Understanding the Utilization of Credit Guarantee System

(Summary)

This paper conducted an analysis of the utilization of the Credit Guarantee System using loan-level data collected through the Common Data Platform. A machine learning approach was employed to identify the key factors influencing whether a loan is guaranteed. Among borrower-related factors, sales and capital ratio were found to have relatively large effects—borrowers with higher values for these indicators were less likely to utilize credit guarantees. While this analysis does not aim to assess the appropriateness of credit guarantee usage—given that such usage varies depending on borrower characteristics and other various factors—it did reveal that the tendency to use guarantees differs significantly based on whether a borrower is in excess liabilities, and difference across industries are also observed.

I. Introduction

This paper conducted an analysis using granular data from the Common Data Platform to better understand the actual usage of credit guarantees provided by Credit Guarantee System[1]—an important scheme that supports smooth financing for small and medium-sized enterprises (SMEs). As shown in Figures 1 and 2, the utilization of guarantees varies depending on the borrower's industry and financial condition. However, it should be noted that these figures also include large enterprises and overseas entities, which are outside the scope of the Credit Guarantee System.

[1] While other forms of guarantees exist—such as personal guarantees by business owners or guarantees by parent companies—this paper does not address those. Unless otherwise specified, the term "(credit) guarantee" refers solely to credit guarantees provided by Credit Guarantee Corporations.

Figure 1: Guaranteed ratio by type of industry

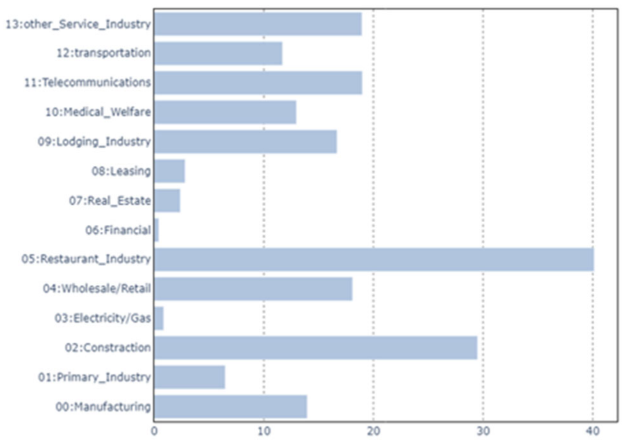
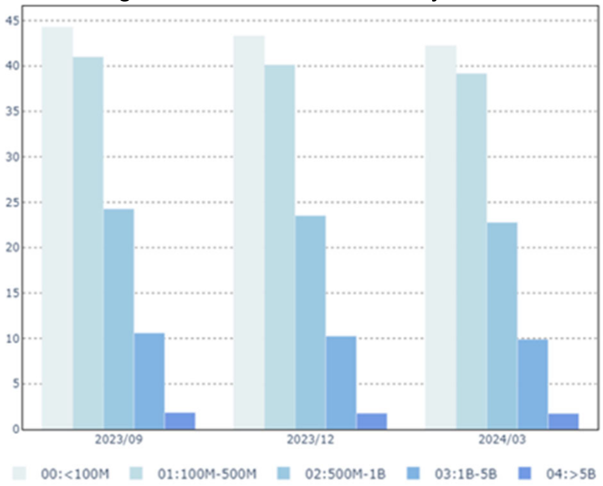


Figure 2: Guaranteed ratio by sales



(Note) Some industry categories such as finance and real estate are excluded. Borrowers with 0 sales are also excluded.
(Source) FSA Analytical Notes (2025.1) vol.2: Analysis of Borrower Classifications Assigned to Shared Borrowers, "Box: Utilization of Credit Guarantee System"

This paper conducts a quantitative analysis to identify the factors influencing the use of credit guarantees and associated trends, based on a more precisely defined set of conditions. It should be noted that the analysis presented here does not intend to assess the appropriateness of such usage—given that credit guarantees are utilized based on a comprehensive assessment of borrower characteristics and banks' credit policies.

This analysis utilizes loan-level data from 62 member banks of the Regional Banks Association of Japan (the number of banks is as of the end-December 2024), specifically for the reference dates at the end of September 2023, December 2023, March 2024, June 2024, and September 2024. Given that the Credit Guarantee System is intended for designated types of small and medium-sized domestic enterprises in specific industries, the scope of analysis is limited to firms that meet the criteria outlined in Figure 3, thereby approximating the eligible population for credit guarantees[2].

