ON THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL

Probability of Bankruptcy Prediction on Credit Rating, CDS and Stock Markets

Our prediction model based on financial statement data can measure the financial health of a company, which serves as a good indicator of its future performance or reliability. Analysing financial health is important since it affects various derivatives markets. In particular, CDS have become increasingly liquid and growing in recent years since they are more liquid and transparent than other credit derivatives. Specifically, CDS are an insurance that compensates

buyers in the event of default; therefore, they are closely related to the risk of firm bankruptcy and an active area of research.

Jorion and Zhang (2007) analyse the intra-industry transfer effect of Chapter 11 and Chapter 7 bankruptcies by measuring the response of CDS spreads and stock prices. Further, several studies analyse the impact of credit rating and CDS spreads. Earlier studies indicated that CDS spreads react significantly to credit rating downgrades but not upgrades (Norden and Weber, 2004; Hull, 2004). More recent studies find that CDS move a few months before negative rating events (Finnerty, 2013; Norden, 2017). Additionally, many find a relationship between CDS and the stock market (Acharya and Johnson, 2007; Han, 2017), confirming the predictive power of CDS spreads for future stock markets.

In this section, given the increasing evidence of the importance of CDS analysis for forecasting the future activities of derivatives markets, we investigate the extent to which the probability of bankruptcy influences CDS in our model. In particular, we examine whether the proposed model can forecast CDS, which are considered to measure the liability of a company. Hence, we examine both the explanatory and the predictive power of the probability of bankruptcy of our model on CDS. We also analyse the influence of credit rating and stock markets with our prediction. Therefore, we estimate the following regression model:

$$CDS_i = \beta_0 + \beta_1 Prediction_i + \beta_2 Rating_i + \beta_3 Stock_i + \epsilon_i \quad (3)$$

where *Prediction*, *Rating* and *Stock* indicate the probability of the bankruptcy of a company, credit rating and stock value, respectively. Our analysis is based on annual values for simplicity. To analyse the explanatory power of our model, we show that the prediction probability in year Y reflects the CDS in the same year. We also examine the predictive power of our year Y prediction by estimating equation (2) with the following year of CDS.

Our analysis is based on the annual financial statements of listed Japanese companies from 2010 to 2014. We collect monthly CDS and stock values from Bloomberg and Orbis in the same period. Credit rating is also collected from CREDIT Surfer (but is only available from 2011 to 2014). Table 6 reports the results of the regressions. Our results show that the probability of bankruptcy is significantly and positively related to CDS spreads. If the probability of bankruptcy increases, CDS also increase. Adding credit rating also shows the significant effect on CDS: An increase in the probability of bankruptcy and a decrease in credit rating increase the CDS value. However, stock value shows no significant effect on CDS, consistent with the findings in the literature (Acharya and Johnson, 2007; Han, 2017), which claim that the spill over effects runs from CDS to the stock market rather than from the stock market to CDS. Finally, we conduct similar experiments based on the probability of bankruptcy with a logistic model. As shown in Table 7, we find that credit rating has significantly more influence than the prediction of the logistic model.

Table 6 INFLUENCE OF THE PROBABILITY OF BANKRUPTCY OF THE RANDOM FORESTS MODEL ON CDS					
	$CDS_i = \beta_0 + \beta_1 Prediction_i + \beta_2 Rating_i + \beta_3 Stock_i + \epsilon_i$				
	Estimate	Std. Error	t value	Adj. R2	#data
Intercept	54.12	6.55	8.26***	0.06	725
Prediction	182.03	25.49	7.14***		

