Factors Influencing Loan Default: An Empirical Analysis Based on Microscopic Evidence

Jingwen Xu

The Affiliated International School of Shenzhen University, Shenzhen, China
Email: xjwen200733@outlook.com (J.W.X.)

Manuscript received November 18, 2024; revised December 17, 2024; accepted January 9, 2025; published January 20, 2025.

***Abstract*—With the rapid development of the loan industry, issues such as loan defaults have become hot topics of social concern. This article constructs a research model on the influencing factors of loan default from three dimensions: personal characteristics, credit characteristics, and loan characteristics. Using logit and other models, the predictive ability of different factors on loan default was tested on a massive dataset. The research results indicate that annualization and employment length negatively affect the probability of loan default, debt to income ratio, loan terms, loan amount, and loan interest rate positively affect the occurrence of loan default. There is an inverted U-shaped relationship between the homeownership and loan default. Moreover, we also divided the dataset for loan purposes and found differences in the influencing factors between real estate loans and non-real estate loans. The research conclusion helps to regulate lending behavior and has a good reference value for financial risk prevention and control.**

***Keywords*—credit attribute, default risk, loan attribute, personal attribute**

I. Introduction

Loans have a significant impact on economic growth. As of 2023, the banking giants at Goldman Sachs reported a staggering total lending amount of $7.6 trillion in the United States, underscoring the pivotal role of loans in fueling GDP expansion. However, this surge in lending activity is mirrored by a concerning rise in loan defaults, reaching levels unprecedented in recent years. The outbreak of financial risks associated with loan defaults has the potential to disrupt credit stability, affecting not only the financial sector but also the broader national economy and the livelihoods of its citizens. This backdrop has brought to the forefront the pressing need to identify and understand the factors influencing loan default, drawing extensive attention from various quarters, including academia, policy-makers, and the financial industry.

The financial landscape is characterized by a rich and diverse array of data (Chen *et al.*, 2018; Covin & Slevin, 1989). However, it is plagued by the challenge of information overload, making it a formidable task to distill valuable insights from this complexity. A growing body of research suggests that personal creditworthiness stands as a critical reference point, with individuals possessing adverse credit attributes found to be at higher risk of loan default (Gao *et al.*, 2022; Lee *et al.*, 2021; Park *et al.*, 2022). Building upon existing literature and research findings, this study embarks on a multidimensional exploration of the influences on loan defaults, categorizing them into three core dimensions: personal characteristics, credit attributes, and loan attributes. By doing so, we endeavor to provide a comprehensive understanding of these intricate dynamics, focusing on empirical evidence and analysis. We proposed three hypotheses during the research process and validated them using microscopic datasets. What's more, we also conducted a more detailed classification analysis of different types of loans.

The research may be of great significance in understanding loan defaults and forecasting. This paper seeks to provide an in-depth understanding of the complex dynamics behind loan defaults by comprehensively examining the multifaceted influences on this phenomenon. Through empirical analysis based on micro-level evidence, it is aspired to enhance the understanding of how to predict and manage loan default risks effectively.

The rest of the paper is organized as follows. In Section III, this paper covers the related theoretical background and provides a review of prior literature in this area. It proposed the theoretical framework and developed hypotheses. In Section IV, this paper introduces the empirical data. It conducted empirical tests on the impact of personal attributes, credit attributes, and loan attributes on loan defaults. Moreover, the differences in the influencing factors of different loan types. In Section V, this paper summarizes and analyzes the empirical results. Then, it proposed the theoretical and practical implications of the research. In Section VI, this paper summarized the gains and growth in the writing process.

II. Literature Review and Hypothesis

A. Personal Attribute Impacts on Loan Default

In the research on factors affecting loan defaults, personal attributes have attracted the attention of many scholars. Henager and Wilmarth (2018) used data from the national financial capacity research database, the influencing factors of academic education on loan health were analyzed, and positive and significant results were obtained. Nalić and Švraka (2018) discussed the credit evaluation methods for borrowers at the level of personal factors such as age and gender and demonstrated good accuracy and stability in practical applications. The willingness of young people to default on loans is significantly influenced by their financial literacy, materialism, and risk perception, but is almost unaffected by emotions or personal debt (Thomas *et al.*, 2023). The research results have reference significance for understanding the lending behavior of young people, especially for in-depth exploration of lending issues in the real estate market. Some scholars have also conducted research on the probability of loan default from the perspective of individual personality traits and found that the information left by defaulting borrowers in loan applications, such as loan intentions, emotional tendencies, and family

doi:10.18178/joebm.2025.13.1.841

characteristics, can help determine the borrower's default propensity (Netzer *et al.*, 2019). Jung and Kim (2020) studied the impact of lender work conditions on loan default rates.

