ADBI Working Paper Series

CREDIT RISK REDUCTION EFFECT ON SMALL AND MEDIUM-SIZED ENTERPRISE FINANCE THROUGH THE USE OF BANK ACCOUNT INFORMATION

Naoko Nemoto, Naoyuki Yoshino,
Yutaka Okubo, Daimei Inaba,
and Kentaro Yanagisawa

No. 857
August 2018

Asian Development Bank Institute

Naoyuki Yoshino is Dean and CEO of the Asian Development Bank Institute. Naoko Nemoto is a financial economist at the Asian Development Bank Institute. Yutaka Okubo is CEO of the Risk Data Bank of Japan. Daimei Inaba is Executive Director of the Risk Data Bank of Japan. Kentaro Yanagisawa is Group Manager of the Risk Data Bank of Japan.

Suggested citation:

Nemoto, N., N. Yoshino, Y. Okubo, D. Inaba, and K. Yanagisawa. 2018. Credit Risk Reduction Effect on Small and Medium-Sized Enterprise Finance through the Use of Bank Account Information. ADBI Working Paper 857. Tokyo: Asian Development Bank Institute. Available: https://www.adb.org/publications/credit-risk-reduction-effect-sme-finance-through-bank-account-information

Please contact the authors for information about this paper.

Email: nnemoto@adbi.org

The authors would like to thank Soheon Lee, a research associate at ADBI, for her research assistance.

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

Abstract

This paper verifies the impact of bank account information, such as information on deposits and withdrawals, that is not necessarily fully accounted for in conventional internal ratings and that can affect the accuracy of the default predictions of small and medium-sized enterprises (SMEs). The analysis demonstrates that the accuracy of default predictions improves when a model based on bank account information is used in addition to the default prediction model based on traditional financial information. The analysis also shows that the degree of improvement increases when the size of the company is small. For small companies, the quality of financial data is generally assumed to be low, but the bank account information model can complement the incomplete data. In addition, for small firms, the bank account information model shows better default prediction capability compared to the financial model, which implies the possibility that banks could extend loans even if only the bank account information is available. The correlation coefficients of the financial model and the bank account model are higher than 50% but not very high, suggesting that these models evaluate borrowers from different perspectives.

This study suggests the possibility of analyzing credit risk more easily without past financial information, especially for small enterprises. If the bank account information model is utilized, banks can reduce credit costs and loan review times and costs, which will make SME financing more efficient and smooth. The empirical analysis in this paper focuses on SMEs in Japan, but the results can also be applied to other countries, particularly emerging countries in Asia.

Keywords: small and medium-sized enterprise finance, credit risk analysis, big data, bank account information

JEL Classification: G2, G21

Contents

1. INTRODUCTION

The growth of small and medium-sized enterprises (SMEs) is a critical issue for the economic development of Japan and other Asian countries. SMEs rely on bank loans for their external financing, but they sometimes have difficulties in obtaining sufficient funds from banks in a timely manner. This paper examines whether the utilization of bank account information, such as information on deposits and withdrawals, which has not been fully accounted for in Japanese banks' loans to SMEs, can enhance the accuracy of credit risk measurements.

SMEs play an important role in the development of the Japanese economy and account for about 40% of Japan's total gross domestic product (GDP) and 74% of all employers. Japanese companies mainly rely on indirect financing by banks, but the trend is even stronger for SMEs, and banks make up much of external borrowing. The lending volume of national banks for SMEs was ¥248 trillion in 1998. The figure has since declined but has been recovering since 2012 and remained at ¥212 trillion as of the end of 2017. Bank lending as a proportion of lending also declined from 49.3% in 1998 to 42.0% in 2016.

The lack of public information on SMEs' corporate activities compared to large enterprises, and the large asymmetry of information between borrowers and lenders, makes it difficult for banks to manage the credit risk of SMEs (Yoshino and Yamagami 2013). Therefore, there is a tendency to rely on collateral, such as real estate, personal assets, and the guarantees of CEOs, rather than judging the creditworthiness of the company itself (Financial Services Agency 2003). Small companies also tend to rely on their own funds, funds from acquaintances, and public funds, etc. without depending on banks (Small and Medium Enterprise Agency 2015; Yamori 2003).

Analysis of the creditworthiness of SMEs by banks has made much progress since the bad loan problems of the 1990s. One example is the implementation of the internal credit rating system that ranks companies according to financial strength. In addition to major financial indicators, such as the capital adequacy ratio, there are cases where qualitative factors, such as management's abilities and financial transparency, are also taken into account. Also since, the 2000s, the financial scoring model has become pervasive (Financial Services Authority 2003). Scoring is a lending model constructed by statistical methods that estimates the probability of bankruptcy of loan claims and uses the probability to determine loan extension and loan rate spreads. The scoring method does not manage risks on a case-by-case basis but manages the risks on loans throughout the portfolio control based on the law of large numbers. Therefore, the accuracy tends to increase as the data pool becomes larger, so the construction of the database is important. The Japan Risk Data Bank (RDB)[1], comprised of major banks and regional banks, was established in 2000 as the first data consortium in Japan, and, in 2001, the SME Credit Risk Information Database (CRD)[2] was established under the initiative of the Ministry of Economy, Trade and Industry, and the Small and Medium Enterprise Agency. These data are used for loan reviews, interest

---

1 The Japan RDB has more than 60 financial institutions, such as major Japanese banks and regional banks, as members and shares the credit risk information of 650,000 client companies on an anonymous basis. In addition to financial information, operational risk information and bank account dynamics information, etc. are also gathered.

2 Approximately 190 financial institutions, including credit unions, are members of the SME Credit Risk Information Database (CRD) and share credit risk information on about 2 million credit companies on an anonymous basis. The average size of SMEs is slightly smaller than in the RDB.

rate setting, and portfolio management, etc. and have contributed to the advancement of credit risk analysis by banks.

On the other hand, there are limits to internal ratings and scoring by banks (Hirata 2005). First, in many cases, there are problems with the quality of the financial statements of SMEs. According to a survey by the Small and Medium Enterprise Agency, only about 30% of firms are considering preparing accounts based on appropriate accounting for strengthening financing capability (Small and Medium Enterprise Agency 2004). Second, there is a time lag of information. There have been many cases where the latest financial statements for the settlement dates acquired for examination were from 3–15 months ago, so the current state of the companies was unclear. In addition, monitoring after financing is not sufficient. With financial statements alone, banks have difficulties in grasping the situations of its clients, which can change daily throughout the fiscal year.

To solve these problems, banks can detect the window dressing of accounts by interviewing and researching changes in the business environment after the settlement date to counteract the weaknesses of ratings and scoring. However, under prolonged monetary easing, the profitability of financial institutions has deteriorated, and efficient management is required now more than ever. Financial technology (fintech) companies that provide loans to individuals and SMEs quickly and easily via personal computers and mobile phones are also entering the market. Under such an environment, it is getting harder for banks to continue traditional labor-intensive lending.

Utilizing bank account information, as considered in this paper, could increase the ability of banks to analyze the credit risk of SMEs and contribute to reductions in the time and costs required for review. In Japan, most money transfers and settlements, which are the result of corporate activity, are done through bank accounts, and “account history information” contains abundant information on businesses—specifically, deposit account withdrawal information, inward and outward remittance data, bills for collection, discounted commercial bills, electronically recorded monetary claims, and loan execution collection details, etc. By continuously monitoring these data, it is possible to continuously capture the actual business situations of clients over time and predict changes in their credit situations.

Numerous academic studies have proved that SMEs can be ranked by credit risk using models that utilize corporate financial data or bank lending data, such as delinquency conditions (Behr, Guttler, and Plattner 2004; Yoshino and Taghizadeh-Hesary 2014). There are also several empirical studies on the changes in the accuracy of the model depending on the choice of variables and whether to incorporate qualitative information when constructing the credit risk model. However, although the importance of bank account information is recognized in banking practices, research that testifies the validity of bank account information has been rare. The contribution of this paper is to show empirical analysis of bank account information that utilizes a large pool of both financial and bank account information (big data) held by the RDB.

