

Credit Scoring Model for Business Start-ups

Kenzo Ogi†

Risk Management Department
Japan Finance Corporation, Tokyo, Japan
Tel: (+81) 3-3270-1408, Email: ogi-k@jfc.go.jp
&
Graduate School of Science and Technology
Keio University, Yokohama, Japan
Tel: (+81) 45-566-1635, Email: ogi-k@keio.jp

Yuichi Utsumi

Risk Management Department
Japan Finance Corporation, Tokyo, Japan
Tel: (+81) 3-3270-1408, Email: utsumi-y@jfc.go.jp

Norio Hibiki

Faculty of Science and Technology
Keio University, Yokohama, Japan
Tel: (+81) 45-566-1635, Email: hibiki@ae.keio.ac.jp

Abstract. The Japanese government is urging Japanese banks to promote loans for business start-ups. The difficulty of managing their credit risk becomes a more serious because of the increase in loans for business start-ups. Most Japanese banks utilize credit scoring models linked to the correlations between financial indicators and default occurrence in order to evaluate their debtor's credit risk. However, we cannot use existing models for evaluating business start-ups because of the lack of financial statements. To the best of our knowledge, there are no models that can measure their credit risk. In this paper, we analyze the impact on the default occurrence of non-financial variables grouped into three categories, or financial, human and industry factors, by using the data set of 34,470 Japanese business start-ups. Then, we construct a logistic regression model of credit scoring for business start-ups and test the robustness from a practical perspective. As a result of analyses, we confirm that some explanatory variables of three categories are statistically significant, and that the accuracy ratio (AR) of our model is about 57%. We believe that our research will help Japanese banks to make practical credit scoring models for business start-ups aiming sound banking.

Keywords: credit scoring model, credit risk, business start-ups, default, logistic regression model

1. INTRODUCTION

Since the latter half of the 1980's, the number of enterprises in Japan has been decreasing. The Japanese government is urging Japanese banks to promote loans for business start-ups (consisting of those that have not yet started and those that are within a year from the start-up). The difficulty of managing their credit risk becomes a more serious because of the increase in loans for business start-ups. It is an important problem not only for banks but also for the Micro Business and Individual Unit of Japan Finance Corporation (JFC-Micro), a policy-based financial institution that aims to contribute to the promotion of small sized firms and solo proprietors. Most Japanese banks utilize credit scoring models linked to the correlations

between financial ratios and default in order to evaluate their debtor's credit risk. However, we cannot use existing models for evaluating business start-ups, because they do not have financial statements at the beginning of their business. To the best of our knowledge, there are no studies concerning credit scoring models that can measure their credit risk.

There are some previous studies that analyzed some factors of default or business-closing in business start-ups. Gonçalves et al. (2014) revealed some determinants of default on business start-ups by using a panel data of 1,430 Portuguese business start-ups. They tested the impact of default by using variables grouped into three categories which are financial capital, human capital and industry dynamics. From a financial point of view, they concluded

Table1: Comparison with previous studies

		Gonçalves et al.(2014)	Suzuki(2012)	Our study
Credit scoring model		×	×	○
Firms		After business start up	After business start up	Before business start up
Explanatory variables		Financial and Non-financial	Financial and Non-financial	Non-financial
Evaluation factors	Human factor	○	○	○
	Financial factor	○	×	○
	Industrial factor	○	○	○
Number of data	maximum	1,430	2,897	34,470
	minimum		783	1,718
Years of making loan		2005-2006	2006	2011-2013

that the supports provided by partners in the financing activity, such as the intensity of use of assets management and reduced debt pay-back period, surely reduce the risk of default. In addition, they found that the default rate would be affected by the quality of human capital such as educational background and management experience. On the other hand, variables of industry dynamics such as industry growth, industry's entry rate/barriers to entry and market concentration, are not statistically significant at the 5% level.