The results indicate that compared to non-self-employed individuals, the loan default rate of self-employed individuals is significantly influenced by their work conditions. In recent years, student loan defaults have also been an important issue of social and academic concern, with a large number of students defaulting on loans due to personal factors. Cox *et al.* (2020) found that issues such as interest penalties caused by defaulting on the minimum repayment amount increase the probability of students defaulting. These burdens weaken individuals' repayment ability, reduce their future credit opportunities, and also lead to an increase in bank financial risks. Distance factors, education levels, and monthly budget availability have a decisive impact on loan defaults, but income levels and gender have almost no impact (Chong, 2021). Based on the data of 1 million personal loans issued by Lending Club from 2007 to 2018, it was found that the probability of default can be determined by housing ownership, employment status, and other factors when there are no obvious issues with the borrower's obvious risk characteristics, loan characteristics, and local economic factors (Croux *et al.*, 2020). Then, the recent increase in the number of severely delinquent accounts can be attributed to changes in the age distribution of borrowers, and the ratio of debt to income has a significant impact on whether lenders will default. Duarte *et al.* (2018) found that the area where the lender resides can also have a certain impact on the probability of loan default in some cases.

Therefore, the attributes of personal attributes are an important basis for predicting default behavior, and scholars generally believe that personal attributes can effectively prevent foreseeable risks through the selection of different characteristics (Gapko & Smíd, 2019; Ju & Sohn, 2017). Therefore, we propose the following assumptions:

H1: The higher a personal's annual income and employment length, the lower the risk of loan default.

B. *Credit Attribute Impacts on Loan Default*

Credit refers to the total amount of personal credit established based on credit, which refers to social records based on individual manifestations and history (Ballester *et al.*, 2020). Due to the foreseeable solvency and willingness to repay, credit provides corresponding trust and evaluation, enabling the citizen to conveniently obtain financial, material, and other economic support in their economic life., It is the foundation of the credit of the entire society.

Credit attributes are important reference indicators for financial transactions such as loans. Research has shown that lenders with non-performing credit attributes can significantly increase loan default rates, posing significant risks to the financial industry (Looney & Yannelis, 2022). There are also studies indicating that lenders attach great importance to their credit status. If they are informed before the loan is made that the loan will be recorded in the credit system, the default rate will significantly decrease (Liao *et al.*, 2023). De Giorgi *et al.* (2023) found that an additional credit limit would result in a 5.9 percentage point decrease in the default rate of high-scoring borrowers in previous loans. However, for borrowers with lower scores, it will increase by 19 percentage points. The former uses new credit to smoothly pay off existing loans, while the latter increases their total debt. Therefore, adopting differentiated loan strategies based on different credit attributes is an important means to reduce the risk of loan default. Hard financial information such as credit scores and property ownership may lead to default, and there is a negative correlation between credit scores and the likelihood of default (Ravina, 2019). There is a negative correlation between collateral, credit score, and default. A low credit score increases the likelihood of default, while a high credit score reduces the probability of default (Duarte *et al.*, 2018). After maintaining the same credit score, credit history, income, employment status, and property ownership, personal attributes have a significant impact on the likelihood of obtaining funds and loan terms (Ravina, 2019). Hard financial information such as income and employment status can have an impact on the likelihood of default. There is a negative correlation between default and the likelihood of default. Borrowers without stable jobs are more likely to default.

A loan evaluation model based on personal credit attribute is an effective risk prevention and control method. Nalić and Švraka (2018) established a credit rating system that evaluates the feasibility of loans based on personal history and current credit information. Li *et al.* (2022) established a default prediction model based on information such as the borrower's historical loan limit, account balance, and credit rating when data is unbalanced. The research results can provide reference for reducing the external risk of default on online lending platforms. Ma (2019) used the XGBoost model to explore the impact of credit attributes such as financial indicators on individuals' willingness to repay credit cards. Some scholars have also attempted to identify the false credit attributes of lenders, and research has found that loan officials may also be involved due to monetary incentives. These false credit attributes are more frequent at the end of the year, leading to a decrease in bank profitability, resulting in a decrease of approximately 1.5 percentage points in the return on equity (Berg *et al.*, 2020). Some scholars have also explored whether extending loan terms can alleviate default risks (Lu & Wang, 2012). In recent research, in order to address the sensitive dependence of small lenders on financial risks, a credit evaluation model based on spatial random effects was developed, which can reduce the increase in loan default rates caused by changes in the financial environment (Medina-Olivares *et al.*, 2022).

In summary, the credit attributes of individual loans are important factors in exploring loan default risk issues (Li *et al.*, 2023; Luong & Scheule, 2022; Lyócsa *et al.*, 2022). The health of individual credit can effectively reveal the possibility of fulfilling repayment obligations in the future. Therefore, we propose the following assumptions:

H2: The higher homeownership and the lower their debt-to-income ratio, the lower the risk of loan default.

C. *Loan Attribute Impacts on Loan Default*

Loans have different attributes, and the attributes and treaties of loans can measure the safety of loans, the difficulty of payment, and performance of loans. The main attributes of a loan include the number of loan terms, loan amount, loan interest rate, and whether there is a guarantee. Different scholars have explored various dimensions of loan attributes for loan default problems in different scenarios, to predict and judge the degree of risk of loan default.