[2] In addition to capital size and industry category, eligibility for the Credit Guarantee System is also subject to conditions such as the number of employees and detailed industry subcategories. However, due to data limitations, these criteria were not incorporated into the analysis. Furthermore, records with missing values in key fields related to credit guarantee usage were excluded from the dataset.

Figure 3: Scope of analysis

	Description
Region	Domestic firms
Firm size	Small and medium-sized enterprises, including sole proprietors; however, entities such as other types of corporations and firms with capital of 300 million yen or more are excluded.
Industry	Manufacturing, Construction, Wholesale & Retail, Food services, Accommodations, Healthcare & Welfare, Information & Communication, Transportation, Other services

II. Key feature analysis by machine learning

This section constructs a classification model using machine learning to predict the presence or absence of credit guarantees[3] based on various features, including bank characteristics and borrower attributes. The contribution of each feature was then examined to identify factors that influence the use of credit guarantees. The analysis employs CatBoost, a type of gradient boosting method based on decision trees. The loan data used as input for the machine learning model is limited to discounted bills or deed loans with an outstanding balance of at least 1 million yen.

1. Dataset

The features used in the machine learning model are listed in Figure 4. For model development, 75% of all records were randomly selected as training data, while the remaining 25% were used as test data to evaluate model performance. Since the Credit Guarantee System has a limit on the guaranteed amount, whether a newly originated loan receives a guarantee may be influenced by the amount already guaranteed to the borrower. To eliminate this effect, when multiple loan records existed for the same reference date, the loan with the earliest transaction start date was selected. Furthermore, to capture recent trends, only loans whose transaction start date was within three months of the reference date were included (sample size: N = 64,845). It should be noted that even when the model was trained on all newly originated loans as of the reference date without such

[3] This analysis is based on a binary classification of whether a credit guarantee is present or absence, and does not take into account variations in the guarantee coverage ratio—such as whether the guarantee covers 80% or 100% of the loan.

restrictions, the overall accuracy declined, but the top-ranking features in terms of importance remained broadly consistent.

Figure 4: List of features[4]

Category	Feature	Description
Bank attributes	Bank name	Name of the lender bank
	FI-HHI	HHI of the loan by financial institutions (including credit associations and credit cooperatives) in each prefecture
Loan attributes	Loan amount	Outstanding loan amount
	Loan maturity	Number of days from the transaction start date to the final repayment date
	Loan type	Discounted bills / deed loans
Borrower attributes	Core company flag	Set to "1" if the borrower has a separate parent company or main operating entity, "0" if otherwise
	Newly founded flag	3 categories: Whether the date of incorporation is less than three 3 years/3 years or more but less than five 5 years/5 years or more from the record date
	Cross-border flag	Set to "1" if the borrower prefecture and lender prefecture is different, "0" if otherwise.
	Industry	Manufaturing / Construction / Wholesale & Retail / Foods services / Accomodations / Healthcare & Welfare / Information & Communication / Transportation / Other services
	Prefecture dummy	47 prefectures
Borrower financial information	Sales	Sales revenue
	Capital ratio	= 100 * capital / total assets
	Operating profit ratio	= 100 * operating profit / sales
	Current profit	Current net profit
	ICR	= (operationg profit + interest income + divident) / interest and discounts expenses
	ROA	= 100 * oridinary profit / total assets
	Cash and deposits over debt ratio	= 100 * cash and deposits / debt
	Cash and deposits ratio	= 100 * cash and deposits / current liabilities
	Cash and cash equivalent ratio	= 100 * cash and cash equivalents / current liabilities
	Current ratio	= 100 * current assets / current liabilities
	Fixed ratio	= 100 * fixed assets / capital

(Ref) Number of samples in Train Data and Test Data

	Gurantee	23/09	23/12	24/03	24/06	24/09
Train	Absence	6,106	5,215	5,525	4,047	5,959
	Presence	4,634	4,343	4,582	3,669	4,553
Test	Absence	2,093	1,726	1,834	1,387	1,911
	Presence	1,492	1,504	1,544	1,245	1,476

[4] The Herfindahl-Hirschman Index (HHI), an indicator used to assess the level of market competition, is calculated here as the sum of the squared loan amount shares held by all financial institutions within the prefecture where the head offices of the financial institutions located. The “newly founded flag” is based on the basic information of entities granted corporate numbers, as published by the National Tax Agency. The variable “prefecture dummy” accounts for potential differences in credit guarantee policies—such as the criteria for issuing guarantees and the extent of coverage—across prefectural Credit Guarantee Corporations, as well as varying degrees of utilization of such guarantees depending on the region.