Suzuki (2012) clarified the reasons for business-closing by analyzing a panel data of 2,897 Japanese business start-ups for the periods from 2006 to 2010. He investigated factors of business-closing of business start-ups by using variables grouped into three categories; founder's ability, business environment and business strategy. As a result, he pointed out the fact that the business-closing rate is affected by founder's background, particularly for his/her previous work experience such as management experience and experience in the same industry. Furthermore, he showed that the business environmental factors such as the number of staffs at the beginning and the business-closing rate of the same industry, and that business strategy factors such as novelty of business, are statistically significant at the 5% level. However, these previous studies did not make credit scoring models. Moreover, their analyses have used financial variables obtained after starting business. In this paper, we use only non-financial variables that are obtained before starting business because enterprises before business starting up do not have financial statements. We categorize these variables into three types; financial, human and industrial factors. We analyze their impact on the default by using the data set of 34,470 Japanese business start-ups, owned by JFC-Micro. Then, we construct a logistic regression model of credit scoring for business start-ups and examine the robustness of the model from a practical perspective.

Table 1 shows the difference between our study and the previous ones. The feature of our study is the fact that

we use only non-financial variables for the enterprises before business starting up, and analyze more number of data than those of previous studies.

As a result, we confirm that some explanatory variables of three factors are statistically significant, and that the accuracy ratio (AR) of our model is about 57%. Here, the AR is commonly used as an evaluation measure. Yamashita and Miura (2011) show that the AR of the logistic regression model is between 60% and 70% for medium or large sized firms. Hibiki et al. (2010) state that it is between 35% and 45% for small sized firms. Therefore, it is expected that our research also helps Japanese banks to make practical use of credit scoring models for business start-ups on sound banking.

The rest of this paper is organized as follows. Section 2 explains the framework of the analysis. Section 3 shows the results of logit analyses for finding statistically significant explanatory variables, and empirically examines the robustness. We state the final remarks in Section 4.

2. METHODOLOGY AND DATA EMPIRICAL RESULTS

The purposes of our study are to construct a credit scoring model for business start-ups by using the non-financial data set of 34,470 Japanese business start-ups which JFC provided loans from 2011 to 2013, and to examine the possibilities of using the model from a practical perspective.

There are various types of credit scoring models. It is common to utilize the logistic regression model (logit model) in order to calculate the credit scores, linked to the correlations between financial ratios and default occurrence rate (DR). We use a binominal logit model which explains the probability of default (PD) by estimating a linear equation of the natural logarithm of the odds ratio of two possible events (default/non-default).

The definition of DR used is crucial for the results, and therefore studies based on different concepts of default cannot be easily compared. In this paper, the

Table2: Main candidates of explanatory variables

Category	Main candidates of explanatory variables
Industrial factor	Default rate by industrial classification Number of business start-ups by industrial classification etc
Financial factor	Cash on hand of founder* Real-estate holding(dummy)* Annual income of the entrepreneur before starting business* Amount of start-up capital* Amount of debt of the founder Repayment performance etc
Human factor	Work experience in the same industry(number of years)* Career before business start-ups(number of years)* Management experience(number of years)* Schooling(Education level)* Franchise store* Frequency of career change* Validity of business plan(dummy)* Age of founder etc

Table 3: Data

	Variable categories	Years of making loan	Years of default	Number of data
DB1	Industrial factor	2011-2013	2012-2014	34,470
	Financial factor(unmarked)			
	Humanl factor(unmarked)			
DB2	Financial factor(marked)	2011-2012	2012-2013	1,718
	Humanl factor(marked)			

concept of default corresponds to the credit event that overdue for more than 90 days.

Also, DR at time t is calculated as follows:

$$DR(t) = \frac{D(t)}{ND(t) + D(t)}, \quad (1)$$

where $ND(t)$ is the number of the non-default firms at time t , and $D(t)$ is the number of the default firms at time t .

