

Measuring and Managing Credit Risk for Chinese Microfinance Institutions

Jie Li¹ & Zhenyu Sheng¹

¹ College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence: Jie Li, College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Tel: 86-137-7669-8523. E-mail: jieli@nuaa.edu.cn

Received: April 17, 2018

Accepted: May 28, 2018

Online Published: June 5, 2018

doi:10.5539/ijef.v10n7p56

URL: <https://doi.org/10.5539/ijef.v10n7p56>

Abstract

Chinese microfinance institutions need to measure and manage credit risk in a quantitative way in order to improve competitiveness. To establish a credit scoring model (CSM) with sound predictive power, they should examine various models carefully, identify variables, assign values to variables and reduce variable dimensions in an appropriate way. Microfinance institutions could employ both CSM and loan officer's subjective appraisals to improve risk management level gradually. The paper sets up a CSM based on the data of a microfinance company running from October 2009 to June 2014 in Jiangsu province. As for establishing the model, the paper uses Linear Discriminant Analysis (LDA) method, selects 16 initial variables, employs direct method to assign variables and adopts all the variables into the model. Ten samples are constructed by randomly selecting records. Based on the samples, the coefficients are determined and the final none-standardized discriminant function is established. It is found that Bank credit, Education, Old client and Rate variables have the greatest impact on the discriminant effect. Compared with the same international models, this model's classification effect is fine. The paper displays the key technical points to build a credit scoring model based on a practical application, which provides help and references for Chinese microfinance institutions to measure and manage credit risk quantitatively.

Keywords: default risk, microfinance institutions, credit scoring

1. Introduction

As an institutional arrangement of financial innovation, microfinance institutions (MFIs) have developed rapidly in recent years. May 2008, China Banking Regulatory Commission (CBRC) and the People's Bank of China jointly issued "Guidance on microfinance institutions", which implied microfinance institutions into China's financial system as a formal financial arrangement, and microfinance institutions have gradually entered a commercially operational stage. In recent years, microfinance industry has grown quickly and has been regarded as a sunrise industry with huge market space. Commercial microfinance institutions are thriving with a consistent manner. By the end of June 2014, the number of microfinance institutions has reached 8394 and the loan balance has reached 881.1 billion Yuan (Note 1).

Meanwhile, the competition in microfinance industry is increasing. Village banks, community banks and other institutions positioning in microfinance have also been set up. In addition, policy banks, large commercial banks, small and medium commercial banks, foreign banks, have also shown great enthusiasm and a great intension in the field of microfinance investment (Shusong, Yongfeng, & Xingliang, 2012). Fullerton Financial Holdings, entirely owned by Temasek, has invested hundreds of million Yuan to establish a number of microfinance organizations, rural banks and other institutions, in order to make a national layout for seizing the SME loan market. Improving efficiency, reducing costs and controlling risk effectiveness (Blanco, Pino-Mejías, Lara, & Rayo, 2013) are necessary for microfinance institutions to survive and keep growing in the fierce competition.

From the experience of foreign countries, MFIs should measure credit risk in a quantitative way in order to run well. In developed countries, credit scoring model (CSM) has become an important tool for risk management, and being used widely (Rhyne & Christen, 1999). In the United States, CSM shortens the loan time of small enterprises from 2 weeks to 12 hours; some network banks (such as Banc One) even entirely rely on CSM to screen and issue loans under \$ 35,000. In Canada, CSM cuts personal loan time from 9 days to 3 days.

Still lots of MFIs in China have not introduced CSM. In addition, due to ignorance of the importance of CSM, they worry that weak data may make it difficult to establish CSM. In fact, this is bias. One can't wait to develop a CSM until the data base is perfect. In most of the microfinance institutions the data accumulation is poor, but that does not mean they cannot establish a CSM. On one side, the modern classification and measurement instruments can effectively deal with qualitative data. On the other side, any good model cannot be done overnight, they could be constantly updated and improved with the time going.