2. Result and implications

The performance[5] of the model was first evaluated, and as shown in Figure 5, all evaluation metrics indicate a sufficiently high level of performance. This suggests that the features used in the model provide a reasonable explanation for the presence or absence of credit guarantees.

Next, both the SHAP values[6] and feature importance[7] shown in Figures 6 and 7, respectively, indicate that loan maturity and loan amount have a relatively significant impact on the presence or absence of credit guarantees. According to the SHAP values in Figure 6, longer loan maturities tend to have positive SHAP values—contributing to the likelihood of a guarantee—whereas shorter maturities tend to yield negative SHAP values, indicating a stronger association with loans without guarantees. This is consistent with the general understanding that longer-term loans carry greater risk.

In contrast, loan amount exhibits the opposite pattern: smaller balances are associated with positive SHAP values (i.e., more likely to be guaranteed), while larger balances show negative SHAP values (i.e., more likely not to be guaranteed). This suggests that the loan amount may act as a proxy for firm size, implying that borrowers with smaller balances tend to be smaller in scale and generally possess lower creditworthiness. Additionally, as shown in the feature importance in Figure 7, the bank name ranked second in terms of contribution. This indicates that beyond loan attributes and borrower attributes, lender-specific factors—such as credit assessment policies—also have a material impact on whether or not a loan is guaranteed.

In addition, it was confirmed that borrower financial information such as sales and capital ratio also exert a certain level of influence. The next section provides a more detailed examination of the

[5] The ROC curve is a graph that plots the true positive rate (TPR = true positives / [true positives + false negatives]) on the vertical axis against the false positive rate (FPR = false positives / [false positives + true negatives]) on the horizontal axis. It illustrates how the TPR and FPR vary as the classification threshold of the model is adjusted. The AUC represents the area under the ROC curve, where a higher value indicates better predictive performance. An AUC of 1 signifies perfect prediction accuracy, while a value of 0.5 corresponds to a completely random model.

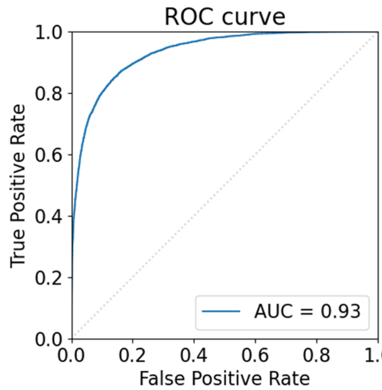
[6] The concept of SHAP (SHapley Additive exPlanations) values is based on calculating how each feature contributes to a prediction by sequentially adding features to the model. Specifically, it begins by computing the prediction when only feature *x* is included, then recalculates after adding feature *y*, and so on—assessing how the prediction changes as each feature is added. SHAP values apply the Shapley value concept from cooperative game theory to machine learning. Because the contribution of each feature can vary depending on the order in which features are added, SHAP values are calculated as the average contribution across all possible feature orderings. For more details, refer to sources such as the Bank of Japan's Working Paper Series *"Application of Machine Learning to a Credit Rating Classification Model: Techniques for Improving the Explainability of Machine Learning" (March 2023)*. Figure 6 visualizes the degree of contribution each feature makes to the actual prediction. In the figure, features represented in color use a scale where red indicates a high value of that feature, and the farther to the right the point lies, the more it contributes to a "presence" of guarantee classification; the farther to the left, the more it contributes to a "absence" of guarantee classification. Features shown in gray are categorical variables for which numeric ordering is not applicable.