It is important to select adequate variables for the model construction. The quality of explanatory variables affects the adequacy of the model, and the performance can be improved greatly by incorporating only one effective variable. Ogi et al.(2015) conclude that their model increased 7% AR-value by adding the firm age variable. We categorize variables into three types and analyze them. Main variable candidates in each category are as follows.

- i) Human factor: management experience, work experience in the same industry, validity of business plan, and so on.
- ii) Financial factor: private assets and debts of the founder, amount of total fund on their business, annual income of the founder before starting business, and so on.
- iii) Industrial factor: default rate of each industry, the number of business start-ups in each industry, and so on.

2.1 Data set

Table 2 shows the main variables of each category used for the analysis. The marked variables are not prepared for every business start-ups on the data set. Table3 shows the data set for the analysis. DB1 is the data set of 34,470 Japanese business start-ups which JFC provided loans from 2011 to 2013. DB2 is the data set of the marked variables of 1,718 Japanese business start-ups randomly selected from all data.

2.2 Analytical procedure

As mentioned above, two kinds of the data sets (DB1/DB2) provide different explanatory variable candidates. Therefore, we construct three models as in Table 4. First, we construct each logistic regression model for three categories. We combine variables of each factor which are statistically significant in each analysis, and estimate parameters of the model I. Second, we construct model II, using variables concerning human factor and financial factor in the DB2.

Table 4: Models

Type	Data set	Variable categories	Goal
Model I	DB1	Industrial factor Financial factor(unmarked) Human factor(unmarked)	Selection of variables(unmarked)
Model II	DB2	Financial factor(marked) Human factor(marked)	Selection of variables (marked)
Model III	DB2	Credit scores of Model I Credit scores of Model II	Construction of model for practical use

Finally, we construct model III by combining the credit scores of model I and model II.

Each model is described in the following, respectively.

(Model I)

PD of business start-ups is calculated by the following regression equation:

$$p_{1,i} = \frac{1}{1 + e^{z_{1,i}}}$$

$$z_{1,i} = \alpha_{1,0} + \sum_{j=1}^J \beta_{1,j} h_{1,i,j} + \sum_{k=1}^K \gamma_{1,k} f_{1,i,k} + \sum_{m=1}^M \delta_m g_{i,m}, \quad (2.1)$$

where $h_{1,i,j}(i = 1, \dots, N_1; j = 1, \dots, J)$ is a variable of business start-ups i with respect to human factor j on DB1, $f_{1,i,k}(i = 1, \dots, N_1; k = 1, \dots, K)$ is a variable of business start-ups i with respect to financial factor k , and $g_{i,m}(i = 1, \dots, N_1; m = 1, \dots, M)$ is a variable of business start-ups i with respect to industry factor m . $\alpha_{1,0}$ is a constant term of Model I. We estimate the parameters, $\beta_{1,j}(j = 1, \dots, J)$, $\gamma_{1,k}(k = 1, \dots, K)$, and $\delta_g(m = 1, \dots, M)$ by employing a maximum likelihood method. Then, $p_{u,i}$ is a PD of business start-ups i for model u , where N_1 is the number of business start-ups in the DB1. The lower PD becomes, the larger $z_{u,i}$ is.

(Model II)

$$p_{2,i} = \frac{1}{1 + e^{z_{2,i}}}$$

$$z_{2,i} = \alpha_{2,0} + \sum_{x=1}^X \beta_{2,x} h_{2,i,x} + \sum_{y=1}^Y \gamma_{2,y} f_{2,i,y} \quad (2.2)$$

where $h_{2,i,x}(i = 1, \dots, N_2; x = 1, \dots, X)$ is a variable of business start-ups i with respect to human factor x on DB2, $f_{2,i,y}(i = 1, \dots, N_2; y = 1, \dots, Y)$ is a variable variables of business start-ups i with respect to financial factor y . We estimate the parameters, $\beta_{2,x}(x = 1, \dots, X)$ and $\gamma_{2,y}(y = 1, \dots, Y)$, by employing a maximum likelihood method, where N_2 is the number of business start-ups in the DB2.