Our paper complements and extends the previous work along three dimensions. Firstly, since MFIs in China emerged as late as 2008, the research on their CSMs is quite little. Our work enriches the existing studies. Secondly, we use an actual case to explore modeling techniques for Chinese MFIs for the first time as far as we know, such as choosing model, identifying initial variables, assigning value to variables, estimating effect. Finally, our model has quite good effect on classifying customers. The paper shows that Bank credit, Education, Old client and Rate are the main determinative factors for credit risk of Chinese MFIs, which has not been addressed in other studies.

This paper will explore modeling techniques on CSM based on the experience of establishing CSM from a microfinance firm in Jiangsu Province and discuss how to combine CSM and corporate risk management effectively. We hope to make a contribution to control credit risk of microfinance institutions.

2. Key Techniques of CSM Establishment

In order to establish a CSM with sound predictive power, some key techniques need to be examined. We should choose a proper model, identify variables, assign values to variables and reduce variable dimensions in an appropriate way.

The first step is to choose an appropriate model. There are many models in the field of credit risk measurement, Linear Discriminant Analysis(LDA), Logit Regression(LR), Neural Network Model, KMV model, JP Morgan's Credit Metrics, etc. Some high-level models are based on default database or company shares price, such as KMV model, JP Morgan's Credit Metrics, which are not suitable for microfinance institutions. According to international experience, traditional models like LDA and LR are most widely used, say Viganò (1993), Dinh and Kleimeier (2007), Schreiner (2004). LR is characterized by giving the probability of default while LDA gives only classification results. But LR has some limitations in application that it requires a larger sample size. In order to obtain more stable and reliable results, typically there must be more than 2000 samples. Domestic microfinance institutions can choose the right model according to its own data characteristic and model classification effect.

Second, we need to identify initial variables appropriately. There are a variety of approaches for choosing the type and quantity of CSM's initial variables (Dinh & Kleimeier, 2007). First Data Resources' credit scoring model uses 48 initial variables; top credit scoring company Fair Isaac' model uses 50-60 initial variables (Mester, 1997); Viganò (1993), when developing model with date of Burkina Faso microfinance institutions, used 53 initial variables; Schreiner (2004) developed the model based on Bolivia microfinance institutions' data, using only 9 initial variables; and Dinh and Kleimeier (2007) developed a model based on Vietnamese retail banking's data, using 22 initial variables. The most important thing is to ensure authenticity of the data, rather than the number of variables. Many model developers prefer financial data, but according to our practice, most microfinance institutions' customer financial data is not so reliable. We can use more qualitative data, such as education, housing situation and so on. Our model has a total of 16 variables.

Careful consideration is required when assigning values to variables in developing a model. From international experience, there are two methods, direct method and grouping method to give values to variables. Direct method uses specific values directly. For example, use specific number for age and use specific month or year for loan period. Grouping method is slightly more complex. The records need to be divided into several groups in terms of the absolute value of a variable. Variables in the same group will be assigned the same value according to the good records ratio in that specific group. Viganò (1993) and Blanco et al. (2013) used direct method, while Dinh and Kleimeier (2007) used grouping method. When establishing a CSM, domestic microfinance institutions may choose the method according to their own data characteristic. In our experience, if the sample size is large and the default rate is high, grouping method could be used. When establishing the model, we use the direct assignment method.

Attention is required for the issue of reducing variables. From international experience, directly using all variables is also feasible when establishing a CSM. Whether dimension-reduction treatment does is mainly based on the number of variables, correlation and prediction accuracy of the model. Viganò (1993), Dinh and Kleimeier (2007) carried out dimension-reduction, while Schreiner (2004) and Blanco et al. (2013) used all

variables directly. And there are diverse methods to reduce variable dimensions, correlation analysis, principal component analysis, factor analysis, cluster analysis, step-by-step method. Using which methods should base on data characteristic, whether these data can go through tests required by relevant methods. Viganò (1993) used correlation analysis and factor analysis, Dinh and Kleimeier (2007) used backward stepwise method. After analyzing the effect of whether reducing dimensions or not, we put all the variables into the model.