[7] Feature importance refers to a numerical score that indicates the relative significance of each feature in building a machine learning model. In general, the more frequently a feature is used for decision splits (e.g., in decision trees), the higher its importance score tends to be. These scores are normalized so that the total across all features sums to 100.

relationship between financial condition and the presence or absence of credit guarantees.

Figure 5: Model performance

Indicator	Description	Score
Accuracy	The proportion of correct predictions made by the model across all data	0.856
Precision	The proportion of actually guaranteed data among data predicted as "presence" of the guarantee by the model	0.842
Recall	The proportion of data predicted as "presence" of the guarantee by the model among actually guaranteed data	0.833
F1-score	Harmonic mean of Precision and Recall = 2 ×（precision×recall）/（precision＋recall）	0.838



*Scores are calculated from Test Data

Figure 6: SHAP values

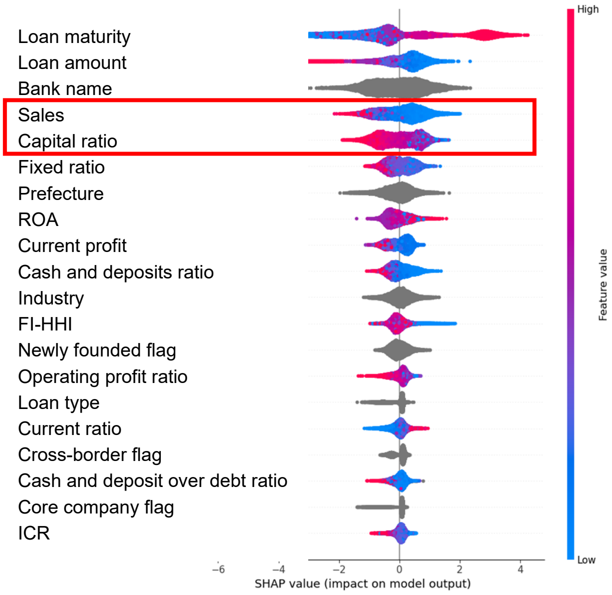
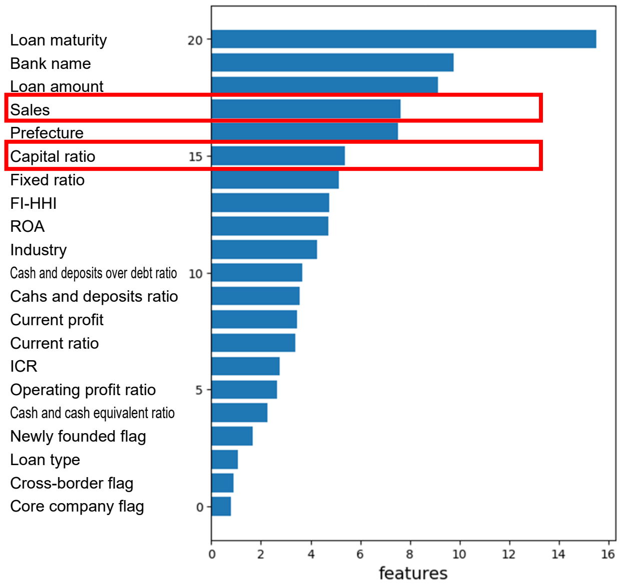