This paper demonstrates that the accuracy of default prediction improves when a model based on bank account information is used in addition to the default prediction model based on traditional financial information. There is a tendency for the improvement to increase when the size of a company is small. If the size of the company is small, the quality of financial data is generally assumed to be low, but the bank account information model can complement this. In addition, for small firms, the bank account information model shows better default prediction compared to the financial model, which implies the possibility that banks could extend loans even if only the bank account information is available. The correlation coefficients of the financial model and the bank account model are higher than 50% but not very high, suggesting these models evaluate borrowers from different perspectives.

If the utilization of the bank account information model spreads, banks can reduce their credit costs and reviews time and costs, and make loans to SMEs more efficiently. The empirical analysis in this paper is targeted at SMEs in Japan, but the results could also be relevant for other countries, particularly in emerging countries in Asia.

SMEs in Asian countries have not obtained sufficient funds for growth, and according to a survey by the International Finance Corporation, the supply shortage of loans to SMEs as of the end of 2014 was $706 billion for East Asia and $2,060 billion for South Asia. In Asia, the loan balance for SMEs accounts for 19% of bank loans overall, which is lower than in other regions. There are several reasons why SMEs do not seek bank loans, such as the collateral requirements, complicated application procedures, bank lending conditions not meeting their needs, and high loan interest rates (ADB 2015). The asymmetry of information is large, especially in low-income countries, and the quality of the financial information on SMEs is low and many are not audited.

Regarding a common database, there are public credit information agencies in 8 of the 20 ADB member countries. Information collected by credit information agencies includes (1) information on business, (2) information on bank transactions (including default information), and (3) information on firms' financial situation. However, companies that do not have bank transactions do not have information on business and bank transactions, and the credibility of the information provided by member banks is not accurate enough in many cases, so it appears that banks only use this information for supplementary purposes when making loans. In most of the countries, there have not been solid systems for banks to share the credit risk information of their clients or provide risk information, such as the estimated default rates of companies such as by the RDB.

In Asia, the proportion of companies receiving bank loans is as small as 15.4% for small companies, but 79.4% have bank accounts and use bank accounts for settlements, etc. (ADB 2015). If bank account information were to be used systematically, we could expect the expansion of loans for SMEs.

The structure of this paper is as follows. Section 2 will look at the findings from previous research. Section 3 shows the theoretical background; and Section 4 explains the credit risk prediction method; the variables of the respective models using financial information, bank account information, and both methods together; and the results of the constructed credit models. Section 5 shows verification of the credit risk models in terms of the default prediction ability, and Section 6 concludes. The Appendix supplements the explanation of the methods used for verification.

2. LITERATURE REVIEW

There have been many studies, inside and outside Japan, which introduce credit risk models for SMEs. Most of them demonstrate that credit risk models utilizing financial statements appear to be useful in differentiating SMEs according to their credit risk and thus in banks' loan decisions (Behr, Guttler, and Plattner 2004; Yoshino and Taghizadeh-Hesary 2014). Behr, Guttler, and Plattner (2004) quantify the credit risk of German SMEs based on a logistic regression model. The variables selected in their model are mainly the financial statements of SMEs (e.g., the equity ratio), but the model also incorporates qualitative features of companies, such as location, corporate structure, and line of business. The default probability of large firms estimated by this model shows a high correspondence with the credit ratings disclosed on them by global credit rating agencies. Also, the accuracy ratio (AR) (see Appendix) for the credit scores estimated by this model is higher than 80, which proves that the model has a good prediction ability. Since the information on the firms tends to be concentrated on major banks in Germany, thereby hindering the diversification of the funding channels and, thus, the growth of SMEs (the hold-up problem), another advantage of the scoring model would be that it could mitigate this problem. By providing an objective benchmark, the scoring model could eventually contribute to the diversification of the funding channels of SMEs.

Grunert et al. (2005) demonstrate that the combination of a firm's quantitative information and qualitative information, such as on management efficiency, results in a mild increase in the model's default prediction ability. Similarly, a model by Anzai (2015) incorporates qualitative information on technology levels. It introduces a technology level index for SMEs and shows that firms with a high level of technology are capable of relatively swift recovery from the downturn of profitability. This implies a correlation between SMEs' financial statements and their business competitiveness. On the other hand, Freichs and Wahrenburg (2003) point out that the model's estimation ability can be improved enough to overcome the limitations from the lack of qualitative information when the data pool covers vast amounts of quantitative information from many banks.

Although most of the studies on the credit risk model focus on the case of advanced countries, there exist a few studies that cover SME finance in emerging countries. Yoshino and Taghizadeh-Hesary (2014) examine the effectiveness of credit risk assessments of Iranian companies based on financial statements held by banks. More specifically, they present a model estimating the default probabilities of SMEs, with the relevant variables selected through principal component analysis and cluster analysis (clustering). The results of the study indicate that banks can classify SMEs according to their credit risk based on the data of financial statements held by banks, and this can be utilized in loan decisions and the pricing of loan rates. The study also emphasizes the importance of a corporate credit information database. Similarly, Yoshino et al. (2016) develop a model for assessing the default probability of SME lending in Thailand through principal component analysis. Based on a database of SME lending provided by the public credit information agency of Thailand, the model classifies SMEs according to their estimated default probability.

On the other hand, there are studies which point out the limitations and room for improvement of the scoring model while admitting the effectiveness of the approach itself. Hirata (2005) examines the advantages and the disadvantages of the scoring model. For the advantages of the scoring model, the study mentions that (1) banks can streamline the troublesome process of loan examination in SME lending; (2) by quantifying risk, banks can monitor the quality of portfolios as a whole, which can also be utilized in new customer acquisition; and (3) the scoring model supports market-based finance as it makes securitization easier. On the other hand, the paper points out the disadvantages of the scoring model that (1) there will be sample selection bias since the model only covers the information of firms who applied for bank loans; (2) the accuracy of the model is limited due to the low quality of SMEs' financial statements; (3) the model does not necessarily guarantee more efficiency since it often requires an additional process, such as in-person interviews, due to the imperfect information.

This paper builds a credit risk model based on financial statements as well as a model based on bank account information, such as the change in deposits or the level of deposits and loans in comparison with sales. In building the credit risk model, the paper refers to the methodology used in Yoshino and Taghizadeh-Hesary (2014) and Yoshino et al. (2016). Then, we examine the accuracy of the default prediction of the models. The results indicate that the accuracy of the default prediction of the model based on financial statements is improved when combined with the model based on bank account information. Especially for the case of micro-sized firms, who are generally assumed to have low-quality financial statements, the model tends to show a large degree of improvement of the default prediction accuracy. This implies that the model based on bank account information could mitigate the information quality issue. We also find that the creditworthiness predicted by each model is correlated with each other and that the accuracy of default prediction by the model based only on dynamic bank account information is comparable to the model based on financial statements in some cases. Bank account information could reveal the actual condition of a firm's cash flow, and its time lag is much shorter than financial statements, which makes it a good complement to financial statements of low quality. Moreover, it could reduce sample selection bias since it also contains the information of the firms without bank loans.

Many studies have indicated that the accuracy of the credit risk model depends on the selection of the variables. However, the effectiveness of the utilization of dynamic bank account information has not been examined before this paper, despite the growing perception of its importance among bankers and financiers. Another new contribution of this paper is that it utilizes big data derived from the RDB's database, which collects both financial and bank account information from Japanese banks. In addition, this paper applies principal component analysis, which has the advantage of reducing multidimensional datasets to lower dimensions of analysis while minimizing the loss of information. Principal component analysis also ensures objective and quantitative assessment in selecting variables.