Intercept	217.90	18.17	12.00***	0.23	428
Prediction	160.33	28.08	5.71***		
Rating	-18.90	1.98	-9.55***		
Intercept	219.80	19.00	11.57***	0.23	428
Prediction	160.54	28.11	5.71***		
Rating	-19.29	2.28	-8.46***		
Stock	0.10	0.29	0.35		
	$CDS_{nextyear,i} = \beta_0 + \beta_1 Prediction_i + \beta_2 Rating_i + \beta_3 Stock_i + \varepsilon_i$				
	Estimate	Std. Error	t value	Adj. R2	#data
Intercept	44.49	6.01	7.40***	0.09	725
Prediction	202.00	23.40	8.68***		
Intercept	180.14	14.01	12.86***	0.21	428
Prediction	79.02	21.66	3.65***		
Rating	-14.85	1.53	-9.73***		
Intercept	185.69	14.63	12.69***	0.21	428
Prediction	79.63	21.64	3.68***		
Rating	-15.99	1.76	-9.11***		
Stock	0.29	0.22	1.31		
*** represents significance at the 0.001 level; ** represents significance at the 0.01 level; * represents significance at the 0.05 level					

Table 7					
INFLUENCE OF THE PROBABILITY OF BANKRUPTCY OF THE LOGISTIC MODEL ON CDS.					
	$CDS_i = \beta_0 + \beta_1 Prediction_i + \beta_2 Rating_i + \beta_3 Stock_i + \varepsilon_i$				
	Estimate	Std. Error	t value	Adj. R2	#data
Intercept	5.86	22.54	0.26	0.03	725
Prediction	216.32	53.61	4.04***		
Intercept	198.94	29.95	6.64***	0.19	428
Prediction	122.13	50.37	2.43*		
Rating	-18.59	2.08	-8.93***		
Intercept	200.27	30.06	6.66***	0.18	428
Prediction	126.297	50.94	2.48*		
Rating	-19.22	2.36	-8.14***		
Stock	0.171	0.30	0.57		

	$CDS_{nextyear.i} = \beta_0 + \beta_1 Prediction_i + \beta_2 Rating_i + \beta_3 Stock_i + \varepsilon_i$				
	Estimate	Std. Error	t value	Adj. R2	#data
Intercept	18.84	17.21	1.09	0.02	725
Prediction	132.93	40.93	3.25**		
Intercept	171.68	22.70	7.56***	0.19	428
Prediction	58.363	38.18	1.53		
Rating	-14.71	1.58	-9.32***		
Intercept	177.86	22.87	7.78***	0.19	428
Prediction	63.44	38.16	1.66.		
Rating	-16.45	1.82	-9.03***		
Stock	0.46	0.25	1.89.		
<i>Note: *** represents significance at the 0.001 level; ** represents significance at the 0.01 level; * represents significance at the 0.05 level; . represents significance at the 0.1 level</i>					

CONCLUSION

This study introduces a simple framework for assessing the vulnerability of industrial economic activities based on the financial statements of companies. We build a model to identify the bankruptcy of a company by using the random forest method. Then, based on the predicted status of companies, we analyse the vulnerability of the economic activities of different industries in Japan. Moreover, we validate the accuracy of our model by using more than 500,000 companies and find that our random forest model outperforms conventional methods such as logistic and tree models.

The probability of the bankruptcy of our random forest model can be used for vulnerability analysis and risk management by analysing financial data and/or importance variables with respect to the probability. This enables the detailed analysis of both the industrial and the regional base. The reliability of these analyses is directly connected to the accuracy of the proposed bankruptcy model.

We also analyse the influence of the probability of bankruptcy prediction on derivatives on market. We find probability of prediction of our model significantly affects CDS, which indicate that our model is good indicator to forecast future CDS.

Although the data used in this study are limited to Japan, our methodology does not limit the usage of different types of data. Our framework can easily be extended to analyse different countries and regions by using data on the target countries. Future work should therefore aim to forecast global industrial economic activity to analyse cross-country differences as well as conduct detailed analyses of the predictive effect of bankruptcy prediction by using more frequent financial data.

ENDNOTE

- Figures A.1a and A.1b in Appendix A.1 show the detailed distribution of total operating revenue and the ratio of operating revenue in the 10-percentage-point interval of the predicted probabilities of failure.
- Figures A.2a and A.2b in Appendix A.2 show the detailed distribution of the predicted probabilities of failure with a 10-percentage-point interval based on the total number of employees and ratio of the number of employees.
- Figures A.3a and A.3b in Appendix A.3 show the detailed distribution of the total number of companies and

ratio of the number of companies in a 10-percentage-point interval of the predicted probabilities of failure.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers (A) 17H00983 and 17K18564.