(Model III)

$$p_{3,i} = \frac{1}{1 + e^{z_{3,i}}}, \quad z_{3,i} = A + Bz_{1,i} + Cz_{2,i} \quad (2.3)$$

We estimate the parameters B and C , by combining $z_{1,i}$ and $z_{2,i}$.

3. RESULTS

In this section, we discuss the results of estimation of the credit scoring models (presented in Table 4, Models I, II and III).

3.1 Construction of Model I

Constructing procedure of model I is as follows.

- i) Concerning industrial factor, we divide all business start-ups of DB1 into some industry groups, based on the number of firms and default rate of respective industry. We select candidates of explanation variables by each industry group.
- ii) About financial factor, we utilize logistic regression method in order to choose candidates of explanation variables with respect to assets and debts of the founder especially.
- iii) Regarding human factor, we attempt to find candidates of significant variables with respect to the age of founder.

Table 5: Estimation results with respect to industrial factor

Business group	Standardized regression coefficient	p-value	Default rate
Other Services	▲ 0.08	<0.001	High
Professional and Technical services I	0.25	0.961	Low
Professional and Technical services II	▲ 0.02	0.304	High
Professional and Technical services III	0.09	0.041	Low
Services to the people's daily lives	0.16	<0.001	Low
Healthcare/Social assistance I	0.16	<0.001	Low
Healthcare/Social assistance II	0.00	0.888	Low
Healthcare/Social assistance III	▲ 0.02	0.423	High
Food services/ Accommodation	▲ 0.20	<0.001	High
Transportation and Warehousing I	▲ 0.03	0.082	High
Transportation and Warehousing II	0.03	0.402	Low
Wholesale trade/Retail trade	▲ 0.13	<0.001	High
Educational services	▲ 0.03	0.181	High
Construction	▲ 0.03	0.188	High
Information	▲ 0.01	0.488	High
Real estate	0.19	<0.001	Low

Table 6: Entry rate and value-added to sales ratio of each business group

Business group	Default rate	Entry rate to the business group (%)	Added-value rate (%)
Food services/ Accommodation	High	4.5	44.0
Other Services	High	2.2	30.6
Wholesale trade/Retail trade	High	2.0	10.4
Services to the people's daily lives	Low	1.8	55.5
Professional and Technical services III	Low	2.4	52.6
Healthcare/Social assistance I	Low	3.1	51.7
Real estate	Low	0.5	31.2
Average	—	2.4	39.4

iv) We construct model I consisting of selected variables, and calculate the AR.

3.1.1 Selection of explanatory variables concerning industry factor

Suzuki (2012) showed that business-closing rates of Food services/Accommodation, Retail trade and Information are high, whereas those of Services for individuals, Traffic and Healthcare/Social assistance are low. It suggests that default rates by industrial classification have potential for effective explanation variables in Japanese business start-ups.

We classify DB1 into 87 business groups, according to the degree of similarity in order to select effective industry variables. Next, we bring together 87 groups into 16 groups in such a way that the number of firms in each business classification becomes a hundred or more. Then,

we examine the results of 16 business groups by logit regression in Table 5. Seven business groups (bold) among 16 groups are statistically significant at the 5% level. The default rates of Food services/Accommodation, Wholesale trade/Retail trade and Other services are high, whereas those of Services to the people's daily lives (for example, beauty salon business, laundry business and so on), Professional and Technical services (for example, legal services, accounting and bookkeeping services and so on), Real estate and Healthcare/Social assistance are low.

In addition, we examine the difference on the default rate in each business classification by using business entry rate in the same industry (ratio of the number of business start-ups to the number of all firms) and value-added to sales ratio.