Figure 7: Feature Importance



III. Relationship between borrower financial condition and utilization of Credit Guarantee System

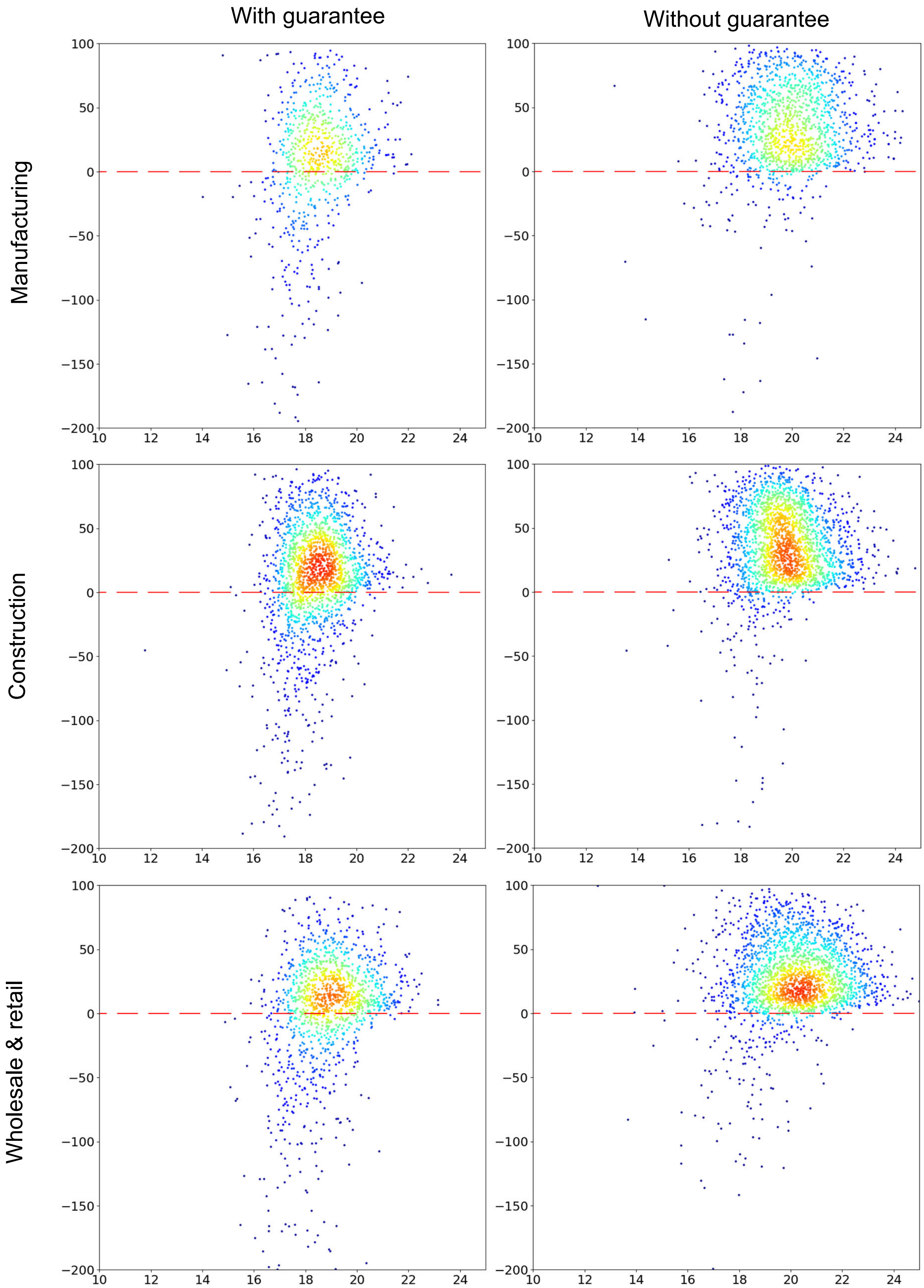
Among the financial information of borrowers, the previous section confirmed that “sales” and “capital ratio” have a certain degree of influence on the presence or absence of credit guarantees. This section examines the relationship between these indicators and the use of guarantees in more detail by borrower industry. The analysis is limited to data as of the end-September 2024, and graphs are presented only for industries with sufficient data: manufacturing, construction, wholesale & retail, food services, and transportation.[8]