APPENDIX

Appendix A.1 Operating Revenue

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	235,186	2,299,754	5,033,570	2,620,345	1,674,658	1,260,725	4,106,810	293,132	136,263	46,216	
B - Mining and quarrying	1,802,930	4,482,183	1,783,129	727,166	11,726,528	880,029	5,858,545	83,107	94,032	25,477	
C - Manufacturing	547,662,501	1,530,074,150	1,007,674,172	340,051,674	406,141,589	216,934,396	52,205,718	46,650,506	2,946,453	5,068,318	
D - Electricity, gas, steam and air conditioning supply	19,614,091	31,645,466	124,975,456	9,413,994	3,623,519	62,373,360	15,981,179	67,333	54,247	1,363	
E - Water supply; sewerage, waste management and remediation activities	811,977	5,129,715	5,643,507	4,646,858	2,879,953	2,514,537	2,046,207	1,820,182	294,479	73,983	
F - Construction	40,249,689	305,673,497	282,950,752	165,577,076	104,374,469	63,130,717	45,142,823	16,207,746	8,178,593	3,496,370	1,488
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	86,134,847	632,754,703	763,040,189	563,060,056	349,058,122	326,136,986	167,625,790	75,600,207	71,186,982	1,345,642	2,037
H - Transportation and storage	33,073,122	141,742,772	126,271,710	74,891,342	62,879,742	52,052,909	16,529,536	6,122,753	3,949,230	851,817	
I - Accommodation and food service activities	855,658	5,940,421	26,961,258	21,393,258	15,036,254	11,651,696	5,323,147	1,683,912	79,580	7,821	
J - Information and communication	14,279,488	298,503,296	108,749,542	152,318,495	27,911,910	15,040,783	6,350,775	2,531,617	1,681,395	126,970	
K - Financial and insurance activities	15,749	15,933,725	13,130,600	21,832,678	47,217,892	44,285,556	42,184,856	2,043,161	504,721	130,589	
L - Real estate activities	1,271,909	40,991,395	77,769,684	45,355,174	36,961,443	34,285,523	15,687,515	6,941,696	4,722,671	1,336,527	15,287
M - Professional, scientific and technical activities	3,321,259	26,241,760	57,604,625	45,475,679	47,797,054	14,363,212	17,443,553	4,895,106	2,067,640	870,929	
N - Administrative and support service activities	23,430,217	28,148,513	40,865,919	46,721,780	47,610,586	22,949,040	9,756,454	7,889,225	2,507,261	634,562	
O - Public administration and defence; compulsory social security			178,280		27,854	9,189					
P - Education	51,092	818,655	5,165,017	1,195,310	3,023,994	4,195,426	2,433,615	487,208	338,674		
Q - Human health and social work activities	9,427,152	28,450,727	39,821,905	53,009,678	32,626,943	18,825,898	3,358,611	1,197,164	539,384	8,231	
R - Arts, entertainment and recreation	30,014	7,619,089	14,230,172	27,857,036	25,600,997	33,994,034	10,384,059	5,075,955	1,590,214	45,870	
S - Other service activities	1,491,402	12,839,579	32,211,952	22,557,562	29,057,731	20,102,640	19,836,323	4,877,158	651,665	259,731	
	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G	0G 1G 2G 3G

FIGURE A.1a

DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL BASIS ON TOTAL OPERATING REVENUE (MILLION US\$) FOR EACH NACE CODE