Table 6 shows the results of entry rate and value-added to sales ratio. We find that business groups of low default rate tend to show high value-added to sales ratio

Table 7: Results concerning financial factor

Variable	Estimates	Standardized regression coefficient	p-value
Constant term	3.94	—	<0.001
Asset status of founder (I)	0.56	0.22	<0.001
Debt status of founder (I)	▲0.29	▲0.07	0.105
Debt status of founder (II)	▲0.14	▲0.15	<0.001
Debt status of founder (III)	▲0.63	▲0.07	<0.001
Debt status of founder (IV)	▲0.11	▲0.06	<0.001
Debt status of founder (V)	▲0.52	▲0.16	<0.001

and low industry's entry rate. It is expected that those businesses are not in severely competitive. On the other hand, business groups of high default rate tend to have low value-added to sales ratio and high industry entry rate. Those businesses seem to be very competitive. Moreover, when we construct the model with the variables of industry factor, it gives the AR of 35.4%.

3.1.2 Selection of explanatory variables concerning financial factor

Ogi et al.(2014) described that assets of founder is one of the most important factors in credit risk evaluation for small sized firms. Gonçalves et al.(2014) verified that the support provided by partners in the financing of the company's activity was decisive in mitigating risk of default. Referring to these papers, we ran logit regression analysis using variables with respect to private assets and debts of founder.

First, we construct a simple logit regression model of each variable among 85 variables concerning financial factor, and choose statistically significant variables at the 5% level. Second, we select variables under the condition that the coefficient of correlation between any two variables is not beyond 0.5, and therefore reduce the number of the candidates of explanatory variables to 15.

Table 7 shows six variables chosen by the stepwise selection. Five variables represent the debt status of founder. The score becomes lower when each of five variables is higher due to negative coefficients. Also, a variable represents founder assets information. The score becomes higher when the variable is higher because of a positive coefficient. In addition, when we construct the model with variables of only financial factor, it gives the AR of 38.7%.

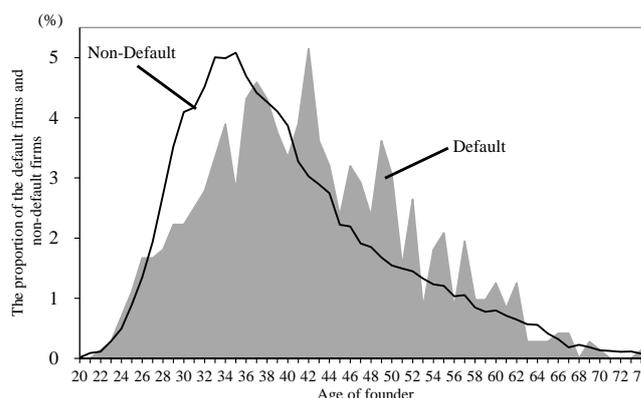


Figure 1: Proportions of default/non-default firms by age of founder

3.1.3 Select explanatory variables concerning human factor

An explanatory variable of human factor on DB1 is only the age of founder. However, it is expected that the age of founder have a great impact on the performance of business start-ups. Genda and Takahashi (2003) revealed that the performance level of business start-ups is the highest when the age of the founder is around 42 years old. Suzuki (2012) also reported that intellectual and physical strength was decreasing with advancing age, and had an impact on the business activity.

Figure 1 indicates the proportion of default and non-default firms in each age of founder. We point out the fact that the proportion of default firms exceeds that of non-default firms around 40 years old, and it is consistent with the previous studies. Based on the result, we choose a dummy variable whether the age of the founder is less than 40 or not.

3.1.4 Evaluating model I

We construct model I by using the explanatory variables selected in the section 3.1.1, 3.1.2 and 3.1.3. Table 8 shows the results. All variables except for a dummy variable of Professional and Technical services are chosen by stepwise selection, and the AR is 51.2%. We confirm that it is higher than the AR of about 45% by our credit scoring model for small sized firms. The result gives us the possibility that we can utilize model I in practice.