3. THEORETICAL BACKGROUND

Before delving into an empirical analysis of credit risk, here we describe the theoretical background regarding the relationship between a bank's profit maximization and the improvement of credit risk analysis. A bank's profit maximization behavior can be written as follows, where Π denotes the bank's profit.

$$Max\ \Pi = r_L(L)L - \rho(CRA, Z_i\ L)L - r_D D - C(L, D)$$

s.t. *kDeposit* + *Loan* = *Deposit* + *Capital*

Assumptions:

$r_L(\cdot)$ denotes the loan rate, which is a function of the amount of the loan

$\rho(\cdot)$ denotes the default risk, which is a function of the credit risk analysis (*CRA*), the information of individual company(Z_i), and the loan amount (L)

r_D denotes the interest rate on the deposit (D)

$C(\cdot)$ denotes the operating cost, which is a function of the loan(L) and the deposit (D)

For simplicity, the capital cost is abstracted from the model, and we assume that the bank uses all of the deposits (D) in making loans (L). Also, instead of assuming perfect competition, we assume that banks are in oligopolistic competition. That is, the loan rate is not given but is determined by the supply and demand of the bank's loanable funds.

$$\Pi = r_L(L) \cdot L - \rho(L, Z) \cdot L - r_D(1-k)L - C(L, D) \tag{1}$$

$$\frac{\partial \Pi}{\partial L} = \left(\frac{\partial r_L}{\partial L} \cdot L + r_L\right) - \left(\frac{\partial \rho}{\partial L} \cdot L + \rho\right) - (1-k)r_D - \frac{\partial C}{\partial L} = \tag{2}$$

$$L \cdot \left(\frac{\partial r_L}{\partial L} - \frac{\partial \rho}{\partial L}\right) = -r_L + (1-k)r_D + \frac{\partial C}{\partial L} + \rho \tag{3}$$

$$L = \frac{1}{\left(\frac{\partial r_L}{\partial L} - \frac{\partial \rho}{\partial L}\right)}\left(-r_L + (1-k)r_D + \frac{\partial C}{\partial L} + \rho\right) \tag{4}$$

$$r_L = -\frac{\partial r_L}{\partial L} \cdot L + \left(\frac{\partial \rho}{\partial L} \cdot L + \rho\right) + (1-k)r_D + \frac{\partial C}{\partial L} \tag{5}$$

Equation (1) is the profit function of the bank: the loan revenue deducted by the credit cost that occurs at the time of borrower's default, the funding cost of deposits, and the operating cost. The bank attempts to maximize its profit, which requires equation (2) to hold true as a first-order condition. Equation (2) can be developed into equation (4), keeping the left side as the loan (L). If the bank could reduce the default risk with the newly introduced credit risk analysis, the marginal default risk, $\partial \rho$, will decrease, and the bank loan (L) will increase. Also, if the credit risk analysis is conducted systematically and effectively, the marginal operating cost, ∂C, will also decrease, which will lead to an increase in the bank loan.

Equation (2) can also be developed into equation (5), keeping the left side as the interest rate (r). If the bank can reduce the default risk with the newly introduced credit risk analysis, the marginal default risk, $\partial \rho$, will decrease. Also, if the credit risk analysis is conducted systematically and effectively, the marginal operating cost, ∂C, will also decrease, which will lead to a decrease in the interest rate, r. Therefore, it can be concluded that, other things being equal, the new methods of credit risk analysis introduced in this paper could lead to an increase in the banks' SME loans and a decrease in interest rates.

4. ANALYSIS OF THE SME CREDIT RISK MODEL

4.1 Credit Risk Models Using Financial Information and Bank Account Information

For our analysis, we develop three credit risk models aimed at assessing the credit quality of companies: (1) a model using financial information, such as balance sheets and financial statements (financial model); (2) a model using bank account information (bank account model); and (3) a model using both financial information and bank account information (hybrid model).

We employ the following logistic regression model to develop the credit risk assessment model in this paper. The logistic regression model has been used by previous research (Behr, Guttler, and Plattner 2003) and for Japanese banks' credit risk models.

$$\ln\left(\frac{1-\rho}{\rho}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

In this equation, $\ln(\cdot)$ is the natural logarithm; ρ is the probability of default; β_0 is the intercept; $x_i, i = 1, \cdots, k$ are financial (dynamic) indicators, which are explanatory variables; $\beta_i, i = 1, \cdots, k$ are coefficients; and k represents the number of indicators used. In the process of developing the model, we assigned 1 to companies that defaulted and 0 to companies that did not default during the observation period. Then, we used these variables including $x_i, i = 1, \cdots, k$ and estimated $\beta_i, i = 0, \cdots, k$, using the maximum likelihood approach. In addition, the following s is called the "beta value", where the estimated values of the intercept and coefficients are $\hat{\beta}_i, i = 0, \cdots, k$.

$$\mathrm{s} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

The next step is to identify the variables. For this study, we follow Yoshino and Taghizadeh-Hesary (2014) and Yoshino et al. (2016), who employed principal component analysis for the credit risk model of SMEs. We applied component analysis to the financial data and bank account data and then selected the components that had eigenvalues of more than 1. Principal component analysis is a technique for simplifying and reducing multidimensional datasets to lower dimensions of analysis while minimizing the loss of information. This method has the additional advantage of being able to select variables using quantitative and non-discretionary methods. Each component is designed to be uncorrelated so that multicollinearity issues are detected and the prediction is reliable.

The principal component Z is shown by the following equations:

$$Z_1 = \mathrm{a}_{11}\mathrm{x}_1 + \mathrm{a}_{12}\mathrm{x}_2 + \cdots + \mathrm{a}_{1P}\mathrm{x}_P = \sum \mathrm{a}_{1i}\mathrm{x}_i \,; \sum \mathrm{a}_{1i}^2 = 1$$

$$Z_k = \mathrm{a}_{k1}\mathrm{x}_1 + \mathrm{a}_{k2}\mathrm{x}_2 + \cdots + \mathrm{a}_{kP}\mathrm{x}_P = \sum \mathrm{a}_{ki}\mathrm{x}_i \,; \sum \mathrm{a}_{ki}^2 = 1$$

Here, *x* is an indicator of the financial information and bank account information, *a* is the loading vector of the indicator, and *k* is the number of components. In the process of creating the hybrid model, we apply principal component analysis for both the financial and bank account information and select components with higher eigenvalues. As the hybrid model incorporates both financial and bank account information, it is expected that the model will have higher default predictability compared to the model using only one factor.

4.2 Characteristics of Financial Information and Bank Account Information

Financial information refers to financial statements consisting of a balance sheet and an income statement, which represent a brief picture of a company's operational performance and financial activities during a specific period. In general, Japanese financial institutions obtain the most recent financial information for multiple fiscal periods from their prospective corporate borrowers prior to loan approvals. Corporate financial statements are highly reliable because companies are required to prepare them pursuant to the accounting standards set forth by law. They also contain a great deal of information. However, financial statements are generally prepared only annually,[3] and financial institutions need to wait for another three months or so to obtain such reports summarizing business results for the previous 12-month period. Therefore, one could argue that banks may fail to identify changes in their customers' business conditions that might have occurred over around the past 15 months.

Bank account information refers to information related to deposit account balances, transaction amounts, outstanding loan balances, and loan extensions and collections. In the case of a company that has accounts with multiple banks, the information that each bank can obtain is only that of the company's account with that bank. That is, if the bank is the main bank of the company and is used by the company for a significant share of its activities, it can grasp the overall status of the company's business because it is able to accumulate broader bank account information than other banks. In contrast, the bank account information is less useful if a bank has a weak business relationship with the target company and may even be useless for a bank with which the target company has no account.

For a financial institution that serves as the main bank for the target company, bank account information on the company has been already accumulated within its database, so acquisition costs are low, and the data is more up-to-date compared with financial information. As such, the institution could use them effectively in monitoring the target company's day-to-day changing status.

In our financial model, we use 77 financial indicators, as shown in Table 1. Previous research (Yanagisawa et al. 2007) used 91 financial indicators, from which we selected 82 indicators and omitted those with high proportions of zero value or values beyond the limits of 2.5% and 97.5%. Furthermore, we omitted indicators with absolute values to have consistency between the financial model and bank account model.