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	1.33%	12.99%	28.43%	14.80%	9.46%	7.12%	23.19%	1.66%	0.77%	0.26%	
B - Mining and quarrying	6.56%	16.32%	6.49%	2.65%	42.70%	3.20%	21.33%	0.30%	0.34%	0.09%	
C - Manufacturing	13.18%	36.82%	24.25%	8.18%	9.77%	5.22%	1.26%	1.12%	0.07%	0.12%	
D - Electricity, gas, steam and air conditioning supply	7.33%	11.82%	46.68%	3.52%	1.35%	23.30%	5.97%	0.03%	0.02%	0.00%	
E - Water supply; sewerage, waste management and remediation activities	3.14%	19.84%	21.62%	17.97%	11.14%	8.72%	7.91%	7.04%	1.14%	0.29%	
F - Construction	3.89%	29.53%	27.34%	16.00%	10.08%	6.10%	4.36%	1.57%	0.79%	0.34%	0.00%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	2.84%	20.84%	25.13%	18.55%	11.50%	10.74%	5.52%	2.49%	2.34%	0.04%	0.00%
H - Transportation and storage	6.38%	27.34%	24.96%	14.45%	12.13%	10.04%	3.19%	1.18%	0.76%	0.16%	
I - Accommodation and food service activities	0.96%	6.68%	30.32%	24.06%	16.91%	13.10%	5.99%	1.89%	0.09%	0.01%	
J - Information and communication	2.28%	47.57%	17.33%	24.27%	4.45%	2.40%	1.01%	0.40%	0.27%	0.02%	
K - Financial and insurance activities	0.01%	8.51%	7.01%	11.66%	25.21%	23.55%	22.53%	1.09%	0.27%	0.07%	
L - Real estate activities	0.48%	15.45%	29.31%	17.09%	13.93%	12.92%	5.91%	2.62%	1.78%	0.50%	0.01%
M - Professional, scientific and technical activities	1.51%	11.92%	26.17%	20.66%	21.72%	6.53%	7.93%	2.22%	0.94%	0.40%	
N - Administrative and support service activities	10.16%	12.21%	17.73%	20.27%	20.65%	9.96%	4.23%	3.42%	1.09%	0.28%	
O - Public administration and defence; compulsory social security			82.80%		12.94%	4.27%					
P - Education	0.29%	4.62%	29.17%	6.75%	17.08%	23.69%	13.74%	2.75%	1.91%		
Q - Human health and social work activities	5.03%	15.19%	21.26%	28.31%	17.42%	10.05%	1.79%	0.64%	0.29%	0.00%	
R - Arts, entertainment and recreation	0.02%	6.03%	11.26%	22.03%	20.25%	26.89%	8.21%	4.01%	1.26%	0.04%	
S - Other service activities	1.04%	8.92%	22.39%	15.68%	20.20%	13.97%	13.79%	3.39%	0.45%	0.18%	
	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%

FIGURE A.1b

DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL BASIS ON THE RATIO OF OPERATING REVENUE FOR EACH NACE CODE

Appendix A.2 Number of Employees

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	944	7,597	13,949	10,647	8,488	6,121	5,379	1,396	790	236	
B - Mining and quarrying	2,484	5,677	3,984	2,630	4,905	1,354	1,244	256	293	80	
C - Manufacturing	1,440,441	4,965,234	3,224,911	928,916	912,831	321,156	143,583	58,935	13,557	2,363	
D - Electricity, gas, steam and air conditioning supply	17,425	42,878	136,626	10,628	2,179	47,161	21,391	54	177	11	
E - Water supply; sewerage, waste management and remediation activities	3,490	23,367	29,383	20,048	14,026	14,656	10,288	4,179	1,741	416	
F - Construction	150,395	802,023	790,261	518,526	447,390	254,134	134,545	65,197	33,374	12,549	11
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	198,528	879,338	1,022,803	655,172	389,466	377,502	109,414	35,608	11,875	1,988	6
H - Transportation and storage	113,334	445,387	631,748	262,087	180,596	314,407	74,525	38,928	21,373	6,748	
I - Accommodation and food service activities	1,638	22,505	88,152	75,351	59,920	30,269	13,345	8,079	1,264	68	
J - Information and communication	50,754	666,750	374,752	261,135	115,687	52,576	30,068	16,659	8,837	974	
K - Financial and insurance activities	75	28,385	29,672	32,604	78,135	56,489	9,013	1,207	514	54	
L - Real estate activities	2,705	68,329	176,625	89,123	79,044	65,667	23,854	8,758	6,764	1,544	11
M - Professional, scientific and technical activities	14,575	117,972	193,818	164,575	157,184	65,194	44,996	24,449	12,739	4,781	
N - Administrative and support service activities	102,864	138,561	181,960	135,292	110,215	85,613	65,737	38,680	18,638	2,941	
O - Public administration and defence; compulsory social security			944		89	96					
P - Education	475	4,326	30,898	8,440	18,989	16,293	10,860	4,627	1,781		
Q - Human health and social work activities	80,719	257,886	342,953	341,884	230,530	147,055	30,001	11,710	5,131	90	
R - Arts, entertainment and recreation	121	11,808	25,524	25,530	34,267	17,479	7,786	2,684	1,008	36	
S - Other service activities	5,521	28,596	50,185	56,443	66,910	58,823	22,791	5,798	1,659	536	
	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M	0M 5M 10M