The study of loan attributes and default risk is a hot topic in fields such as finance. Based on the data of 1 million personal loans issued by Lending Club from 2007 to 2018, it was found that the probability of default can be determined by the loan term, loan purpose, and other factors when there are no obvious issues with the borrower's obvious risk attributes, loan attributes, and local economic factors (Croux *et al.*, 2020). Meanwhile, collateral has a positive impact on default, but collateral has a negative impact on default (Duarte *et al.*, 2018). Some scholars have also found that the urgency of short-term loans is an important indicator for predicting default. At the same time, applying text mining methods to feature mining the text information described by borrowers can help improve the accuracy of the prediction model (Liu *et al.*, 2021). It is worth noting that there may be differences in loan default issues under different market scenarios, and loan attributes need to be chosen based on specific cultural and economic backgrounds (Saha *et al.*, 2022). With the deepening of theoretical research and the development of computer technology, some scholars have attempted to combine existing theoretical guidance with advanced technology to develop intelligent models that can automatically identify loan default risks. Features were extracted from loan-related information using the transferor method, and the reliability of the model was tested on real US market datasets (Zhang *et al.*, 2020). The algorithm achieved optimal performance under AUC and G-means indicators. A corresponding network loan default risk assessment model was established using Backpropagation Neural Networks (BPNN), and the network loan default assessment model of the BPNN model was simulated (Li, 2022). The model was compared with support vector machines and regression models. The experimental results show that the BPNN model based on loan features and other data performs better on multiple indicators than support vector machines and regression models. Further demonstrates the applicability of deep learning in predicting default risk. At the same time, finding more predictive features that can be input into neural networks remains a hot topic for scholars in related fields.

Overall, loan attributes are an important feature dimension of traditional econometric theory models and cutting-edge deep learning models in this field. Loan default issues are significantly influenced by various loan attributes (Do *et al.*, 2018; Lin *et al.*, 2011). Therefore, we propose the following assumptions:

H3: The shorter the term of the loan, the smaller the amount, the lower the interest rate, and the lower the risk of loan default.

D. Research Model on Factors Influencing Loan Default Risk

Overall, the factors that affect loan default risk can be mainly divided into credit characteristics, loan characteristics, and personal attributes. This article summarizes the relevant literature as follows, as detailed in Table 1. Among them, Lu and Wang (2012) only involve the study of credit and loan features, while Liu *et al.* (2021) involves the extraction and analysis of loan features in quantitative research. It can be seen that there has been a lot of research on credit characteristics, loan characteristics, and personal attributes in existing literature, but there is a lack of unified and comprehensive analysis. This study comprehensively considers credit characteristics, loan characteristics, and personal attributes, and conducts data analysis through a combination of quantitative and qualitative methods. Based on the data results, the problem of loan default has been deeply explored.

Table 1. Summary of related literature review

Paper	Method	Personal attribute	Credit attribute	Loan attribute
Lu and Wang (2012)	Quantitive		√	√
Looney and Yannelis (2022)	Quantitive	√		√
Liu *et al.* (2021)	Quantitive			√
Elberry *et al.* (2023)	Qualitative		√	√
This paper	Quantitive+ Qualitative	√	√	√

Meanwhile, this paper proposes three hypotheses, H1, H2, and H3, based on a literature review and theoretical review. The research theoretical model can be viewed through Fig. 1.

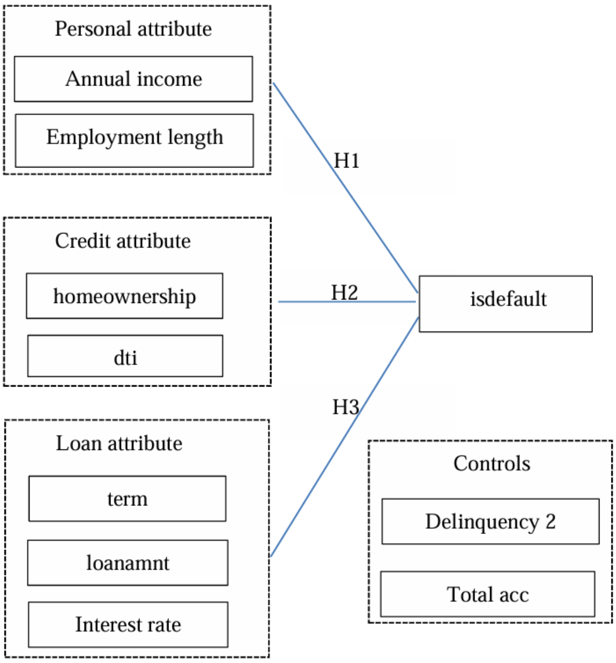

Fig. 1. Factors influencing loan default research model.

III. Data Collection and Analysis

A. Data and Methods

Loans are important matters related to financial security and personal privacy, so the corresponding data is generally not publicly available, which brings difficulties to data acquisition. However, financial institutions may disclose some anonymous historical data for risk prevention and scientific research purposes at certain times. We collected data on financial loans from a subsidiary of a Fortune Global 500 group in China (https://tianchi.aliyun.com/dataset/140861). This data discloses information on loan records in Chinese history, including whether there is default, annual income, employment length, homeownership, dti, term, loanamnt, interest rate, and delinquency 2years, totalacc, and other data.

On this basis, we established the following empirical research model (1) by combining the research framework and theoretical reasoning:

$$\text{Logit}(\text{isdefault}) = \beta_0 + \beta_1\text{annualincome} + \beta_2\text{employmentlength} + \beta_3\text{homeownership} + \beta_4\text{dti} + \beta_5\text{term} + \beta_6\text{loanamnt} + \beta_7\text{interestrate} + \beta_8\text{delinquency\_2years} + \beta_9\text{totalacc} \quad (1)$$

The dependent variable isdefault in the model belongs to 0, class 1 discrete variable. Based on the attributes of the dependent variable, this article selects the logit model for experiments, β_0 represents other unobservable or unobservable factors in this model, and $\beta_i, i = 1,2,3,4,5,6,7,8,9$ represents the degree of influence of each independent variable on the dependent variable. The last two variables in the model are the control variables, used to enhance the explanatory power of the model.