The bank account model uses the balances of liquid deposits, fixed deposits, and outstanding loans as well as the amount of inflows and outflows of liquid deposits. Checking and savings accounts are treated as liquid deposits, and term deposits are treated as fixed deposits. In addition, the sum of the liquid deposit balance and the fixed deposit balance is defined as the total deposit balance; the total deposit balance minus the outstanding loan balance is defined as the net deposit balance. For the use of the bank account model, 79 indicators are created in the form of ratios. To calculate the ratios, the numerators are the amounts of deposits, loans, and net deposits at the end of the month. We also use the minimum and maximum amounts and the standard deviations in a certain period as numerators. The denominators are the total sales or loan amounts. The indicators include the percentage of growth and the reduction of deposits and loans.

[3] Some corporations, such as listed companies, are required by law to prepare quarterly financial reports. However, most SMEs prepare statutory financial statements only once a year in Japan.

Table 1: List of Financial Indicators

Indicators		
Cash ratio	Interval measure	Debt to equity ratio
Net-debt turnover period	Net interest-bearing debt turnover period	Gross profit to interest and discount expense ratio
Cash plus marketable securities ratio	Interest burden to sales ratio	Net interest burden to sales
Interest coverage ratio	EBIT to interest and discount rate expense ratio	Cash to debt ratio
Cash flow to sales ratio	Cash flow to total assets ratio	Cash flow to total liabilities ratio
Cash flow to interest-bearing debt ratio	Cash flow to debt ratio	Credit interval
Cash flow to total expense ratio	Cash flow to operating cost	Value-added to sales ratio
Labor share	Capital investment efficiency	Depreciation costs/ordinary earnings
Break-even point ratio	Ordinary profit and loss ratio	Cash asset ratio
Working capital ratio	Quick (acid test) ratio	Working capital to sales ratio
Trade payables to trade receivables ratio	Equity ratio	Fixed assets to equity ratio
Fixed assets to long-term liabilities ratio	Leverage	Long-term debt ratio
Debt ratio	Equity to total asset ratio	Equity to total liabilities ratio
Capitalization rate	Debt to sales ratio	Quick ratio
Price book-value ratio	Working capital to total asset ratio	Gross profit to sales ratio
Sales, general, and administrative expenses to sales ratio	Earnings before interest and taxes to sales ratio	Ordinary income to sales ratio
Income before tax to sales ratio	Current income to sales ratio	Return on total assets
Earnings before interest and taxes to total assets ratio	Ordinary income to total assets ratio	Current income to total assets ratio
Return on assets	Unappropriated retained earnings to total assets ratio	Current ratio
Return on equity	Earnings before interest and taxes to equity ratio	Ordinary income to equity ratio
Current income to equity ratio	Cash and deposits/short-term borrowings	Total asset turnover
Fixed asset turnover	Receivables turnover period	Inventory turnover period
Trade payable turnover period	Net interest-bearing debt to assets ratio	Reimbursement period
Interest-bearing debt to cash	Interest-bearing debt ratio	Debt to equity ratio
Earnings before interest depreciation and amortization to interest-bearing debt	Debt capacity ratio	Turnover period of interest-bearing debt

Source: Authors.

4.3 Sample Data

The data for both financial information and bank account information for this statistical analysis were provided by the RDB, one of the leading data consortiums in Japan. The study also covers data on loan classifications, which are assigned by member banks of the RDB for internal credit risk analysis. The base dates are set at the end of 2015 and 2016. We apply the credit scoring model to all companies and compare the status of those companies after a 1-year period to judge the accuracy of the credit scoring. The financial information is annual data that were submitted to member banks between June 2014 and September 2016. The bank account information is basically daily data from which we use 400 observation points during 1 December 2014–31 December 2016.[4] The study covers entities that meet all the following conditions:

1. Nondefault status[5] as of 31 March 2015 or 31 March 2016 (referred to as the base dates);
2. Granted loans from the data-providing banks as of the base date;
3. Able to provide financial information for a period of 3–18 months prior to the base date;
4. Able to provide bank account information for a period of 13 months prior to the base date;
5. Sales below ¥10 billion;[6] and
6. Share of deposits at the bank[7] of above 50% but below 200%.

In the next section, we show the default predictability of each model that has been described in this section. The default observation period is defined as 1 year from the base date in that calculation. We classify companies that made full repayments of their loans to the data-providing bank during the observation period as nondefaulted as of the end of the fiscal term. To show the distribution of the sample data, Figure 1 shows the common logarithms of the sales of all companies used for the analysis. The median for sales is ¥150 million, and the peak of the distribution is around that point.

---

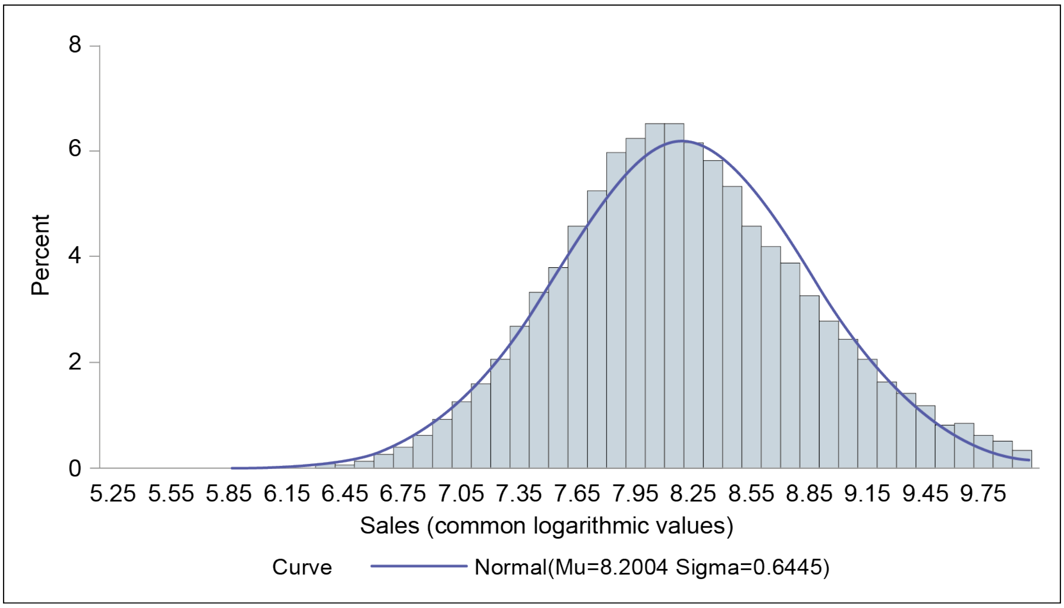
[4] It could be argued that we could collect more information if we were to extend the observation periods. On the other hand, longer periods means higher costs for acquiring the data. Therefore, we choose 13 months as the period is not too long, and we can calculate the annual growth rate.

[5] With respect to the classification of obligor status, we include firms lower than the category of “in danger of bankruptcy” under the Financial Restoration Law. This category is defined as those firms with a negative net worth or that are vulnerable to a change of environment. Under this standard, default includes obligors classified as “in danger of bankruptcy,” “de facto bankrupt,” and “bankrupt.” We consider obligors classified as “normal (*seijo-saki*),” “need special attention,” and “substandard” to be “nondefault.”

[6] This is because the focus of this study is on SMEs.

[7] The share of deposits represents the share of the target company’s account for settlement at the data-providing bank. It is calculated by dividing the amount of “business-related” deposits for a period of 12 months prior to the end of the fiscal year by the total amount of sales in the financial statement for that period. Business-related deposits are the amount of deposits that were deposited in the entity’s account at the data-providing bank minus the amounts of deposits or withdrawals that are not directly related with business, such as loans granted, withdrawals of term deposits, and transfers within the entity, leaving only the transactions that are the related sales and payments of the entity. The higher the share of deposits or withdrawals, the more effectively the model functions.

Figure 1: Distribution of Targeted Firm Sales



Note: Total number = 42,654; default = 441, non-default = 42,213.

Source: Authors' calculations.

In the process of developing a credit risk assessment model, it is common to divide the target data into two sets, one for in-sample data and another for out-sample data. The model can then be created based on the in-sample data and applied to the out-sample data to check the validity of the model. We follow this approach in our analysis here, dividing the target data into two sets in a way that the number of defaults and the number of nondefaults in the in-sample data and the out-sample data become 1-to-1.[8] Table 2 shows the results of the default ratios of the target companies.

