International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 5, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijgem.v5n1.01
Journal homepage: https://wepub.org/index.php/IJGEM/index

Impacting Factors of Default Risk of Personal Housing Loans in China: Based on MCLP Model

Li Zhao, Guangrui He

School of Economics and management, Shanghai Maritime University, China

ABSTRACT

A large-scale default on personal housing loans will have a significant negative impact on the stability of the financial system and the smooth operation of the macro-economy. After the pandemic, preventing personal housing loan risks has become particularly important. Based on personal housing loans from commercial banks in China, a personal housing loan default risk model was constructed using the MCLP model, and the prediction results of the MCLP model were compared with those of the traditional logistic model to identify indicators that may affect default risk of personal housing loans. It was found that the MCLP had higher accuracy than the traditional logistic model. Empirical findings indicated that the education level, employment status, monthly household income, and census register of housing loan borrowers were primary factors influencing customer defaults.

KEYWORDS

Personal housing loans; Loan default risk; MCLP model

1. INTRODUCTION

The "4 trillion yuan" economic stimulus plan in 2008 sparked a trend of "housing speculation", significantly increasing demand for home purchases. Many people became ensnared in the grip of "housing slavery", and the scale of housing loans continued to expand. Currently, personal housing loans for residents in China constitute the largest proportion of personal loans. Consequently, a large-scale default due to borrowers' inability to repay loans due to lack of income could adversely affect the stability of the financial system and the smooth operation of the macro economy.

During the pandemic, numerous instances of delayed mortgage repayments emerged in Chinese banks. In the short term, the global spread of COVID-19 primarily impacted banks through a decline in credit demand, bad debts from some customers, and losses incurred from interest concessions. Simultaneously, the pandemic led to an increase in social unemployment, a decrease in personal income, and a reduction in debt repayment capacity. This situation is expected to elevate the default risk of personal housing loans for banks and augment the cost of collecting commercial debts. This article analyzes real-world data on personal housing loans from commercial banks in China, identifies factors that may lead to loan customer defaults, and constructs a personal housing loan default risk model based on the MCLP model to empirically test the factors that affect loan performance. The effectiveness of the model is analyzed and compared, and relevant suggestions for effectively identifying personal housing loan default risks are proposed.

The structure of this article is arranged as follows: in addition to the introduction, relevant literature is described in the second part of this article, relevant data and indicators and the main models are introduced in the third part, and the forth part summarizes the entire article and puts forward some constructive suggestions.

2. RELATED LITERATURE

Some related studies on the default risk of personal housing loans, such as Berry and Dalton's research on customer loan data of Australian financial institutions in early foreign studies found that marital status, long-term unemployment of family members, and sudden decline in household income are important factors leading to default risk [1]; In domestic research, Xu and Zhao (2011) used BP filtering method to empirically analyze the relationship between the fluctuation of non-performing loan ratio of Bank of Communications and GDP [2].

The earliest commonly used tool for analyzing bank credit risk was discriminant analysis [3-4], and in subsequent research, scholars began to use more accurate methods such as multiple linear regression and Probit analysis [5]. After entering the 1990s, Glover further expanded the applicability and flexibility of linear discriminant analysis models, and since then more methods have been used for credit risk discrimination [6].

Multi-factor Linear Regression Analysis (MCLP) is a credit analysis method deemed potentially superior to existing methods in the field of credit scoring. Compared to traditional models, the MCLP model excels in accuracy, sensitivity, and other metrics. This approach has been effectively employed in the analysis of credit card customers. For MCLP, b serves as the dividing line between "good customers" and "bad customers", with the value of b significantly impacting the accuracy of the MCLP model. Li, YU, and Lu utilized German credit card customer data from the UCI database to conduct a sensitivity analysis on how the value of b affects classification accuracy, ultimately concluding that b=1 yields the highest classification accuracy [7]. Shi, Wise, Luo, and others applied MCLP to the study of default risk among credit card customers, classifying cardholders into "good" and "bad" categories based on their default risk levels [8]. Kou, Liu, Peng, and others extended this method by further segmenting customers based on their default risk [9]. The above studies all suggest that MCLP can produce better results than other methods. Li, Shi, and He adopted three MCLP based methods to improve the accuracy of identifying "bad customers" [10].

To sum up, it becomes evident that the various influencing factors of personal housing mortgage loans bear significant similarities to those of credit card customers. Consequently, employing the MCLP model to study the default risk of personal housing loans is expected to enhance the accuracy of the model's predictive outcomes. This article explores the risk factors contributing to personal housing loan defaults against the background of the COVID-19 pandemic in China.

We employed a logistic regression model to examine the factors influencing the default risk of personal housing loans in sample banks. Subsequently, we compared the analysis results of the MCLP model and the logistic regression model, aiming to demonstrate the applicability of the MCLP model in studying the default risk of personal housing loans.

3. COMPREHENSIVE DISCRIMINANT ANALYSIS OF CUSTOMER DEFAULT RISK BASED ON MCLP MODEL

3.1. Research Hypothesis

To simplify the analysis, we first make several simple assumptions:

In the bank credit market, there exists a credit information sharing mechanism, where all banks share all credit information of borrowing individuals with one another. This mechanism mitigates adverse selection in the market.