B. *Descriptive Statistics*

This study first conducted a descriptive statistical analysis of the data. Descriptive line analysis can perform prospective testing on data, identifying outliers in the data, and conducting prior analysis on research issues. As shown in Table 2, this experiment analyzed 100000 sample data, and there was no sample missing problem for each variable. From the perspective of specific variables, the average default rate is 0.197, indicating that the overall problem of loan default is within a controllable range. The per capita working experience is 6.059 years, reflecting that the loan population is mainly distributed among those who have worked for several years. Meanwhile, we can observe that the annual income variance of the population is relatively large (Std (annualincome) = 76349.721), this indicates that there is still a strong wealth gap in the sample data. Meanwhile, the mean and variance of the number of real estate owned by individuals in the sample remain at a relatively concentrated level. And Mean (homeownership) = 0.61, std (homeownership) = 0.673, indicating that the asset possession status of society is relatively average. At the same time, Mean (delinquency 2years) = 0.325 indicates that this indicator is a feature that differs from 0 and may be a good reference feature for default prediction with differences in the population.

Table 2. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
isdefault	100000	0.197	0.398	0	1
Annual income	100000	77683.972	76349.721	2000	10999200
Employment length	100000	6.059	3.569	1	10
homeownership	100000	0.61	0.673	0	5
This paper	Quantitive+ Qualitative	√	√	√	
dti	100000	18.298	8.995	0	489.16
term	100000	3.496	0.864	3	5
loanamnt	100000	14710.022	8766.236	1000	40000
interestrate	100000	13.256	4.818	5.31	30.99
delinquency 2years	100000	0.325	0.896	0	27
totalacc	100000	25.165	12.043	2	141

C. *Correlation Analysis*

Econometrics requires empirical testing to show that there is no high correlation between independent variables. Therefore, we used Python language to plot the visual results of variable correlation, as shown in Fig. 2. Among them, the diagonal of the image represents the degree of correlation between each variable itself, while other regions represent the degree of correlation between variables and non self variables. The overall data appears blue (low correlation), and there is no correlation issue. The fixed housing assets and the overall available credit limit show extreme irrelevance, but the number of loan terms and loan amount shows a slight correlation. We also conducted the VIF method in empirical testing to distinguish the correlation between variables, which is a common correlation discrimination method in the empirical field. The results showed that the VIF values were all less than 5, supporting the rationality and robustness of this experiment.

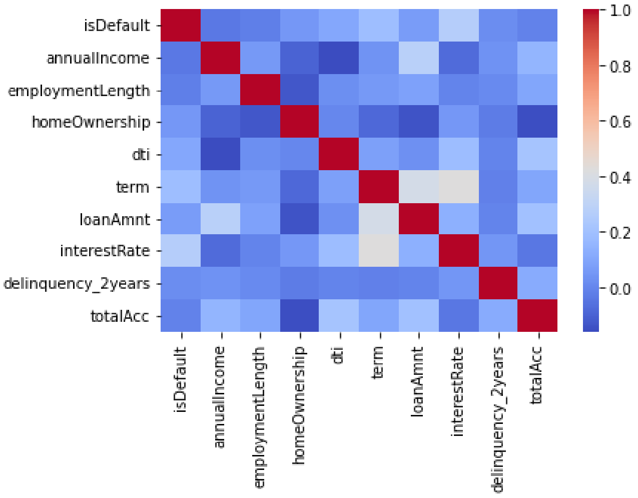

Fig. 2. Correlation of variables.

D. *Empirical Results and Analysis*

After descriptive row statistics and correlation analysis, we conducted empirical tests based on the research model (1), and the test results are shown in Table 3. Column (1) first experimentally examined the case where only the main independent variables were included. Column (2) added control variables to column (1) to enhance the explanatory power of the model, which can be seen through indicators such as R2(2) = 0.0819> R2 (1) = 0.0813. The absolute values of AIC and BIC are both decreasing, indicating that the model of column (2) is better than that of column (1).

The experimental results are shown in Table 3, $\beta_1 = -0.000$ *** and $\beta_2 = -0.013$ *** significant negative impact on the probability of default occurrence, H1 hypothesis verified. $\beta_3 = 0.186$ *** indicates that the number of real estate ownership positively affects the occurrence of default events, which is inconsistent with some descriptions in the H2 hypothesis. $\beta_4 = 0.015$*** indicates that the debt-to-income ratio positively affects the occurrence of default events, which is consistent with some descriptions in the H2 hypothesis. $\beta_5 = 0.250$***, $\beta_6 = 0.000$ *** and $\beta_7 = 0.105$ *** indicate that the hypothesis related to loan attributes in H3 has been validated, that is, the shorter the loan term, the less the amount, and the lower the interest rate, the lower the probability of loan default occurring.