Table 2: Default Frequency Data

	Default Frequency Data			
	Total	**Nondefault**	**Default**	**Default Ratio**
In-sample data	21,328	21,107	221	1.04%
Out-sample data	21,326	21,106	220	1.03%
Total	42,654	42,213	441	1.03%

Source: Authors' compilation.

The number in the "default" column indicates the number of companies that defaulted in the observation period, and the "nondefault" column shows the number of companies that did not default in the same period. Table 3 shows the sales comparisons for the data.

8 To make sure that no large differences in the characteristics exist between the in-sample data and the out-sample data, we compared the sales of the two sets of data. The p-value in the t-test is sufficiently large, which confirms that there is not a big discrepancy between the average sales in the in-sample data and those in the out-sample data.

Table 3: Comparison of Sales for the In-Sample and Out-Sample Data

		Corporate Sales Amounts (¥ million)						
	Cases	**Average**	**STD**	**25%**	**50%**	**75%**	**t-statistic**	**p-value**
In-sample data	21,328	492	1,043	58	147	413	–0.04	0.9711
Out- sample data	21,326	492	1,048	58	145	418		
Total	42,654	492	1,045	58	146	415		

Source: Authors' compilation.

4.4 Analysis Results

Using the datasets explained in Sections 4.2 and 4.3, we develop (1) a financial model using the financial information of SMEs, (2) a bank account model using the account information from data-providing banks, and (3) a hybrid model using both the financial and bank account information. The model development procedures are as follows:

1. Set up the upper and lower limits for the indicators[9]
2. Apply principal component analysis to the indicators and find components and eigenvalues of each component.
3. Select those factors accounting for more than 10% of the variance (eigenvalues are larger than 1), which are regarded as significant components for explaining the variance
4. Develop the logistic model using factors as explanatory variables and find the estimated coefficients of the variables

The results of the principal component analysis for the financial model are shown in Table 4, the bank account model results in Table 5, and the hybrid model in Table 6. The first process through the third process are performed on the overall data including the in-sample data and the out-sample data.

Table 4 shows that the first component with the highest eigenvalue has 22.6% variance, which means it explains 22.6% of the total variance of the financial ratios. Factors 1–16 are used for the financial model as they have eigenvalues higher than 1. Factors 1–11 are used for the bank account model, and factors 1–26 are used for the hybrid model. The cumulative variances of the models are 84.1%, 91.2%, and 86.9%, respectively, indicating that a large part of the total variance is explained by the components.

[9] In setting the upper and lower limits for each indicator, we replace large (small) values that could impair the estimation accuracy of coefficients of the logistic regression model with the upper (lower) limit. The lower limit is set at 2.5% and the upper limit is set at 97.5%.

Table 4: Results of Principal Component Analysis: Financial Information (Balance Sheets and Financial Statements)

Principal Component	Eigenvalue	Share of Variance (%)	Cumulative Variance (%)	Characteristics of Major Components
1	**17.37**	**22.6**	**22.6**	Reliance on borrowing, such as total assets/borrowings.
2	**11.86**	**15.4**	**38.0**	Profitability measurement, including earnings versus expenses.
3	**7.03**	**9.1**	**47.1**	Profitability and efficiency, such as break-even point ratio.
4	**5.78**	**7.5**	**54.6**	Liquidity at hand, such as total liquid deposits/total assets.
5	**3.72**	**4.8**	**59.4**	Costs and expenses in comparison with sales.
6	**3.22**	**4.2**	**63.6**	Ability to cover debt payment (earnings/interest expense).
7	**2.47**	**3.2**	**66.8**	
8	**2.09**	**2.7**	**69.5**	
9	**1.94**	**2.5**	**72.0**	
10	**1.71**	**2.2**	**74.3**	
11	**1.51**	**2.0**	**76.2**	
12	**1.38**	**1.8**	**78.0**	
13	**1.27**	**1.7**	**79.7**	
14	**1.18**	**1.5**	**81.2**	
15	**1.15**	**1.5**	**82.7**	
16	**1.11**	**1.4**	**84.1**	
17	0.99	1.3	85.4	
18	0.93	1.2	86.6	
19	0.86	1.1	87.7	
20	0.79	1.0	88.8	
21	0.70	0.9	89.7	
22	0.60	0.8	90.5	
23	0.56	0.7	91.2	
24	0.51	0.7	91.8	
25	0.48	0.6	92.5	
26	0.43	0.6	93.0	
27	0.42	0.5	93.6	
28	0.37	0.5	94.0	
29	0.34	0.4	94.5	
30	0.33	0.4	94.9	

Source: Authors' calculations.

Table 5: Results of Principal Component Analysis: Bank Account Information

Principal Components	Eigenvalue	Share of Variance (%)	Cumulative Variance (%)	Characteristics of Major Components
1	**24.73**	**38.6**	**38.6**	Total deposit amount versus sales, reflecting liquidity and cash flow.
2	**11.99**	**18.7**	**57.4**	Total borrowing versus total sales.
3	**5.16**	**8.1**	**65.4**	Combination of deposits/sales and loan/deposits.
4	**4.19**	**6.5**	**72.0**	Total borrowings versus total deposits.
5	**3.12**	**4.9**	**76.9**	Increase and decrease of loans.
6	**2.28**	**3.6**	**80.4**	
7	**1.73**	**2.7**	**83.1**	
8	**1.49**	**2.3**	**85.5**	
9	**1.43**	**2.2**	**87.7**	
10	**1.22**	**1.9**	**89.6**	
11	**1.05**	**1.6**	**91.2**	
12	0.87	1.4	92.6	
13	0.60	0.9	93.5	
14	0.55	0.9	94.4	
15	0.45	0.7	95.1	
16	0.35	0.6	95.6	
17	0.31	0.5	96.1	
18	0.29	0.5	96.6	
19	0.27	0.4	97.0	
20	0.19	0.3	97.3	
21	0.18	0.3	97.6	
22	0.17	0.3	97.8	
23	0.16	0.3	98.1	
24	0.15	0.2	98.3	
25	0.14	0.2	98.5	
26	0.11	0.2	98.7	
27	0.11	0.2	98.9	
28	0.10	0.2	99.0	
29	0.08	0.1	99.2	
30	0.08	0.1	99.3	

Source: Authors' calculations.

Table 6: Results of Principal Component Analysis: Hybrid

Principal Component	Eigenvalue	Share of Variance (%)	Cumulative Variance (%)	Characteristics of Major Components
1	**32.26**	**22.9**	**22.9**	Deposits versus total sales (bank account).
2	**22.22**	**15.8**	**38.6**	Reliance on borrowings (financial and bank account).
3	**11.59**	**8.2**	**46.9**	Profitability, such as breakeven point (financial).
4	**7.28**	**5.2**	**52.0**	Capital adequacy (financial).
5	**5.53**	**3.9**	**55.9**	Total borrowings versus deposits (bank account).
6	**5.01**	**3.6**	**59.5**	Increase and decrease of deposits (bank account).
7	**4.33**	**3.1**	**62.6**	
8	**3.58**	**2.5**	**65.1**	
9	**3.13**	**2.2**	**67.3**	
10	**2.96**	**2.1**	**69.4**	
11	**2.43**	**1.7**	**71.1**	
12	**2.26**	**1.6**	**72.7**	
13	**2.24**	**1.6**	**74.3**	
14	**1.86**	**1.3**	**75.6**	
15	**1.80**	**1.3**	**76.9**	
16	**1.61**	**1.1**	**78.1**	
17	**1.53**	**1.1**	**79.1**	
18	**1.43**	**1.0**	**80.2**	
19	**1.39**	**1.0**	**81.1**	
20	**1.33**	**0.9**	**82.1**	
21	**1.28**	**0.9**	**83.0**	
22	**1.17**	**0.8**	**83.8**	
23	**1.14**	**0.8**	**84.6**	
24	**1.12**	**0.8**	**85.4**	
25	**1.08**	**0.8**	**86.2**	
26	**1.05**	**0.7**	**86.9**	
27	0.98	0.7	87.6	
28	0.96	0.7	88.3	
29	0.91	0.6	89.0	
30	0.86	0.6	89.6	

Source: Authors' calculations.