3. How to Use CSM in Credit Risk Management

After establishing a CSM, the next question is how to use the model and how to combine it with MFIs' risk management practice? CSM should be in its integration with loan officer's subjective appraisals. CSM's main function is to reduce costs. It is a standardized and quantitative method with advantage of objectivity and scientific basis. But the appraisal of borrower's repayment ability and willingness is comprehensive. Complete information cannot be reflected in the model. For example, some industry environment is constantly changing along with economic cycle. If the empirical results based on previous data are used to assess the creditworthiness of current customers, it will inevitably lead to a significant deviation. Therefore, microfinance institutions could employ both CSM and loan officer's subjective appraisals to improve risk management level.

There are two ways to combine CSM and loan officer's subjective appraisals. One is superposition method, which adds a scoring model for filter to the original loan program. Just like Schreiner pointed out, using the scoring model to screen customers who have passed loan officers' audits. In this way, credit scoring model is equivalent to a firewall.

The other is embedding method, which does not change the dominant position of the loan officer, but embeds CSM in loan officer's decision-making process. First, CSM gives a score. Next, loan officer focuses on customers near the critical value and then further information will be collected for analysis and decision in detail. As for customers near the critical value, the scope will be up to the risk preference and cost-benefit assess of the microfinance institution. First National Bank of Chicago uses this method: combining subjective appraisals, in accordance with the scoring model, loan officer reanalyzes suspicious customers, and 25% of customers rejected by CSM will get a loan after reevaluation, at the same time, about 25% of customers accepted by CSM will finally be refused by loan officer.

Embedding method is more suitable for introducing CSM into microfinance institution. It has flexibility, and model's function in decision-making process can be improved gradually. Integration with microfinance institution's loan policy is also feasible. At the beginning of the introduction of CSM, model reliability is uncertain; a large range of critical value can be set. When the model effect gradually confirmed, the range can be appropriately reduced to lower the cost.

4. An Example of Credit Risk Assessment Model of a Microfinance Institution in Jiangsu Province

4.1 Choose Model and Process Data

Due to limited amount of samples, we use Linear Discriminant Analysis (LDA) method. LDA is the earliest statistical model applied to personal credit scoring model and is considered to be one of the most widely used statistical techniques in the field of classification models today (Sung, Chang, & Lee, 1999). The sample comes from a microfinance institution in Jiangsu Province, from establishment time October 2009 to June 2014 for a total of 4 years and 8 months. The total number of sample is 393, which contains 24 default records (three months overdue) and 369 normal records. This company's customers are given priority to small enterprises, and its industry covers agricultural and sideline products manufacturing, wholesale business, building materials industry, trade and service industry, etc. Its data includes basic information and related certificates provided by borrowers, bank card transactions, records and other copies, financial management data such as financial statement, purchasing and shipping documents etc., and report submitted by credit manager after field study.

To identify variables, we use the set of variables Viganò (1993), Schreiner (2004) and Dinh and kleimeier (2007) as the basic variable group, then adjust them according to their data quality, and ultimately determine 16 initial variables, see Table 1. There are two points worth noting: First, the number of financial data provided is small (less than one third) and the data is incomplete, so the reliability is difficult to guarantee. Therefore, we don't choose financial data variables. Second, the microfinance institution is not a deposit institution. Most of the information of deposit accounts is missing, so we have no initial variables in this respect compared to Viganò (1993) and Dinh and kleimeier (2007). In conclusion, we are closer to Schreiner (2004) in the identification of initial variables, mainly basing on the qualitative data of borrowers, such as education, housing and so on. These characteristics can reflect the borrower's repayment ability and willingness to some extent. The data was collected by authors and managers from the MFI. The descriptive statistics of the data is shown in table 2.