In order to further explore the relationship between the number of real estate ownership and loan default in the H2 hypothesis, this study further added the quadratic term of the number of real estate ownership to volume (3) and found that the coefficient of this term was −0.257***, indicating an inverted U-shaped relationship between the impact of the number of real estate ownership on loan default. That is, people with extremely large and very small numbers of real estate do not bring high default risk, which is unexpected. In theory, people with a large number of real estate assets have a higher ability to repay debts, resulting in a lower risk of default. However, people with a very small number of real estate assets generally make cautious decisions about loan behavior, and the loan amount is often low. This result also confirms the positive relationship between loan amount and loan default.

Table 3. Experimental results about factors influencing loan default risk

	(1)	(2)	(3)
	Main	Main+Control	Further_analysis
Variables	isdefault	isdefault	isdefault
Personal attribute			
Annual income	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)
Employment length	−0.013***	−0.013***	−0.008***
	(0.002)	(0.002)	(0.002)
Credit attribute			
homeownership	0.186***	0.184***	0.628***
	(0.012)	(0.013)	(0.035)
homeownership2			−0.257***
			(0.019)
dti	0.015***	0.016***	0.017***
	(0.001)	(0.001)	(0.001)
Loan attribute			
term	0.250***	0.255***	0.262***
	(0.010)	(0.010)	(0.010)
loanamnt	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
interestrate	0.105***	0.104***	0.103***
	(0.002)	(0.002)	(0.002)
Controls			
delinquency_2years		0.057***	0.060***
		(0.009)	(0.009)
totalacc		−0.003***	−0.002***
		(0.001)	(0.001)
Constant	−4.063***	−4.043***	−4.198***
	(0.046)	(0.047)	(0.048)
Observations	100,000	100,000	100,000
R^2	0.0813	0.0819	0.0838
AIC	91126.99	91077.5	90893.06
BIC	91203.09	91172.63	90997.7

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

E. Robust Check

In order to further enhance the credibility of our research conclusions, we conducted two types of robust tests. Robustness testing is often used in disciplines such as economics and management to enhance the explanatory power and credibility of obtained conclusions and reduce endogeneity issues in research. Common methods include changing measurement methods, expanding sample sets, and instrumental variables.

On the one hand, based on the characteristics of the research and the characteristics of the dataset, we first used the form of replacement econometric methods to test the conclusions. Probit regression is also applicable to discrete forms with dependent variables of 0 and 1. It is usually used in conjunction with logistic regression in cases of uncertain data population distribution to ensure the scientific nature of research conclusions. In addition to transforming the logit regression model into a Probit regression model, column (4–6) is a replication of column (1–3). From the experimental results, it can be seen that the variable coefficients and significance in Table 4 are consistent with the main test, which enhances the credibility of our experimental results.

Table 4. Robust results about factors influencing loan default risk

	(4)	(5)	(6)
	P_Main	P_Main_Control	P_Further_analysis
Variables	isdefault	isdefault	isdefault
Personal attribute			
annualincome	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)
employmentlength	−0.007***	−0.007***	−0.008***
	(0.001)	(0.001)	(0.002)
Credit attribute			
homeownership	0.109***	0.107***	0.628***
	(0.007)	(0.007)	(0.035)
homeownership2			−0.257***
			(0.019)
dti	0.009***	0.009***	0.017***
	(0.001)	(0.001)	(0.001)
Loan attribute			
term	0.145***	0.148***	0.262***
	(0.006)	(0.006)	(0.010)
loanamnt	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
interestrate	0.063***	0.062***	0.103***
	(0.001)	(0.001)	(0.002)
Controls			
delinquency_2years		0.032***	0.060***
		(0.005)	(0.009)
totalacc		−0.002***	−0.002***
		(0.000)	(0.001)
Constant	−2.423***	−2.405***	−4.198***
	(0.025)	(0.026)	(0.048)
Observations	100,000	100,000	100,000

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

On the other hand, we also conducted experiments by expanding the data sample. The purpose of this robustness test is to prevent bias in the sample data during sampling, making it difficult to cover the overall situation, and to avoid potential endogeneity issues that may arise as a result. Column (7–9) is the result of an experiment conducted on 200000 sample sizes, and the experimental method and variables are consistent with Column (1–3), only doubling the sample size on the original dataset. The experimental results are shown in Table 5. It can be observed that the experimental results are still consistent with the main test, and this robust test result can alleviate concerns about sample selection bias in the experiment.

Table 5. Results in expanded dataset

	(7)	(8)	(9)
	Main	Main_Control	Further_analyse
Variables	isdefault	isdefault	isdefault
Personal attribute			
annualincome	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)
employmentlength	−0.012***	−0.012***	−0.007***
	(0.002)	(0.002)	(0.002)
Credit attribute			
homeownership	0.191***	0.188***	0.665***
	(0.009)	(0.009)	(0.025)
homeownership2			−0.277***
			(0.014)
dti	0.014***	0.015***	0.016***
	(0.001)	(0.001)	(0.001)
Loan attribute			
term	0.235***	0.240***	0.248***
	(0.007)	(0.007)	(0.007)
loanamnt	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
interestrate	0.106***	0.105***	0.103***
	(0.001)	(0.001)	(0.001)
Controls			
delinquency_2years		0.047***	0.050***
		(0.006)	(0.006)
totalacc		−0.004***	−0.003***
		(0.001)	(0.001)
Constant	−4.030***	−4.003***	−4.170***
	(0.033)	(0.033)	(0.034)
Observations	200,000	200,000	200,000

Note: *** $p < 0.01$, ** $p <0.05$, * $p <0.1$

F. Heterogeneity Analysis

We conducted heterogeneity analysis mainly for the following two reasons. Firstly, banks are prone to a type preference when approving loans, which means they are more willing to engage in a specific type of loan (Song *et al.*, 2023). Second, the experimental samples of this study are mainly from Chinese Mainland. During the sample period, real estate finance prevailed, which may lead to irrational loans and other problems.