1. Distribution of sales and capital ratio

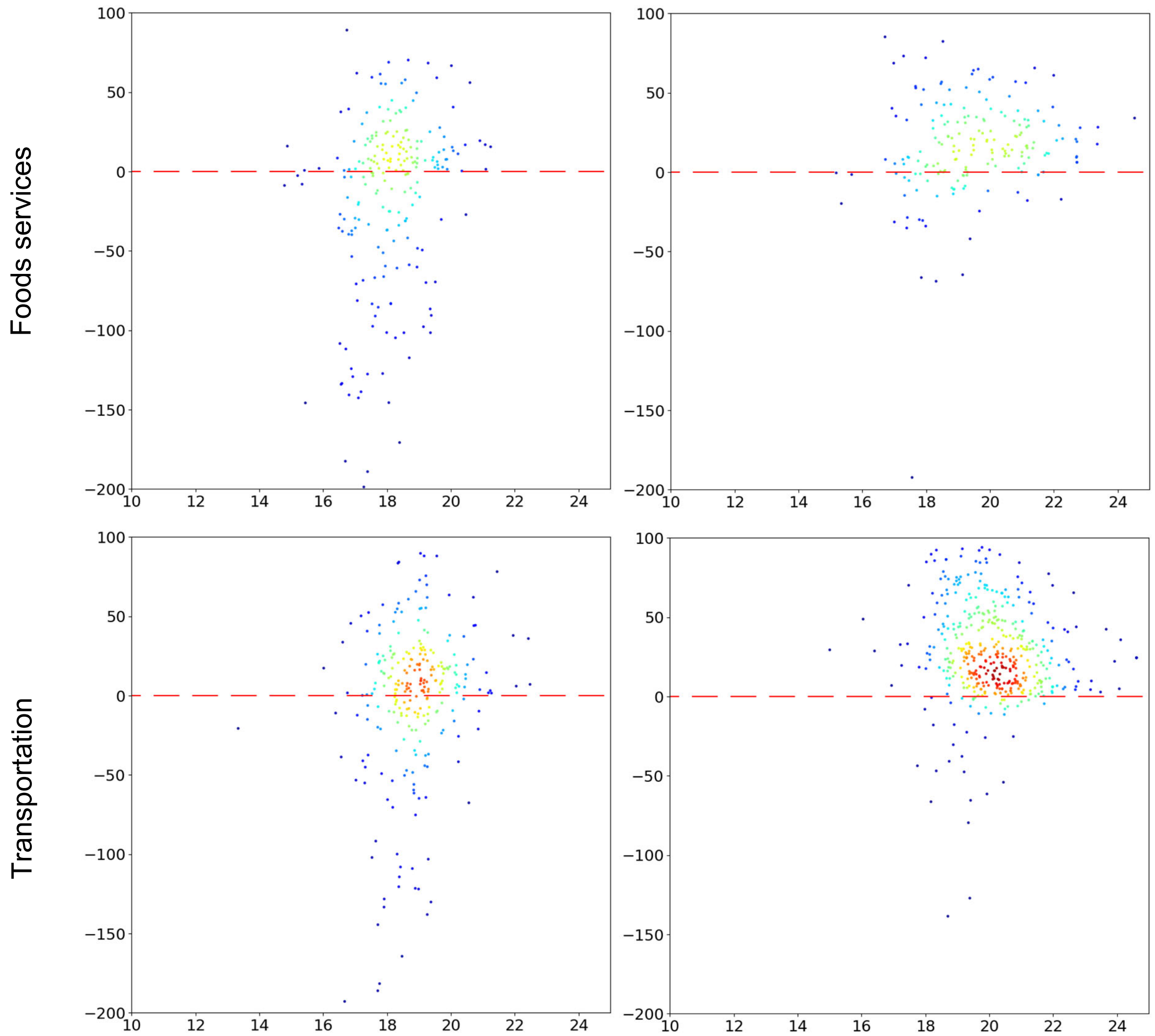
Figure 8 shows the relationship between “sales” and “capital ratio” of each borrower in a scatter plot. The red dashed line indicates a capital ratio of 0%, with points below this line representing debt-overhang firms. Across all industries, the distribution of loans without guarantees drops sharply below the red dashed line, indicating that loans without guarantees are rarely extended to firms with negative equity. Regarding sales, the distribution for loans without guarantees generally appears further to the right, suggesting that borrowers without guarantees tend to have higher sales than those with guarantees.

[8] The accommodation, healthcare & welfare, and information & communications sectors are excluded from this analysis due to an insufficient number of data points. In addition, the "other services" category is not illustrated, as it encompasses a wide variety of industry types.

Figure 8[9] Sales and capital ratio



[9] The X-axis represents the logarithmic value of sales (excluding cases where sales are zero), while the Y-axis represents the capital ratio of the borrower. The degree of concentration is visualized using a color scale based on kernel density estimation (approximated using a Gaussian distribution in this case), with areas of higher density shown in red. The color mapping, based on the output values of the estimated function, is applied using a consistent scale across all panels.