During the pandemic, individuals faced both positive and negative options when it came to repaying their mortgage. This distinction does not reflect the moral attributes of the borrower; rather, it is an intrinsic choice.

The variations in factors affecting the default risk of personal housing loans are attributed to the pandemic.

3.2. Data Selection

The customer's basic personal information encompasses age, gender, census register, education, employment status, and monthly household income. The data was sourced from the personal housing loan data and documentation of a primary branch of a commercial bank spanning from 2019 to 2020. This study ultimately extracted 210 sets of customer data, selecting 59 sets of defaulting customers and 151 sets of customers who fulfilled their obligations, in accordance with the optimal sample ratio requirements of the default rate model [11].

Using SPSS software, we processed the 210 sets of data, considering "performance" as the dependent variable and other parameters as independent variables. We conducted a univariate logistic regression, excluding variables with a significance level of ≥ 5% that were significantly unrelated to breach of contract. The regression results, presented in Table 1 on the following page, indicate that education, employment status, monthly household income, and census register are significantly associated with default, whereas the other two indicators are not significantly related to default and should be eliminated.

3.3. MCLP Model Analysis

From a total of 210 customer groups, this study randomly selected 180 as the training set for model development, reserving the remaining 30 for testing the model's accuracy. We incorporated performance-related indicators from the 180 training sets into the MCLP model. Using Lingo software for model computation, we set α^* to the smallest possible value greater than zero, specifically α=0.001, and β to the largest possible value, namely β^*=99999. Based on research by Li, Yu, and Liu, the optimal value for b is b=1.

Table 1. MCLP Model Parameter Selection

variable	B	S.E.	Wals	df	Sig.	Exp(B)
gender	-0.601	0.542	1.233	1	0.267	0.548
age	-0.328	0.412	0.633	1	0.426	0.72
census register	2.278	0.575	15.71	1	0	9.758
education	1.101	0.309	12.708	1	0	3.007
Employment status	2.28	0.598	14.559	1	0	9.781
monthly household income	0.586	0.276	4.498	1	0.034	1.796

The Lingo calculation results are presented in Table 2. Here, D1 represents $d\alpha^-$, D2 denotes $d\alpha^+$, D3 signifies $d\beta^-$, and D4 indicates $d\beta^+$. X1 stands for education background, X2 for work status, X3 for monthly household income, and X4 for household registration.

The model is expressed as:

$$P=0.2X1+0.4X2+0.2X3+0.6X4 \tag{1}$$

Table 2. Analysis Results of MCLP Model

target value		75.001
Number of contradictory constraints		2.22E-16
variable	number	Reduced Cost
D1	75.001	0
D2	0	2
D3	0	1
D4	0	1
X3	0.2	0
X1	0.2	0
X4	0.6	0
X2	0.4	0

3.4. Model Checking

3.4.1. Prediction Accuracy

We evaluated the model's predictive accuracy by substituting data from the randomly selected 30 test sets. If the resulting P value is equal to or exceeds 1, it is assumed that the customer will fulfill the contract; otherwise, default is anticipated. The results are summarized in Table 3. We observed that the prediction accuracy for compliant customers was 83.3%, for defaulting customers was 75%, and the overall accuracy stood at 80%. The model's overall predictive accuracy is deemed satisfactory.

Table 3. MCLP Model Verification Results

		Predicting Customers	percentage
		performance default	
Original Customer	performance	15 3	83.3%
	default	3 9	75.0%
	summation		80.0%

3.4.2. ROC curve

The ROC curve for the MCLP model demonstrates an AUC value of 0.868, significantly higher than 0.7, and a P-value of 0.001, indicating that the MCLP model exhibits high prediction accuracy for customer default situations.

3.5. Comparison with Logistic Model

3.5.1. Analysis Results of Logistic Model

Table 4 reveals notable differences between the results of the Logistic model and the MCLP model. Based on the analysis, the estimated Logistic model for the sample is as follows:

$$\ln\frac{p}{1-p} = -4.244 + 1.968X_1 + 1.561X_2 + 2.035X_3 + 0.924X_4 - 2.710X_5 \quad (2)$$

Table 4. Logistic Model Parameter Selection

		B	S.E.	Wals	df	Sig.	Exp (B)
step1a	education	1.157	0.217	28.511	1	0.000	3.181
	constant	-0.495	0.324	2.332	1	0.127	0.609
step2b	education	1.170	0.235	24.835	1	0.000	3.220
	census register	1.770	0.465	14.493	1	0.000	5.872
	constant	-1.103	0.386	8.161	1	0.004	0.332
step3c	employment status	1.636	0.489	11.179	1	0.001	5.134
	education	1.138	0.247	21.228	1	0.000	3.122
	census register	1.844	0.492	14.066	1	0.000	6.320
	constant	-1.585	0.435	13.286	1	0.000	0.205
step4d	employment status	1.968	0.586	11.299	1	0.001	7.157
	education	1.561	0.312	25.075	1	0.000	4.765
	census register	2.035	0.597	11.616	1	0.001	7.654
	monthly household income	0.924	0.236	15.366	1	0.000	2.521
	constant	-4.244	0.913	21.589	1	0.000	0.014

3.5.2. Prediction Accuracy of Logistic Model

Table 5 presents the validation of the Logistic model's prediction accuracy using 30 test samples. Evidently, the prediction accuracy of the Logistic model is significantly lower than that of the MCLP model, highlighting that the MCLP method employed in this study offers superior accuracy in analyzing and predicting personal housing loan defaults.