For the financial model, the main components have the following characteristics:

1. The first component has variables that reflect the reliance on borrowing. For the first component, the variables with large loadings are the total assets divided by total borrowings, total assets divided by interest bearing debt, and the capital ratios (capital/ total assets).
2. The second component has variables that reflect profitability. For the second component, the variables with large loadings are the indicators reflecting earnings versus expenses, such as the total expenses cash flow ratio and the ordinary income ratio. Other variables are those reflecting total sales versus total borrowings (e.g., the interest-bearing debt turnover and the total sales borrowing ratio).
3. The third component has variables that also reflect profitability. For the third component, the variables with large loadings are the breakeven point ratio and the return on equity.
4. The fourth component has variables that reflect liquidity at hand. For the fourth component, the variables with large loadings are the total liquid deposits/total assets, the cash ratio, and the quick ratio. Total capital/total earnings also has a large loading, which implies that companies with high earnings tend to have good liquidity.
5. The fifth component has variables that reflect costs and expenses in comparison with sales. For the fifth component, the variables with large loadings are indicators such as operating expenses versus total sales and sales margins. They also include interest expense and discounted notes/gross profits and trade receivables turnover.
6. The sixth component has variables that reflect earnings capacity to cover debt payment. For the sixth component, the variables with large loadings are indicators reflecting operating profits and earnings before interest, tax depreciation, and amortization (EBITDA) compared to interest expenses and interest-bearing debt.

For the bank account model, the main components have the following characteristics:

1. The first component has variables that compare the total deposit amounts and total sales. The level of total deposits versus total sales reflects whether a company has sufficient liquidity to support business activities and has the ability to produce a stable cash flow.
2. For the second component, the variables with large loadings are the indicators reflecting total borrowings versus total sales. The higher total borrowing ratio means higher credit risk for the firm.
3. The third component has variables that are combinations of the deposit/total sales ratios and loan/deposit ratios. The deposit ratios include liquid deposits/total sales and the increase of total deposits/total sales.
4. For the fourth component, the variables with large loadings are indicators reflecting total borrowings in comparison with total deposits.
5. The fifth component has variables that reflect the increase and decrease of loans. The variables with large loadings are the net increase of total loans/total sales and the increase of total loans compared to the level in the previous year.

For the hybrid model, which incorporates both the financial information and bank account information, the main components have the following characteristics:

1. The first component has variables that reflect deposits versus total sales (bank account information).
2. The second component has variables that reflect reliance on borrowings. The variables with large loadings are the total loan amounts and a mixture of financial and bank account information (financial and bank account information).
3. The third component has variables that reflect profitability. The variables with large loadings are the breakeven point and return on equity ratios (financial information).
4. The fourth component has variables that reflect capital adequacy in comparison with total assets and cash flow (financial information).
5. The fifth component has variables that reflect total borrowings compared to deposits (bank account information).
6. The sixth component has variables that reflect the increase and decrease of deposits (bank account information).

The first component, namely deposits versus total sales, has a 22.9% variance. Among the top six components, bank account information is included in four components, which indicates bank account information is critical for default prediction.

Table 7: Results of Estimates: Financial Model (Balance Sheets and Financial Statements)

Variable	Estimated Coefficient	t-statistic	p-value
Intercept	5.7256	1,153.4066	0.0000
Component 1	0.3652	68.3285	0.0000
2	0.1112	11.2091	0.0008
3	–0.0181	0.2495	0.6174
4	0.2026	13.0828	0.0003
5	–0.1048	2.7642	0.0964
6	0.2678	7.6029	0.0058
7	–0.1060	2.8531	0.0912
8	0.1407	6.4802	0.0109
9	0.2534	23.9213	0.0000
10	0.1122	2.5644	0.1093
11	–0.1519	7.2368	0.0071
12	–0.0635	0.9261	0.3359
13	–0.0731	1.1677	0.2799
14	–0.0266	0.1864	0.6660
15	–0.2497	13.0188	0.0003
16	–0.0552	0.6541	0.4186
AIC	2,068.4000		
McFadden R2	0.1728		

Source: Authors' calculations.

Table 8: Results of Estimates: Bank Account Model

Variable	Estimated Coefficient	t-statistic	p-value
Intercept	8.4585	264.0288	0.0000
Component 1	0.7818	98.1440	0.0000
2	–1.4290	55.8695	0.0000
3	–1.2847	42.7388	0.0000
4	2.6597	40.8866	0.0000
5	0.0696	2.1601	0.1416
6	0.1841	11.1913	0.0008
7	0.1363	4.1128	0.0426
8	0.1731	3.5690	0.0589
9	0.1272	0.3009	0.5833
10	0.1257	2.5162	0.1127
11	0.0098	0.0264	0.8709
AIC	2,461.5000		
McFadden R2	0.1329		

Source: Authors' calculations.

Table 9: Results of Hybrid Model

Variable	Estimated Coefficient	t-statistic
Intercept	7.5273	329.6084
Component 1	0.5122	76.3668
2	–0.6521	40.3171
3	0.1391	7.3013
4	0.2580	33.3141
5	1.1070	22.0717
6	0.5399	21.0513
7	–0.6395	26.0072
8	0.3601	12.7396
9	0.0538	1.0830
10	0.4075	17.3367
11	0.7058	17.0777
12	–0.2333	6.5708
13	–0.0460	0.8381
14	0.1398	3.5677
15	0.0421	0.3624
16	0.0787	1.2867
17	0.0135	0.0489
18	–0.2539	9.0493
19	–0.0930	1.6797
20	–0.2502	9.4903
21	0.1906	10.9591
22	–0.0149	0.0416
23	–0.1501	4.4251
24	–0.2931	9.2046
25	0.1472	2.4150
26	0.0542	0.7916
AIC	2,461.5000	
McFadden R2	0.2092	

Source: Authors' calculations.

Tables 7–9 provide the results of the estimation of the coefficients of the logistic regression models based on selected components with eigenvalues of more than 1. The regression was run using the in-sample data. Table 7 shows the financial model, Table 8 the bank account information model, and Table 9 the hybrid model. In Table 7, the hypothesis that the coefficient is zero is rejected at the 1% level for 7 out of 16 components, showing the validity of the model. In Table 9, the hypothesis that the coefficient is zero is rejected at the 1% level for 5 out of 11 components. In Table 9, the hypothesis that the coefficient is zero is rejected at the 1% level for 14 out of 26 components. These results all demonstrate the validity of the model.

Looking at Akaike's Information Criterion (AIC), the hybrid model has a lower AIC than the financial model and bank account model, which shows that the hybrid model has a smaller prediction error compared to other two models.

5. VERIFICATION OF THE CREDIT RISK MODEL

To compare the default prediction abilities of each model, we measured the accuracy ratios (ARs) (see Appendix). The results are shown in Tables 10, 11, and 12. We also computed the ARs of each model by category and by the amount of sales and verified whether the quality of the discrimination of the model differed in terms of firm scale.

Table 10: Accuracy Ratios of the In-Sample Data

					Accuracy Ratio		
Group Segment	Total	Nondefault	Default	Default Ratio	Financial	Bank Account	Hybrid
Less than ¥30 million	2,680	2,643	37	1.4	67.6	63.2	71.8
¥30 million–¥100 million	5,634	5,557	77	1.4	67.6	67.9	74.4
¥100 million–¥300 million	6,329	6,259	70	1.1	72.3	62.1	77.0
More than ¥300 million	6,685	6,648	37	0.6	74.6	69.6	79.9
Total	21,328	21,107	221	1.0	71.6	66.8	76.5

Note: 'Financial' shows the results of the financial model using balance sheets and financial statements. 'Bank account' shows the results of the bank account model. 'Hybrid' shows the results of the hybrid model using both financial data and bank account data.

Source: Authors' calculations.