3.2 Constructing Model II

In this section, we construct model II by using ten variables of DB2, or financial and human factors with asterisks in Table 2. Table 9 shows the result of statistically

Table 8: Results of Model I

Category	Variables	Estimates	Standardized regression coefficient	p-value
	Constant term	3.91	—	<0.001
Industry factor	Food services/ Accommodation	▲ 0.82	▲ 0.20	<0.001
	Other services	▲ 0.71	▲ 0.06	<0.001
	Wholesale trade/Retail trade	▲ 0.57	▲ 0.11	<0.001
	Services to the people's daily lives	0.65	0.13	<0.001
	Healthcare/Social assistance	0.84	0.14	<0.001
	Real estate	0.14	0.18	<0.001
Financial factor	Asset status of founder (I)	0.53	0.20	<0.001
	Debt status of founder (I)	▲ 0.62	▲ 0.07	<0.001
	Debt status of founder (II)	▲ 0.54	▲ 0.16	<0.001
	Debt status of founder (III)	▲ 0.06	▲ 0.04	<0.001
	Debt status of founder (IV)	▲ 0.13	▲ 0.14	<0.001
	Debt status of founder (V)	▲ 0.12	▲ 0.06	<0.001
Human factor	Age of the founder is less than 40 or not	0.50	0.14	<0.001

Table 9: Results of Model II

Variable	Estimation	Standardized regression coefficient	p-value
Constant term	▲ 0.71	—	0.063
Cash on hand of founder	0.3	0.20	<0.001
Work experience in the same industry (dummy)	▲ 0.49	▲ 0.12	<0.001
Validity of business plan (dummy)	▲ 0.50	▲ 0.11	0.006

Table 10: AR-value by work experience in the same industry (number of years)

Dummy	Under 1 year	Under 2 years	Under 3 years	Under 4 years	Under 5 years	Under 6 years	Under 7 years	Under 8 years	Under 9 years	Under 10 years
AR-value	8.4	10.4	11.4	10.9	11.1	9.0	6.6	6.9	8.1	5.9

significant variables chosen by the stepwise selection. We confirm that two variables concerning human factor (work experience in the same industry, validity of business plan) and a variable concerning financial factor (cash on hand of founder) are statistically significant at the 5% level. On the other hand, management experience and schooling are not significant. The variable of work experience in the same industry is a dummy variable that the number of years of experience is less than 5 years or not. Suzuki (2012) found that the business-closing rate becomes lower as the number of years of experience becomes longer. We calculate the AR of each year from 1 to 10 years in order to find it as in Table 10. We find that a dummy variable to obtain the highest AR is less than three years or not. However, we select less than five years as a boundary because the AR of

less than four years and less than five years is almost equal to that of three years. The validity of business plan is a dummy variable subjectively determined by loan officer's judgment. Also, the variable of cash on hand of founder concerning financial factor is amount of money left over after starting business. These two variables are statistically significant at the 5% level. We confirm that the AR of model II is 27.9%

3.3 Constructing Model III (integrated model)

We calculate the credit score $z_{1,i}$ of model I and the credit score $z_{2,i}$ of model II by using DB2, and build model III by combining both values.

Table 11: Results of Model III

Variable	Estimation	Standardized regression coefficient	p-value
Constant term	▲ 0.12	—	<0.001
Credit score of Model I	0.98	0.59	<0.001
Credit score of Model I	0.85	0.22	<0.001

Table 11 shows the result. We find that the standardized regression coefficient of the credit score of model I is 0.59 and that of model II is 0.22.

Figure 2 indicates the proportions of the default firms and non-default firms in each value of model III. We confirm that the AR of the model III is 57.1%, which is much higher than the AR of our model for small sized firms (Ogi et al.(2015)). The result enhances the possibility that we employ model III for practice use.

4. CONCLUSION

In this paper, we analyze the impact on the default occurrence of non-financial variables grouped into three categories which are industrial, financial, and human factors, using the data set of 34,470 Japanese business start-ups. Then, we construct logistic regression models of credit scoring for business start-ups and test the robustness of the models from a practical perspective.