Table 1. Initial variables of the model

Classification	Specific variables
Related to personal information:	age, gender, marriage, education;
Related to business and financial situation:	bankcredit, housing, existyear (year of existing business);
Related to past corporation:	oldclient;
Related to Borrowing situation:	loanpurpose, loanamount, loanperiod, rate, collateral, pledge, guarantee, full (whether collateral of full value or not).

Table 2. Descriptive statistics of the variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	393	39.32	9.24	23	63
Gender	393	0.84	0.37	0	1
Marriage	393	0.13	0.42	0	2
Education	393	1.52	1.16	0	4
Bankcredit	393	0.09	0.33	0	2
Housing	393	0.09	0.29	0	1
Existyear	392	7.58	5.30	0.083	30
Oldclient	393	0.43	0.50	0	1
Loanpurpose	393	0.03	0.16	0	1
Loanamount	393	52.86	61.41	1.9	500
Loanperiod	393	10.30	2.99	1.33	12
Rate	393	1.43	0.11	1.17	1.53
Collateral	393	0.49	0.50	0	1
Pledge	393	0.91	0.28	0	1
Guarantee	393	0.27	0.45	0	1
Full	393	0.52	0.50	0	1

4.2 Choose Technology to Handle Variables

The technical problems that need to be solved in the process of handling variable include: 1) the method of variable assignment, that is, direct method or grouping method; 2) whether variable dimensions are reduced or not, that is, adopting all the variables or selecting some of them. To analyze the performance in different conditions, establishing three experimental samples (original sample and two default samples of large proportion); and then separately using direct method and grouping method to assign values, totally receiving 6 modeling data tables; finally for 6 data tables, respectively using all variable method and step-by-step method (Note 2), 12 discrimination functions (scoring model) are received. When comparing the classification effects of the model, we depend on two widely used indexes: the percentage of correctly classified bad loans PCC_{bad} (PCC , percentage of correctly classified loans) and the percentage of correctly classified good loans PCC_{good} . Comparing the two indexes, PCC_{bad} is much more important than PCC_{good} because if a good customer is wrongly judged as a bad customer, the microfinance institution only loses the opportunity cost of funds, but if a bad customer is wrongly judged as a good customer, the loss may be the entire principal. Therefore, when comparing the discriminant effects, PCC_{bad} is mainly used, and PCC_{good} is auxiliary. The calculation method of the two indexes is shown in Table 3, and the prediction effects under different technologies are shown in Table 4, 5, 6, 7.

Table 3. Percentage of correctly classified borrowers and its calculation method

Actual observation value	Predicted value		PCC_{good}	PCC_{bad}
	good	Bad		
good	Gg	Gb	$PCC_{good}=Gg / (Gg+Gb)$	$PCC_{bad}=Bb / (Bb+Bg)$
bad	Bg	Bb		

From different angles, the classification effects of the above 12 discrimination functions are compared, and the appropriate assignment method and variable determination method will be determined. Table 4 and 5 list the classification results of the three samples under direct method and grouping method respectively. Table 4 shows the results of all variable method and Table 5 shows the results of step-by-step method.

Table 4. Comparison of classification results under direct method or grouping method (all variable method) unit: %

	Original sample		Sample one		Sample two		Synthesis	
	PCC _{bad}	PCC _{good}						
Direct method	58.33	98.64	70.83	95.34	75.00	98.59	68.05	97.52
Grouping method	58.33	96.21	62.50	98.84	66.67	97.18	62.50	97.41

Table 5. Comparison of classification results under direct method and grouping method (step-by-step method) unit: %

	Original sample		Sample one		Sample two		Synthesis	
	PCC _{bad}	PCC _{good}						
Direct method	58.33	98.37	62.50	97.67	62.50	98.59	61.11	98.21
Grouping method	58.33	95.93	58.33	96.51	58.33	98.59	58.33	97.01

Table 4 shows that under all variable method, direct method has a better classification effect. In original sample and sample two, the two indexes under direct method are higher than or equal to that under grouping method. In sample one, PCC_{bad} under direct method is also higher than that under grouping method, only PCC_{good} is the opposite. The calculation of the synthesis is getting through the average value of the three samples classification indexes and reflects the overall classification effect. In synthesis, the two PCC indexes under direct method are superior to those under grouping method.