Therefore, we divided the dataset into real estate loans and non real estate loans based on loan purposes, and tested it based on the research model (1). We found that there is no significant relationship between loan default and loan amount in real estate loans, meaning that loan amount should not be used as an important reference indicator for evaluating the default risk of real estate loans. For non real estate loans, there is no significant relationship between working years and default risk, that is, the risk prevention and control of non real estate loans should not be based on the borrower’s working years as an important reference indicator.

Table 6. Heterogeneity results about factors influencing loan default risk

	(10)	(11)
	Real_estate	No_Real_estate
Variables	isdefault	isdefault
Personal attribute		
Annual income	−0.000***	−0.000***
	(0.000)	(0.000)
Employment length	−0.011***	−0.004
	(0.003)	(0.004)
Credit attribute		
homeownership	0.613***	0.642***
	(0.045)	(0.056)
homeownership2	−0.246***	−0.268***
	(0.025)	(0.030)
dti	0.016***	0.018***
	(0.001)	(0.002)
Loan attribute		
term	0.270***	0.248***
	(0.013)	(0.017)
loanamnt	0.000	0.000***
	(0.000)	(0.000)
interestrate	0.101***	0.106***
	(0.002)	(0.003)
Controls		
delinquency_2years	0.063***	0.054***
	(0.011)	(0.014)
totalacc	−0.001	−0.004***
	(0.001)	(0.001)
Constant	−4.132***	−4.295***
	(0.062)	(0.077)
Observations	58,437	41,563
R2	0.0818	0.0849
AIC	54952.04	35926.58
BIC	55050.77	36021.57

Note: *** $p <0.01$, ** $p <0.05$, * $p <0.1$

These conclusions supplement and improve the understanding of risk prevention and control in the financial industry and financial lending and help to further strengthen the development and upgrading of risk prevention and control indicators for segmented industries. At the same time, it may also alleviate the difficulties faced by small and medium-sized enterprises in bank lending.

IV. Conclusion

In this study, we delved into the influencing factors of loan default risk and conducted extensive analysis based on real loan data disclosed by a subsidiary of a Fortune 500 group in China. The use of descriptive statistical analysis, correlation analysis, and other methods provides a strong foundation for further empirical research. At the same time, we constructed a Logit regression model to explore the impact of different independent variables on loan defaults. Through estimation and analysis of the model, we have drawn the following key conclusions, which are summarized in Table 7.

Firstly, annual income and employment length have a significant negative impact on the loan default rate, that is, with the increase in annual income and working years, the probability of loan default significantly decreases. This means that stable economic conditions and longer work experience have a significant impact on loan default risk.

Secondly, the ratio of debt to income has a positive impact on loan default, and there is an inverse U-line relationship between homeownership and loan default, meaning that people with extremely large or very small holdings are more likely to default. This result is somewhat unexpected, as the quantity of real estate ownership cannot be simply used as a primary variable as an evaluation indicator, and high debt is significantly positively correlated with default risk.

Thirdly, the number of loan terms, loan amount, and loan interest rate also have a significant impact on the default rate. Shorter term loans, lower loan amounts, and lower loan interest rates all reduce the probability of loan default.

Table 7. Summary of hypothesis

Hypothesis	Result
H1: The higher a personal’s annual income and employment length, the lower the risk of loan default.	√
H2: The higher homeownership, the lower the risk of loan default.	×
H2: The lower their debt-to-income ratio, the lower the risk of loan default.	√
H3: The shorter the term of the loan, the smaller the amount, the lower the interest rate, and the lower the risk of loan default.	√

We also conducted robustness tests, using Probit regression models and expanding data samples to validate our conclusions, and the results consistently supported our main findings.

Finally, we divided the dataset into real estate and non real estate loans based on different loan purposes and found that there were differences in key factors among different loan categories. This emphasizes that financial institutions should adopt different strategies based on loan types in risk prevention and loan decision-making.

In summary, this study provides important insights on loan default risk, which can help financial institutions better understand and manage risks, as well as improve loan decision-making. Our conclusion not only has important guiding significance for practitioners in the financial field but also provides useful information for small and medium-sized enterprises and their loan applicants, which is expected to help them better cope with loan default risks. This study provides a new perspective and method for empirical research on loan default risk and also provides valuable reference for future related research.

A. Theoretical Implications

In terms of theory, this study provides important contributions to the field of loan default risk. Firstly, the loan default model we constructed delves into the impact of multiple dimensional factors on default risk, including individual economic conditions, loan characteristics, and the number of real estate holdings. This not only broadens the understanding of loan default risk but also expands the theoretical framework of its influencing factors. This has enlightening significance for the academic community's research on loan default risk and has stimulated more in-depth exploration of the interaction of influencing factors.