FIGURE A.2a
DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL BASIS ON THE TOTAL NUMBER OF EMPLOYEES FOR EACH NACE CODE

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	1.70%	13.68%	25.11%	19.17%	15.28%	11.02%	9.68%	2.51%	1.42%	0.42%	
B - Mining and quarrying	10.84%	24.78%	17.39%	11.48%	5.91%	5.43%	1.12%	1.28%	0.35%		
C - Manufacturing	11.99%	41.34%	26.85%	7.73%	7.60%	2.67%	1.20%	0.49%	0.11%	0.02%	
D - Electricity, gas, steam and air conditioning supply	6.26%	15.39%	49.05%	3.82%	0.78%	16.93%	7.68%	0.02%	0.06%	0.00%	
E - Water supply; sewerage, waste management and remediation activities	2.87%	19.22%	24.16%	16.49%	11.54%	12.05%	8.46%	3.44%	1.43%	0.34%	
F - Construction	4.69%	25.00%	24.63%	16.16%	13.94%	7.92%	4.19%	2.03%	1.04%	0.39%	0.00%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	5.39%	23.88%	27.78%	17.80%	10.58%	10.25%	2.97%	0.97%	0.32%	0.05%	0.00%
H - Transportation and storage	5.42%	21.32%	30.24%	12.55%	8.64%	15.05%	3.57%	1.86%	1.02%	0.32%	
I - Accommodation and food service activities	0.54%	7.49%	29.33%	25.07%	19.93%	10.07%	4.44%	2.69%	0.42%	0.02%	
J - Information and communication	3.22%	42.25%	23.75%	16.55%	7.33%	3.33%	1.91%	1.06%	0.56%	0.06%	
K - Financial and insurance activities	0.03%	12.02%	12.57%	13.81%	33.09%	23.92%	3.82%	0.51%	0.22%	0.02%	
L - Real estate activities	0.52%	13.08%	33.81%	17.06%	15.13%	12.57%	4.57%	1.68%	1.29%	0.30%	0.00%
M - Professional, scientific and technical activities	1.82%	14.74%	24.22%	20.56%	19.64%	8.15%	5.62%	3.06%	1.59%	0.60%	
N - Administrative and support service activities	11.68%	15.74%	20.67%	15.37%	12.52%	9.72%	7.47%	4.39%	2.12%	0.33%	
O - Public administration and defence; compulsory social security			83.61%		7.88%	8.50%					
P - Education	0.49%	4.47%	31.96%	8.73%	19.64%	16.85%	11.23%	4.79%	1.84%		
Q - Human health and social work activities	5.57%	17.81%	23.69%	23.61%	15.92%	10.16%	2.07%	0.81%	0.35%	0.01%	
R - Arts, entertainment and recreation	0.10%	9.35%	20.22%	20.22%	27.14%	13.85%	5.17%	2.13%	0.80%	0.03%	
S - Other service activities	1.86%	9.62%	16.88%	18.99%	22.51%	19.79%	7.67%	1.95%	0.56%	0.18%	
	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%	0% 50%

Figure A.2b
DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-PERCENTAGE-POINT INTERVAL BASIS ON THE RATIO OF THE NUMBER OF EMPLOYEES FOR EACH NACE CODE