Table 5 shows that under step-by-step method, direct method performs better. In the three samples and the synthesis, both PCC_{bad} and PCC_{good} under direct method are equal to or higher than those of grouping method. Combining Table 4 with Table 5, as for assignment method, the effect of direct method is significantly better than grouping method. Therefore, direct method should be used to establish CSM in this paper.

The comparison between Table 4 and Table 5 can reflect the advantage and disadvantage of direct method and grouping method. In order to make the comparison clearer, we use Table 6 and Table 7 to give the results of the all variable method and step-by-step method. Table 6 shows the comparison under direct method, and Table 7 shows the comparison under grouping method.

Table 6. Discriminant effects under all variable method and step-by-step method (direct method) unit: %

	Original sample		Sample one		Sample two		Synthesis	
	PCC _{bad}	PCC _{good}						
all variable method	58.33	98.64	70.83	95.35	75.00	98.59	68.06	97.53
step-by-step method	58.33	98.37	62.50	97.67	62.50	98.59	61.11	98.21

Table 7. Discriminant effects under all variable method and step-by-step method (grouping method) unit: %

	Original sample		Sample one		Sample two		Synthesis	
	PCC _{bad}	PCC _{good}						
all variable method	58.33	96.21	62.50	98.84	66.67	97.18	62.50	97.41
step-by-step method	58.33	95.93	58.33	96.51	58.33	98.59	58.33	97.01

Table 6 shows that if direct method is used, the overall effect of all variable method is slightly better than that of step-by-step method. In original sample and sample two, PCC indexes under all variable method are higher than or equal to those under step-by-step method. In sample one, both methods have their own advantage. PCC_{bad} under all variable method is obviously higher than that under step-by-step method, and PCC_{good} is just the opposite. In synthesis, PCC_{bad} under all variable method is significantly higher than that under step-by-step method, while PCC_{good} under step-by-step method is slightly higher than that under all variable method. From Table 6, we can see that in synthesis, PCC_{bad} under all variable method exceeds that under step-by-step method by nearly 6 percentage, while PCC_{good} under all variable method is only less than 1 percentage lower than step-by-step method. In addition, PCC_{bad} is far more important than PCC_{good}. Therefore, from Table 6, we can see that the classification effect of all variable method is better.

The results in Table 7 show that all variable method works better than step-by-step method if we use grouping method. In the three samples and the synthesis, except PCC_{good} in sample two, indexes under all variable method are all higher than or equal to those under step-by-step method. Therefore, combining Table 6 with Table 7, for the data presented in this paper, all variable method is more appropriate when establishing a CSM.

Taking those result into consideration, we developed a CSM based on the data of this microfinance institution, employing direct method and all variable method.

When determining the final model, it is also necessary to determine the appropriate sample structure. We randomly select some records from the normal records according to the proportion of 1/7, 1/6, 1/5, 1/4, 1/3, plus all default samples, constructing ten samples. Modeling and comparing predictive results in accordance with the above mentioned handling variables technologies, the results are shown in Table 8:

Table 8. Comparison of samples of different structure and discriminant effects

	Number of record	Percent of default samples (%)	PCC_{bad} (%)	PCC_{good} (%)
Sample 1	59	40.68	79.17	88.57
Sample 2	77	31.17	79.17	92.45
Sample 3	81	29.63	75.00	94.74
Sample 4	88	27.27	75.00	92.19
Sample 5	90	26.67	70.83	95.45
Sample 6	95	25.26	75.00	98.59
Sample 7	115	20.87	70.83	96.70
Sample 8	116	20.69	66.67	97.83
Sample 9	145	16.55	58.33	99.17
Sample 10	147	16.33	66.67	99.19