The first key finding of this paper is that some explanatory variables of three categories are statistically significant. The second key finding is that the AR of our integrated model is about 57%. Moreover, to the best of our knowledge, there are no studies about credit scoring model for business start-ups. Thus, we firmly believe that the results of our analyses are very unique.

In our results, we need to pay attention to two aspects. First, it is possible that these findings are weekly biased, because the data set consists of only the firms for which JFC financed. In order to improve the accuracy of our estimation, we might need to update the model using more refined data set. Second, the time-series analysis cannot be conducted in this paper because of the short span of the data. They are to be analyzed in our future research. Because of the JFC's management policy for its own practical credit scoring models, we could not clearly specify some exact name of variables.

Nevertheless, we can conclude that this study enhances the potential of credit scoring models for business start-ups for the practical use aiming sound banking.

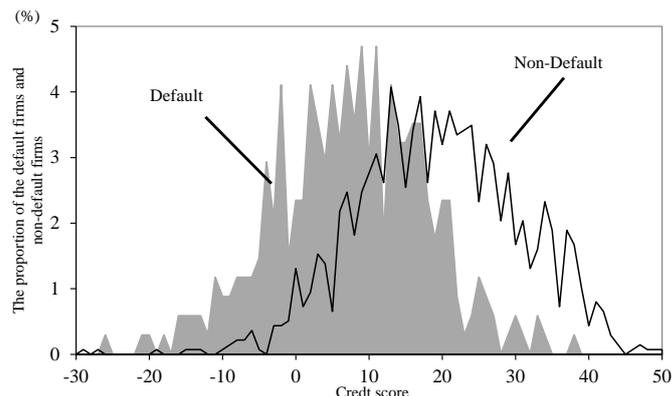


Figure 2: Proportions of default/non-default firms by credit scores of Model III

REFERENCES

- Altman, E.I. and G. Sabato (2007), “Modelling Credit Risk for SMEs: Evidence from the U.S. Market”, *ABACUS*, Vol.43, No.3, 332-357.
- Genda, Y. and Y. Takahashi (2003), “Present and Future of Self-employment”, *Chosa kiho*, No.64, National Life and Finance Corporation (present JFC) Research Institute, 1-27, (in Japanese).
- Gonçalves, V.S., Martins, F.V., and Brandao, E. (2014), “The Determinants of Credit Default on Start-Up Firms. Econometric Modeling using Financial Capital, Human Capital and Industry Dynamics Variables”, *FEP Working Papers*, n.534.
- Ogi, K., Moridaira, S. (2013) The Determinants of Macro Economic Factor that Impact on Default Rate of Small Sized Firms -1factor model approach-, *Collection of Papers in Japan Finance Corporation*, No.20, 71-87, (in Japanese).
- Ogi, K., Toshiro, M. and N. Hibiki (2014), “Effect of Firm Age in the Credit Scoring Model for Small Sized Firms”, *Proceedings of the 15th Asia Pacific Industrial Engineering & Management Systems Conference*.
- Ogi, K., M. Toshiro and N. Hibiki (2015), “Effect of Firm Age in Expected Loss Estimation for Small Sized Firms”, *Proceedings of the 16th Asia Pacific Industrial Engineering & Management Systems Conference*.
- Ogi, K., M. Toshiro and N. Hibiki (2015), “Empirical Study of Recovery Rate Model for Small Sized Firms”, *JAFEE journal*, 168-208 (in Japanese).
- Ohlson, J.A. (1980), “Financial ratios and the probabilistic prediction of bankruptcy”, *Journal of Accounting Research*, Vol.18, No.1, 109-131.
- Suzuki, M. (2012), “Trajectory of Business Start-Ups- Business Performance, Resources and Consciousness Change based on Panel Data-”, *Keiso syobo*, (in Japanese).
- Yamashita, S. and S. Miura (2011), “Prediction Accuracy of Credit Risk Model: Accuracy Ratio and Evaluation Indicator”, *Asakura Shoten* (in Japanese).