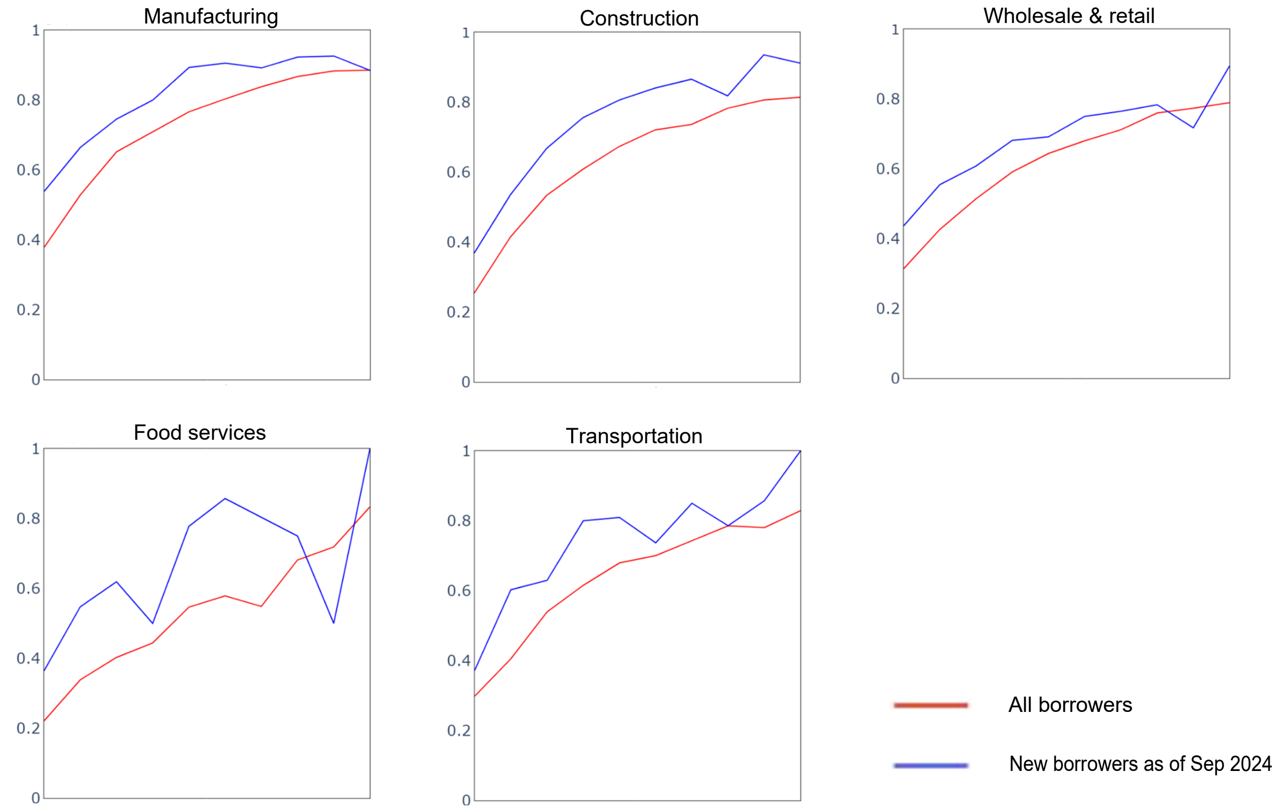


2. Relationship between sales and guarantee

This section provides a more detailed examination of the utilization of guarantees by borrowers, categorized by sales volume[10]. In the following analysis, borrowers for whom the outstanding balance of guaranteed loans accounts for less than 50% of their total loan balance are defined as "non-guarantee-primary borrowers[11]." In Figure 9, the horizontal axis shows the sales categories[12] of borrowers, while the vertical axis indicates the proportion of non-guarantee-primary borrowers.

[10] It should be noted that the numerical data for sales and capital ratios are based on records as of the end of September 2024, and may differ slightly from the financial information available at the time of lending. In addition, to exclude the effects of effectively interest-free and unsecured loans offered by private financial institutions, the analysis is limited to loans executed from 2022 onward and is aggregated at the debtor level.
[11] In this analysis, whether credit guarantees are considered the primary source is determined based on a 50% threshold. However, in practice, approximately 80% of debtors fall into either of the two extremes—entirely without guarantees (0%) or fully guaranteed (100%).
[12] Divided into segments of 100 million yen each, ranging from 0 to less than 1 billion yen.