Appendix A.3 Number of Companies

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	117	503	685	802	1,065	847	415	205	101	24	
B - Mining and quarrying	12	72	126	89	91	96	37	26	19	10	
C - Manufacturing	2,058	9,797	11,064	8,118	6,843	4,851	2,233	905	398	98	
D - Electricity, gas, steam and air conditioning supply	16	68	83	62	104	119	69	8	4	1	
E - Water supply; sewerage, waste management and remediation activities	142	580	778	765	669	542	291	136	83	20	
F - Construction	18,374	53,275	58,577	52,811	57,684	42,704	20,943	9,622	4,728	1,486	2
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	2,071	11,009	15,334	12,829	11,594	7,380	3,245	1,136	441	142	1
H - Transportation and storage	198	1,180	1,994	2,130	2,012	1,577	914	601	441	102	
I - Accommodation and food service activities	18	266	807	991	1,024	613	206	37	7	2	
J - Information and communication	244	1,421	2,234	1,982	2,103	1,500	627	274	131	23	
K - Financial and insurance activities	8	58	255	569	897	823	218	38	8	2	
L - Real estate activities	62	637	2,584	6,499	11,883	9,501	3,141	1,158	505	232	3
M - Professional, scientific and technical activities	475	2,225	3,387	3,954	4,984	4,227	2,056	925	489	151	
N - Administrative and support service activities	189	1,104	1,612	1,704	1,883	1,494	777	433	205	50	
O - Public administration and defence; compulsory social security			1		1	1					
P - Education	2	32	78	136	205	168	42	7	5		
Q - Human health and social work activities	294	1,596	5,442	16,654	9,108	3,928	1,276	237	57	1	
R - Arts, entertainment and recreation	6	91	230	393	589	423	158	49	15	1	
S - Other service activities	46	438	803	1,077	1,420	1,152	461	155	69	20	
	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K	OK 50K 100K

FIGURE A.3a
**DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-
 PERCENTAGE-POINT INTERVAL BASIS ON THE TOTAL NUMBER OF
 COMPANIES FOR EACH NACE CODE**

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A - Agriculture, forestry and fishing	2.46%	10.56%	14.38%	16.83%	22.36%	17.78%	8.71%	4.30%	2.12%	0.50%	
B - Mining and quarrying	2.08%	12.46%	21.80%	15.40%	15.74%	16.61%	6.40%	4.50%	3.29%	1.73%	
C - Manufacturing	4.44%	21.13%	23.86%	17.51%	14.76%	10.46%	4.82%	1.95%	0.86%	0.21%	
D - Electricity, gas, steam and air conditioning supply	3.00%	12.73%	15.54%	11.61%	19.48%	22.28%	12.92%	1.50%	0.75%	0.19%	
E - Water supply; sewerage, waste management and remediation activities	3.54%	14.48%	19.42%	19.10%	16.70%	13.53%	7.26%	3.39%	2.07%	0.50%	
F - Construction	5.74%	16.64%	18.29%	16.49%	18.01%	13.34%	6.54%	3.00%	1.48%	0.46%	0.00%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	3.18%	16.89%	23.52%	19.68%	17.79%	11.32%	4.98%	1.74%	0.68%	0.22%	0.00%
H - Transportation and storage	1.78%	10.58%	17.89%	19.10%	18.05%	14.14%	8.20%	5.39%	3.96%	0.91%	
I - Accommodation and food service activities	0.45%	6.70%	20.32%	24.96%	25.79%	15.44%	5.19%	0.93%	0.18%	0.05%	
J - Information and communication	2.32%	13.49%	21.20%	18.81%	19.95%	14.23%	5.95%	2.60%	1.24%	0.22%	
K - Financial and insurance activities	0.28%	2.02%	8.87%	19.78%	31.19%	28.62%	7.58%	1.32%	0.28%	0.07%	
L - Real estate activities	0.17%	1.76%	7.14%	17.95%	32.82%	26.24%	8.68%	3.20%	1.39%	0.64%	0.01%
M - Professional, scientific and technical activities	2.08%	9.73%	14.81%	17.29%	21.79%	18.48%	8.99%	4.04%	2.14%	0.66%	
N - Administrative and support service activities	2.00%	11.68%	17.06%	18.03%	19.92%	15.81%	8.22%	4.58%	2.17%	0.53%	
O - Public administration and defence; compulsory social security			33.33%		33.33%	33.33%					
P - Education	0.30%	4.74%	11.56%	20.15%	30.37%	24.89%	6.22%	1.04%	0.74%		
Q - Human health and social work activities	0.76%	4.14%	14.10%	43.15%	23.60%	10.18%	3.31%	0.61%	0.15%	0.00%	
R - Arts, entertainment and recreation	0.31%	4.65%	11.76%	20.10%	30.13%	21.64%	8.08%	2.51%	0.77%	0.05%	
S - Other service activities	0.82%	7.76%	14.24%	19.09%	25.17%	20.42%	8.17%	2.75%	1.22%	0.35%	
	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%	0% 20% 40%