Table 11: Accuracy Ratios of the Out-Sample Data

				Default Ratio	Accuracy Ratio		
Group Segment	Total	Nondefault	Default	Total	Nondefault	Default	Total
Less than ¥30 million	2,718	2,675	43	1.6	55.3	59.5	62.6
¥30 million–¥100 million	5,722	5,638	84	1.5	53.4	59.9	62.7
¥100 million–¥300 million	6,116	6,063	53	0.9	72.8	73.7	80.2
More than ¥300 million	6,770	6,730	40	0.6	80.7	58.9	78.4
Total	21,326	21,106	220	1.0	65.1	64.6	71.4

Note: 'Financial' shows the results of the financial model using balance sheets and financial statements. 'Bank account' shows the results of the bank account model. 'Hybrid' shows the results of the hybrid model using both financial data and bank account data.

Source: Authors' calculations.

Table 12: Accuracy Ratios of the Total Data

Group Segment	Total	Nondefault	Default	Default Ratio Total	Accuracy Ratio		
					Nondefault	Default	Total
Less than ¥30 million	5,398	5,318	80	1.5	61.0	61.2	66.9
¥30 million–¥100 million	11,356	11,195	161	1.4	60.2	63.7	68.2
¥100 million–¥300 million	12,445	12,322	123	1.0	72.6	67.1	78.4
More than ¥300 million	13,455	13,378	77	0.6	77.8	64.0	79.2
Total	42,654	42,213	441	1.0	68.3	65.7	73.9

Note: 'Financial' shows the results of the financial model using balance sheets and financial statements. 'Bank account' shows the results of the bank account model. 'Hybrid' shows the results of the hybrid model using both financial data and bank account data.

Source: Authors' calculations.

The followings are evaluations that can be inferred from the results of the verification.

- The ARs tend to be lowest for the bank account model, followed by the financial model. The ARs for the hybrid model are the highest. The combination of the financial information with the bank account information produces a model that has the strongest capability to detect defaults.
- The ARs of the out-sample data in the financial model are about 6.5 percentage points lower than the ARs of the in-sample data. The ARs of the out-sample data in the bank account model are about 2.2% lower than the ARs of the in-sample data. However, for both models, the ARs of the out-sample data are regarded as high at above 65% in comparison with other studies and results of rating agencies. Considering that the differences in the ARs for the in-sample data and the out-sample are not very large, and the absolute level of the ARs of the out-sample data are high, we can confirm the validity of the model created from the in-sample data.
- For those firms with sales less than ¥300 million, the ARs for the bank account model are higher than the ARs for the financial model. The ARs of the bank account model are highest in the category of companies with sales ranging from ¥100 million to ¥300 million. In addition, the improvements in the ARs by adding the bank account information to the financial information are higher for those categories with lower sales. This shows that the bank account information is useful for smaller firms, which tend to have lower quality financial statements.

For the next step, we derived the Pearson correlation coefficients between the financial model and the bank account model. The results are shown in Tables 13, 14, and 15 for the in-sample data, out-sample data, and the total data. The Pearson correlation coefficients of the financial model and the bank account model are about 51%. This suggests that both models have correlation, but the degree of correlation is not very high. This offers supporting evidence that the bank account model evaluates borrowers from a viewpoint that is different from that of the financial model. As a result, the credit discrimination capability may improve through the hybridization of the financial model and the dynamic model. Additionally, banks will be able to expand their target area by adding the bank account information to the financial information.

Table 13: Pearson Correlations: In-Sample Data

	Pearson Correlations		
	Financial	**Bank Account**	**Hybrid**
Financial model	1.0000	0.5637	0.8377
Bank account model	0.5637	1.0000	0.8882
Hybrid model	0.8377	0.8882	1.0000

Source: Authors' calculations.

Table 14: Pearson Correlations: Out-Sample Data

	Pearson Correlations		
	Financial	**Bank Account**	**Hybrid**
Financial model	1.0000	0.5655	0.8385
Bank account model	0.5655	1.0000	0.8888
Hybrid model	0.8385	0.8888	1.0000

Source: Authors' calculations.

Table 15: Pearson Correlations: Total

	Pearson Correlations		
	Financial	**Bank Account**	**Hybrid**
Financial model	1.0000	0.5646	0.8381
Bank account model	0.5646	1.0000	0.8885
Hybrid model	0.8381	0.8885	1.0000

Source: Authors' calculations.

Overall, the analysis verifies that the accuracy of default prediction improves when a model based on bank account information is used in addition to the default prediction model based on traditional financial information. The analysis shows that the degree of improvement increases when the size of the company is small, and the effect is significant for companies with less than ¥300 million in annual sales. For small companies, the quality of financial data is generally assumed to be low, but the bank account information model can complement the incomplete data. Also, for small firms, the bank account model has a higher default prediction ability than the financial model.

For the hybrid model, which incorporates both the financial information and the bank account information, the main components are (1) the deposit amount versus total sales, (2) the loan amount versus total sales, (3) the indicators showing profitability, (4) the capital ratios, (5) the loan amount compared to deposits, and (6) the increase and decrease of deposits. Among the six major components, four are derived from the bank account information, which offers supporting evidence that the bank account information is critical for default prediction.

6. CONCLUSIONS

This paper demonstrates that the accuracy of default prediction improves when a model based on bank account information is used in addition to the default prediction model based on traditional financial information. There is a tendency for the improvement to increase when the size of the company is small. If the size of the company is small, the quality of financial data is generally assumed to be low, but the bank account information model can complement this.

For small firms, the accuracy of the default estimation of the bank account model is superior to that of the financial model, supporting the possibility that banks will have the ability to determine the credit risks of SMEs even if only the bank account information is used.

If the utilization of the bank account information model spreads, banks can reduce credit costs and review times and costs and make loans to SMEs more efficiently. Bank account information cannot be manipulated by an information provider for the purposes of tax returns or loan applications, as is the case with financial information, so the bank receiving the account information can easily examine the data even if the target company is a new customer. Accordingly, concerns over the credibility of the information and the personnel costs to scrutinize it, which are intrinsic issues for financial scoring loans, are eliminated.

In addition, lending based on bank account information enables a bank to easily calculate the upper lending limit**.** By grasping the annual cash flow of the borrower's account, a bank can estimate the realistic amount a borrower can repay. Considering this estimate, the bank can then determine the loan amount. In contrast, the traditional financial scoring model is based on the financial statements of the previous fiscal year, so simulations of the possible lending amounts do not work well in some transactions. As such, this may lead to excessive lending.

One of the limitations of the bank account model is that the level of information depends on the depth of relations between banks and corporate customers. If a bank has a weak relationship with a firm, the bank account information may not grasp the whole picture of the firm's business activities. However, recently, "cloud accounting firms" [10], which can be classified as fintech companies, are providing automatic accounting services to banks. Cloud accounting systems enable accounting firms to have easy access to bank account information from various institutions. Soon, banks with small shares of deposits will not be significantly disadvantaged if they have tie-ups with those accounting firms or corporate customers.

The use of bank account information can be an effective tool for banks in analyzing the business and financial conditions of their customers and providing effective consulting services. In addition, it can identify commercial distribution and potential business opportunities for clients. The speed and accuracy of the data can also be

[10] Users provide the account information of all their creditor banks on the accounting software (applications) of the cloud firm, making a journal entry for each deposit and withdrawal to streamline their monthly accounting work and account settlement procedures. The users can then make inquiries about their financing directly to financial institutions during the accounting processing. By collaborating with this accounting software, financial institutions can obtain the account information of other banks, subject to the prior consent of the user; that is, banks can potentially obtain information on all of the client's accounts.

useful for grasping the impact of macro shocks, such as currency appreciations or natural disasters.[11]

The empirical analysis in this paper is targeted at SMEs in Japan, but the results may also be relevant for other countries, especially emerging countries in Asia. In Asia, the proportion of companies receiving bank loans is as small as 15.4% for small companies, but 79.4% have bank accounts in banks and use bank accounts for settlements, etc. (ADB 2015). If bank account information can be used systematically, we can expect the expansion of loans for SMEs.