Figure 9: Sales (X-axis) and proportion of non-guarantee-primary borrowers (Y-axis)



The red line represents the full dataset used (all borrowers in the scope) in this analysis. Across all industries, the proportion of non-guarantee-primary borrowers tends to be lower for borrowers with smaller sales volumes. However, there was no notable trend indicating a sharp increase at a specific sales threshold. This suggests that the presence or absence of credit guarantees is not uniformly determined based on a fixed level of sales.

To examine the impact of whether a borrower is a new client on the usage of guarantee, the figure also plots in blue the data for borrowers who were new as of end-September 2024[13]. While their behavior generally mirrors that of all borrowers shown in red, the proportion of new borrowers tends to be higher across all industries[14] and sales levels. This may suggest that, prior to decisions on whether to apply guarantees, new borrowers are more likely to be selectively granted loans only if they are financially sound, or that when loans are newly extended to borrowers already served by a main bank, such loans are more likely to be unguaranteed. By industry, the proportion of non-guarantee-primary borrowers is particularly higher in manufacturing for borrowers with smaller sales volumes, and this trend is especially pronounced among new borrowers.

[13] New borrowers as of end-September 2024 refers to borrowers who did not appear in the datasets as of the reference dates of September 2023, December 2023, March 2024, or June 2024, but are present in the dataset as of September 2024.

[14] For the food service and transportation sectors, it should be noted that the number of observations for new borrowers as of end-September 2024 in the sales range of 500 million to 1 billion yen is limited.

3. Relationship between capital ratio and guarantee

Next, Figure 10 plots the proportion of non-guarantee-primary borrowers by borrowers' capital ratio bracket[15]. The green dashed line represents a capital ratio of 0%, with the left side indicating negative equity (excess liabilities) and the right side indicating positive equity (excess assets).

Figure 10: Capital ratio (X-axis) and proportion of non-guarantee-primary borrowers (Y-axis)

Focusing on the red line representing the all borrowers, a consistent trend can be observed across all industries: the proportion of non-guarantee-primary borrowers increases significantly around the 0% capital ratio threshold. On the other hand, when the capital ratio is negative, the proportion of non-guarantee-primary borrowers remains relatively unchanged, even as the absolute value of the negative capital increases. Industry-specific patterns also emerge—construction and food service sectors generally show a lower proportion of non-guarantee-primary borrowers, likely because they tend to consist of smaller firms that make greater use of credit guarantees. In contrast, the

[15] Divided into 5% increments within the range of -50% to 50%.

manufacturing sector shows a relatively high proportion of non-guarantee-primary borrowers, even among borrowers with negative capital.

Additionally, the blue line represents borrowers whose sales are 100 million yen or less. Compared to the red line (which reflects all borrowers), the blue line shows no significant shift around the 0% capital ratio threshold. Across all industries, the proportion of non-guarantee-primary borrowers increases gradually and linearly. This suggests that for smaller-scale borrowers, credit guarantees tend to be utilized even if the firm is not in a state of excess debt as of the reference date.

IV. Conclusion

This paper utilized loan-level data obtained through the Common Data Platform to identify the key factors influencing the presence of credit guarantees using machine learning. In addition, a detailed analysis was conducted on the usage of credit guarantee by various attributes such as borrowers' size and industry. Among borrower-specific factors, sales and capital ratios were found to have relatively significant effects on guarantee usage. While no clear threshold in sales levels was observed that would significantly alter the share of non-guarantee-primary borrowers, a notable shift was seen around the zero percent threshold for the capital ratio, indicating a substantial difference depending on whether the borrower was in a state of excess debt. Differences were also observed across industries, with sectors such as construction and food services showing relatively low shares of non-guarantee-primary borrowers. This may reflect the prevalence of smaller firms in these sectors, which are more likely to utilize credit guarantees.

However, it should be noted that due to data limitations, the scope of this analysis was restricted to a short-term dataset from regional banks, beginning in the September 2023 period. Going forward, as more data becomes available, efforts will be made to enhance the sophistication of the analysis while continuing to monitor guarantee trends and deepen the understanding of the actual conditions.