Secondly, the findings of this study emphasize the complex relationship between different factors and the inverse U-shaped relationship between the number of real estate holdings and loan defaults. This helps to further improve the theoretical framework and provides a new theoretical foundation for the fields of economics, finance, and risk management.

Finally, our robustness testing method and the application of data augmentation also provide examples of empirical research methods, improving the scientific and credibility of the research, which has positive significance for promoting the development of methodology in the field of economics.

B. Practical Implications

At the practical level, this study has significant practical significance for financial institutions, government regulatory authorities, and loan applicants.

Firstly, financial institutions can use our research findings to improve risk assessment and loan decision-making. More accurate default risk prediction will help financial institutions manage non-performing assets more effectively, improve asset quality, and reduce risk exposure.

Secondly, government regulatory agencies can draw on the results of this study to strengthen financial market regulatory policies and maintain the stability of the financial system. A deeper understanding of the factors underlying default risk will help to formulate regulatory policies more scientifically and enhance the resilience of the financial system.

Thirdly, loan applicants can obtain valuable advice from this study and manage their loan risks more wisely. Understanding one's own economic situation, carefully selecting loan characteristics, and carefully evaluating factors such as debt to income ratio will help reduce the likelihood of default, enhance personal financial health, and improve the chances of obtaining loans in the future.

In summary, this study not only expands the theoretical research on loan default risk, but also provides practical guidance and reference for financial institutions, government regulatory departments, and individual loan applicants. It is expected to have a positive impact on the stability of the financial market and the improvement of personal financial condition.

Conflict of Interest

The author has claimed that no conflict of interest exists.

REFERENCES

Ballester, L., González-Urteaga, A., & Martínez, B. 2020. The role of internal corporate governance mechanisms on default risk: A systematic review for different institutional settings. ***Research in International Business and Finance***, 54: 29, Article 101293. https://doi.org/10.1016/j.ribaf.2020.101293

Berg, T., Puri, M., & Rocholl, J. 2020. Loan officer incentives, internal rating models, and default rates. ***Review of Finance***, 24(3): 529–578. https://doi.org/10.1093/rof/rfz018

Chen, Q., Tsai, S. B., Zhai, Y. M., Chu, C. C., Zhou, J., Li, G. D., Zheng, Y. X., Wang, J. T., Chang, L. C., & Hsu, C. F. 2018. An empirical research on bank client credit assessments. ***Sustainability***, 10(5). https://doi.org/ARTN140610.3390/su10051406

Chong, F. 2021. Loan delinquency: Some determining factors. ***Journal of Risk and Financial Management***, 14(7): 320.

Covin, J. G., & Slevin, D. P. 1989. Strategic management of small firms in hostile and benign environments. ***Strategic Management Journal***, 10(1): 75–87.

Cox, J. C., Kreisman, D., & Dynarski, S. 2020. Designed to fail: Effects of the default option and information complexity on student loan repayment. ***Journal of Public Economics***, 192: 19, Article 104298. https://doi.org/10.1016/j.jpubeco.2020.104298

Croux, C., Jagtiani, J., Korivi, T., & Vulanovic, M. 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. ***Journal of Economic Behavior & Organization***, 173: 270–296. https://doi.org/10.1016/j.jebo.2020.03.016

De Giorgi, G., Drenik, A., & Seira, E. 2023. The extension of credit with nonexclusive contracts and sequential banking externalitiest. ***American Economic Journal-Economic Policy***, 15(1): 233–271. https://doi.org/10.1257/pol.20200220

Do, H. X., Rösch, D., & Scheule, H. 2018. Predicting loss severities for residential mortgage loans: A three-step selection approach. ***European Journal of Operational Research***, 270(1): 246–259. https://doi.org/10.1016/j.ejor.2018.02.057

Duarte, F. D., Gama, A. P. M., & Gulamhussen, M. A. 2018. Defaults in bank loans to SMEs during the financial crisis. ***Small Business Economics***, 51(3): 591–608. https://doi.org/10.1007/s11187-017-9944-9

Elberry, N. A., Naert, F., & Goeminne, S. 2023. Optimal public debt composition during debt crises: A review of theoretical literature. ***Journal of Economic Surveys***, 37(2): 351–376. https://doi.org/10.1111/joes.12491

Gao, W. H., Liu, Y., Yin, H., & Zhang, Y. W. 2022. Social capital, phone call activities and borrower default in mobile micro-lending. ***Decision Support Systems***, 159: 10, Article 113802. https://doi.org/10.1016/j.dss.2022.113802

Gapko, P., & Smíd, M. 2019. Modeling credit losses for multiple loan portfolios. ***Finance a Uver-Czech Journal of Economics and Finance***, 69(6): 558-579. <Go to ISI>://WOS:000537833600003

Henager, R., & Wilmarth, M. J. 2018. The relationship between student loan debt and financial wellness. ***Family and Consumer Sciences Research Journal***, 46(4): 381–395.