FIGURE A.3b
**DISTRIBUTION OF THE PREDICTED PROBABILITIES OF FAILURE WITH A 10-
 PERCENTAGE-POINT INTERVAL BASIS ON THE RATIO OF THE NUMBER OF
 COMPANIES FOR EACH NACE CODE**

REFERENCES

- Acharya, V. & Johnson, T. (2007). Insider trading in credit derivatives. *Journal of Financial Economics*, 84, 110-141.
- Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589-609.
- Beaver, W.H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
- Brédart, X. (2015). Bankruptcy prediction model: The case of the United States. *International Journal of Economics and Finance*, 6, 1-7.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and Regression Trees*. Monterey, California: Wadsworth.
- Finnerty, J., Miller, C. & Chen, R. (2013). The impact of credit rating announcements on credit default swap spreads. *Journal of Banking & Finance*, 37, 2011-2030.
- Frankel, J. & Saravelos, G. (2012). Can leading indicators assess country vulnerability? Evidence from the 2008-2009 global financial crises. *Journal of International Economics*, 87, 216-231.
- Han, B., Subrahmanyam, A. & Zhou, Y. (2017). The term structure of credit spreads, firm fundamentals and expected stock returns. *Journal of Financial Economics*, 124, 147-171.
- Hillegeist, S.A., Keating, E.K., Cram, D.P. & Lundstedt, K.G. (2004). Assessing the probabilities of bankruptcy prediction. *Review of Accounting Studies*, 9, 5-34.
- Hull, J., Predescu, M. & White, A. (2004). The relationship between credit default swap spreads, bond yields and credit rating announcements. *Journal of Banking & Finance*, 28, 2789-2811.
- Jorion, P. & Zhang, G. (2007). Good and bad credit contagion: Evidence from credit default swaps *Journal of Financial Economics*, 84, 860-883.
- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modelling*. New York: Springer.
- Lennox, C. (1999). Identifying failing companies: A re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51, 347-364.
- Messier, Jr.W. & Hansen, J.V. (1988). Inducing rules for expert system development: An example using default and bankruptcy data. *Management Science*, 34, 1403-1415.
- Norden, L. & Weber, M. (2004). Informational efficiency of credit default swap and stock markets: The impact of credit rating announcements. *Journal of Banking & Finance*, 28, 2813-2843.
- Norden, L. (2017). Information in CDS spreads. *Journal of Banking & Finance*, 75, 118-135.
- Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- Shirata, C.Y. (1999). *Predictable Information of Bankruptcy in Japan*. Tokyo: Chuokeizai-Sha.
- Shirata, C.Y. (2003). Predictors of bankruptcy after bubble economy in Japan: What can you learn from the Japan case? *Proceedings of the 15th Asian-Pacific Conference on International Accounting Issues*.
- Tanaka, K., Kinkyo, T. & Hamori, S. (2016). Random forests-based early warning system for bank failures. *Economics Letters*, 148, 118-121.
- Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28, 3-28.

This article was originally published in a special issue, entitled: "Management Information, Decision Sciences, and Cognate Disciplines", Edited by Chia-Lin Chang, Michael McAleer & Wing-Keung Wong.