Table 5. Logistic Model Test Results

		Predicting Customers	percentage
		performance default	
Original Customer	performance	12 6	66.7%
	default	5 7	58.3%
	summation		62.5%

4. CONCLUSION AND POLICY IMPLICATIONS

4.1. Conclusion

Empirical findings indicate that the education level, employment status, monthly household income, and census register of housing loan borrowers are primary factors influencing customer defaults. Specifically, borrowers with lower education levels are more prone to loan defaults; employees in state-owned enterprises, public institutions, and civil servants exhibit significantly better credit performance compared to other professionals; a higher household income correlates with a lower likelihood of default; and borrowers with local urban census register demonstrate significantly better repayment performance than others.

In contrast to previous research methodologies, this study introduces the MCLP model into the analysis of factors influencing personal housing loan defaults, revealing that the MCLP model's prediction accuracy is notably superior to methods like the logistic model. Consequently, the analytical model employed in this paper holds significant practical value for accurately assessing the default risk associated with personal housing loans in China. Building on the conclusions of this research, and incorporating advanced credit evaluation models from abroad, it will assist commercial banks in establishing credit evaluation models tailored to their unique business characteristics, fostering a more mature and rational credit evaluation system, and thereby enhancing their risk assessment and management capabilities.

Based on the above findings, the following suggestions are proposed.

First, it is crucial for borrowers to cultivate a consciousness of personal responsibility, trustworthiness, and timely loan repayment. The pandemic has resulted in numerous overdue mortgage repayments, elevating the risk of bad debts for banks and increasing the likelihood of default and dishonesty among borrowers. This has negatively impacted their credit histories, thereby affecting future credit activities.Additionally, borrowers must understand how to borrow prudently, consume wisely, and attain "green credit" within the framework of "green finance".

Next, let's consider commercial banks. They should place greater emphasis on borrowers' characteristics, including education background, household registration, employment status, and monthly household income. They should identify and prioritize customer groups with a high risk of default, adopting targeted measures to effectively mitigate loan default risks. In cases of force majeure, proactive prevention can significantly reduce risk losses.

Finally, concerning the government's input. The government should persist in promoting and educating the system that motives trustworthiness and penalizes dishonesty, striving for everyone to embody "sincerity in their hearts and actions". Additionally, the nation should enhance its credit system to enhance its feasibility and practicality. Government regulations on credit should be genuine and effective, grounded in evidence rather than dogmatism or formalism. It's also essential to broaden their application, ideally making them accessible to the broader society, and ensuring they are highly operational and practical.

REFERENCES

[1] Berry M, Dalton T. Home ownership into the new millennium: a view from the margin [J]. Urban Policy and Research, 2000, 18(4):435-454.

[2] Xu Yanyan, Zhao Zhenglong Empirical analysis and countermeasures research on the pro cyclical credit risk of commercial banks in China [J] New Finance, 2011 (2): 40-44.

[3] Klein B, Leffler K B. The role of market forces in assuring contractual performance [J]. The he Journal of Political Economy, 1981, 89(4):615-641.

[4] Myers J H, Forgy E W. The development of numerical credit evaluation systems [J]. Journal of the American Statistical Association, 1936, 303 (58):799-806.

[5] Glover F, Laguna M. Tabu search [M]. US: Springer US, 1999.

[6] Li A, Yu Y, Liu Z. Sensitivity Analysis of b in MCLP model for credit card-holders classification [J]. Procedia Computer Science, 2013, 17 (1):1023-1031.

[7] Shi Y, Wise M, Luo M, etal. Data mining in credit card portfolio management: a multiple criteria decision making approach [M]. US: Springer Berlin Heidelberg, 2001.

[8] Kou G, Liu X, Peng Y, etal. Multiple criteria linear programming approach to data mining: models, algorithm designs and software development [J]. Optimization Methods and Software, 2003, 18(4):453-473.

[9] Li A, Shi Y, He J. MCLP based methods for improving “bad” catching rate in credit cardholder behavior analysis [J]. Applied Soft Computing, 2008, 8(2):1259-1265.

[10] Jia Shenghua, Li, Hang. Noise traders' expectations and real estate foam: an empirical study —based on 35 large and medium-sized cities [J] Audit and Economic Research, 2014 (3): 85-92

[11] Li A, Yu Y, Liu Z. Sensitivity Analysis of b in MCLP model for credit card-holders classification [J]. Procedia Computer Science, 2013, 17 (1):1023-1031.