Ju, Y., & Sohn, S. Y. 2017. Technology Credit Scoring Based on a Quantification Method. ***Sustainability***, 9(6): 16, Article 1057. https://doi.org/10.3390/su9061057

Jung, H., & Kim, H. H. 2020. Default probability by employment status in South Korea*. ***Asian Economic Papers***, 19(3): 62–84. https://doi.org/10.1162/asep_a_00786

Lee, J. W., Lee, W. K., & Sohn, S. Y. 2021. Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. ***Expert Systems with Applications***, 168: 7, Article 114411. https://doi.org/10.1016/j.eswa.2020.114411

Li, B. D. 2022. Online loan default prediction model based on deep learning neural network. ***Computational Intelligence and Neuroscience***, 9, Article 4276253. https://doi.org/10.1155/2022/4276253

Li, X., Ergu, D., Zhang, D., Qiu, D., Cai, Y., & Ma, B. 2022. Prediction of loan default based on multi-model fusion. ***Procedia Computer Science***, 199: 757–764.

Li, Z. Y., Li, A. M., Bellotti, A., & Yao, X. 2023. The profitability of online loans: A competing risks analysis on default and prepayment. ***European Journal of Operational Research***, 306(2): 968–985. https://doi.org/10.1016/j.ejor.2022.08.013

Liao, L., Martin, X., Wang, N., Wang, Z. W., & Yang, J. 2023. What if borrowers were informed about credit reporting? Two natural field experiments. ***Accounting Review***, 98(3): 397–425. https://doi.org/10.2308/tar-2021-0191

Lin, T. T., Lee, C. C., & Chen, C. H. 2011. Impacts of the borrower's attributes, loan contract contents, and collateral characteristics on mortgage loan default. ***Service Industries Journal***, 31(9): 1385–1404, Article Pii 928028323. https://doi.org/10.1080/02642060903437535

Liu, H., Yuan, M. K., & Zhou, M. L. 2021. How does the urgency of borrowing in text messages affect loan defaults? Evidence from P2P loans in China. ***Security and Communication Networks***, 13, Article 4060676. https://doi.org/10.1155/2021/4060676

Looney, A., & Yannelis, C. 2022. The consequences of student loan credit expansions: Evidence from three decades of default cycles. ***Journal of Financial Economics***, 143(2): 771–793. https://doi.org/10.1016/jafineco.2021.06.013

Lu, S. L., & Wang, M. C. 2012. How to measure the credit risk of housing loans: Evidence from a Taiwanese Bank. ***Emerging Markets Finance and Trade***, 48: 122-138. https://doi.org/10.2753/ree1540-496x48s207

Luong, T. M., & Scheule, H. 2022. Benchmarking forecast approaches for mortgage credit risk for forward periods. ***European Journal of Operational Research***, 299(2): 750–767. https://doi.org/10.1016/j.ejor.2021.09.026

Lyócsa, S., Vasanicová, P., Misheva, B. H., & Vateha, M. D. 2022. Default or profit scoring credit systems? Evidence from European and US peer-to-peer lending markets. ***Financial Innovation***, 8(1): 21, Article 32. https://doi.org/10.1186/s40854-022-00338-5

Ma, Y. 2019. Prediction of default probability of credit-card bills. ***Open Journal of Business and Management***, 8(01): 231.

Medina-Olivares, V., Calabrese, R., Dong, Y. Z., & Shi, B. F. 2022. Spatial dependence in microfinance credit default. ***International Journal of Forecasting***, 38(3): 1071–1085. https://doi.org/10.1016/j.ijforecast.2021.05.009

Nalić, J., & Švraka, A. 2018. Using data mining approaches to build credit scoring model: Case study—implementation of credit scoring model in microfinance institution. ***Proceedings of 2018 17th International Symposium Infoteh-Jahorina*** (INFOTEH),

Netzer, O., Lemaire, A., & Herzenstein, M. 2019. When words sweat: Identifying signals for loan default in the text of loan applications. ***Journal of Marketing Research***, 56(6): 960–980. https://doi.org/10.1177/0022243719852959

Park, K. T., Yang, H., & Sohn, S. Y. 2022. Recommendation of investment portfolio for peer-to-peer lending with additional consideration of bidding period. ***Annals of Operations Research***, 315(2): 1083–1105. https://doi.org/10.1007/s10479-021-04300-z

Ravina, E. 2019. Love & loans: The effect of beauty and personal characteristics in credit markets. Available at SSRN 1107307.

Saha, A., Rooj, D., & Sengupta, R. 2022. Loan to value ratio and housing loan default - evidence from microdata in India. ***International Journal of Emerging Markets***, 20. https://doi.org/10.1108/ijoem-10-2020-1272

Song, Y., Wang, Y. Y., Ye, X., Zaretzki, R., & Liu, C. R. 2023. Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. ***Information Sciences***, 629: 599–617. https://doi.org/10.1016/j.ins.2023.02.014

Thomas, S. S., George, J. P., Godwin, B. J., & Siby, A. 2023. Young adults' default intention: Influence of behavioral factors in determining housing and real estate loan repayment in India***. International Journal of Housing Markets and Analysis***, 16(2): 426–444.

Zhang, W. G., Wang, C., Zhang, Y., & Wang, J. B. 2020. Credit risk evaluation model with textual features from loan descriptions for P2P lending. ***Electronic Commerce Research and Applications***, 42: 15, Article 100989. https://doi.org/10.1016/j.elerap.2020.100989