The policy implications of this paper are that it is crucial for financial institutions to enhance their credit risk assessment and improve service quality by leveraging bank account information. It would be more efficient for financial institutions to have a common database and share information rather than developing systems on their own. As a possible solution for other Asian economies, we provide the example of the Risk Data Bank in Japan and show how to create a credit risk model based on financial and bank account information data. This could be important for policy makers for providing good guidance to financial institutions and supporting the development of the credit information system. If the use of bank account information prevails, it could help SME have easy access to finance and enhance growth and productivity.

[11] The case of Kumamoto Bank, located in Kumamoto prefecture in Japan, is an example where bank account information has been used for disaster management. After the Kumamoto earthquake, Kumamoto Bank fully utilized bank account information to identify which areas and sectors were most hit by the earthquake and took proactive remedy measures in collaboration with local governments.

REFERENCES

Anzai, K. 2015. Issues in the Assessment of Small and Medium-sized Enterprises by Financial Institutions and Suggestions for a New Methodology (in Japanese). PhD dissertation. Sakushin Gakuin University.

Asia Development Bank. 2014. *Asia SME Finance Monitor 2013.*

———. 2015. *Asia SME Finance Monitor 2014.*

Behr, P., A. Guttler, and D. Plattner. 2004. Credit Scoring and Relationship Lending: The Case of German SME, University of Frankfurt, Version 16. March 2004.

Financial Service Agency. 2003. Progress of Action Program for Strengthening Relationship-based Banking (as of the first half of FY2003) (in Japanese).

Goto, Y. 2014. *Macro-performance of Small and Medium-sized Enterprises: Explaining Their Contribution to the Japanese Economy* (in Japanese). Nikkei Publishing Inc.

Grunert, J., L. Norden, and M. Weber. 2005. The Role of Non-financial Factors in Internal Credit Ratings. *The Journal of Banking and Finance,* 29(2): 509–531.

Hergen, F., and M. Wahrenburg. 2003. Evaluating Internal Credit Rating Systems Depending on Bank Size. Working Paper, Goethe University Finance & Accounting No. 115.

Hirata, H. 2005. Recent Development in Japan's Small and Medium-sized Enterprise Finance (in Japanese), *Keiei Shirin, Hosei Journal of Business,* 41(2): 31–51.

Itou, M. and N. Yamori. 2015. Revitalization of Regional Banks: Toward Contribution Banking (in Japanese). Kinyu Zaisei Jijo Kenkyu Kai.

Maehara, Y. 2013. The Role of the Credit Risk Database in SME Finance (in Japanese). RIETI Discussion Paper Series No.13-J-067. RIETI.

Okubo, Y. 2015. Complete Overhaul of Spread Banking (in Japanese). Kinyu Zaisei Jijo Kenkyu Kai.

Okubo, Y, and T. Inaba. 2007. The Era of Credit Rating on Small and Medium-sized Enterprises: Benefits of 'Japan SME Credit Scoring' Exclusively for SMEs (in Japanese). Kinyu Zaisei Jijo Kenkyu Kai.

Simizu, S. 2016. Structural Reforms of Economy and Maintenance of Financial Infrastructure Required for Asian Economies (in Japanese). *RIM Pacific Business and Industries* 16(63).

Small and Medium Enterprise Agency. 2015. White Paper on Small and Medium Enterprise 2015.

Watanabe, T., and I. Uesugi, eds. 2008. *Examining SME Finance* (in Japanese). Nikkei Publishing Inc.

Yamashita, S., N. Kawaguchi, and T. Tsuruga. 2003. Consideration and Comparison of Evaluation Methods for the Credit Risk Model (in Japanese). *FSA Financial Research and Training Center Discussion Paper Series 2003*.

Yamori, N. 2003. Financial Services Desirable for Enterprises and Issues in SME Finance: Focusing on the Annual Survey of Corporate Behavior Regarding Corporate Finance in the Kansai Region (in Japanese). *RIETI Discussion Paper Series,* No.06-J-003.

———. 2018. *The Role of Regional Financial Institutions in Regional Revitalization: Towards an Improvement in the Quality of Financial Intermediation* (in Japanese). Chuo-Keizai Group Publishing.

Yanagaisawa, K., H. Shimoda, E. Okdada, N. Simizu, and M. Noguchi. 2007. An Analysis on the Historical Change of Explanatory Power of Credit Risk Model based on the RDB Database (in Japanese). *Japanese Association of Financial Econometrics and Engineering, Proceedings of 2007 Summer Conference*: 249–263.

Yoshino, N., and F. Taghizadeh-Hesary. 2014. Analytical Framework on Credit Risks for Financing Small and Medium-sized Enterprises in Asia. *Asia-Pacific Development Journal* 21(2): 1–22.

Yoshino, N., F. Taghizadeh-Hesary, P. Charoensivakorn and B. Niraula. 2016. Small and Medium-sized Enterprise Credit Risk Analysis Using Bank Lending Data: An Analysis of Thai SMEs. *Journal of Comparative Asian Development* 15(3): 383–406.

Yoshino, N., and F. Taghizadeh-Hesary, eds. 2016. *Japan's Lost Decade: Lessons for Asian Economy.* Springer.

Yoshino, N., and H. Yamagami. 2017. *Monetary Economics: Practice and Theory* (in Japanese). Keio University Press.

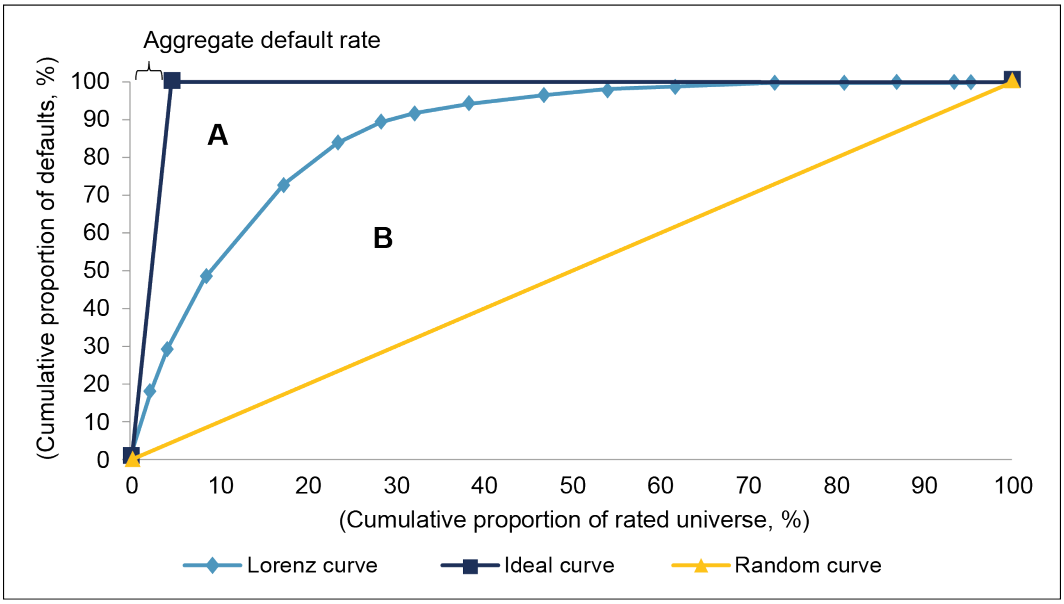
APPENDIX: ACCURACY RATIO

The accuracy ratio (AR) is a summary of the quantitative measure of the discriminatory power in classification models, e.g., credit scoring models. The AR measure expresses the ratio of the area above and under the power curve (cumulative accuracy profile of the model under consideration versus the "perfectly" discriminating models).

An AR can take a value between 0 and 1. The closer AR is to 1 (100%), the larger the excess surface covered by the CAP curve, and the higher the discriminative power of the classification system.

The accuracy ratio is sometimes also denoted as the Gini coefficient. The procedure for calculating the Gini coefficient is illustrated below. Area B is bounded by the random curve and the Lorenz curve, while area A is bounded by the Lorenz curve and the ideal curve. The Gini coefficient (AR) is defined as area B divided by the total of areas A plus B.

Sample Lorenz Curve



Source: S&P Global Fixed Income Research.