Table 8 shows that there are significant differences in classification accuracy index under different sample structures. The higher the default ratio is, the better the classification effect of the bad loan is, and the lower the classification effect of the good loan is. The relationship between PCC and the sample structure can be represented by Figure 1. Regression analysis shows that both PCC_{Bad} and PCC_{Good} have significant correlations with the sample structure, see formula (1) and (2). The two equations and coefficients pass the significance test.

$$Y_{PCC_{bad}} = 52.652 + 0.746structure \quad (1)$$

$$Y_{PCC_{good}} = 106.653 - 0.438structure \quad (2)$$

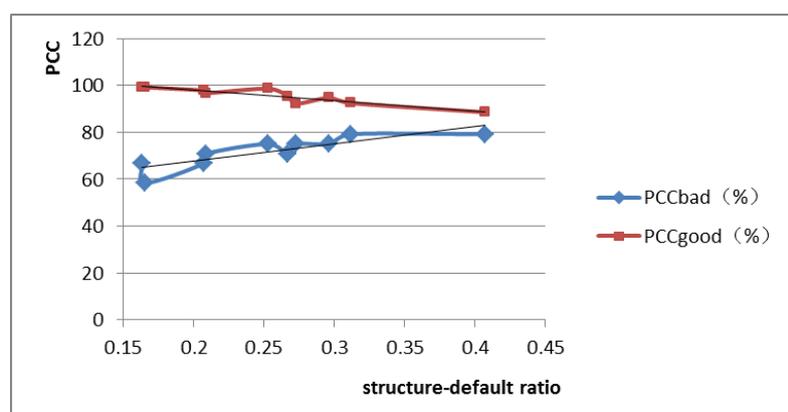


Figure 1. Relationship between PCC and sample structure

Therefore, the determination of the sample structure and final scoring model is based on the two indexes, combined with the company's loan policy, funding opportunity cost and loan loss cost. Using abovementioned method, we can determine the appropriate sample structure: default samples account for about 25%, the number of total records is 95, including 71 normal records, 24 default records, no missing value. Using SPSS for discriminant analysis and then establishing the scoring model, the model's non-standard discriminant function see formula (3):

$$\begin{aligned}
Y = & 4.500 + 0.027 \times \text{age} + 0.537 \times \text{gender} - 0.432 \times \text{marriage} + 0.548 \times \text{education} - 1.952 \times \text{bankcredit} \\
& + 0.024 \times \text{housing} + 1.101 \times \text{oldclient} + 0.028 \times \text{existyear} - 0.001 \times \text{loanamount} - 0.551 \times \text{loanpurpose} \\
& + 0.022 \times \text{loanperiod} - 4.431 \times \text{rate} - 0.141 \times \text{collateral} - 0.671 \times \text{pledge} + 0.713 \times \text{guarantee} - 0.546 \times \text{full}
\end{aligned} \quad (3)$$

To compare the contribution of each variable further and find out variables needing to be focused on, we get the standardized discriminant function, see formula (4):

$$\begin{aligned}
Y = & -0.680 \times \text{bankcredit} + 0.592 \times \text{education} + 0.504 \times \text{oldclient} - 0.437 \times \text{rate} + 0.307 \times \text{guarantee} \\
& - 0.250 \times \text{full} + 0.242 \times \text{age} - 0.190 \times \text{marriage} - 0.185 \times \text{pledge} + 0.180 \times \text{gender} + 0.136 \times \text{existyear} \\
& - 0.08 \times \text{loanpurpose} - 0.067 \times \text{loanamount} - 0.067 \times \text{collateral} + 0.066 \times \text{loanperiod} + 0.008 \times \text{housing}
\end{aligned} \quad (4)$$

The four variables that have the greatest impact on the discriminant effect are as following: Bank credit, Education, Old client and Rate.

Customers who have default records in bank, with low educational levels and new, are more likely to default and they require additional attention, which is consistent with loan officer's experiential appraisals. In addition, customers with high lending rates are more likely to default. Lending rate is set according to the company's customer credit evaluation, so this variable, on the one hand, reflects that the company's credit evaluation of customer is relatively accurate; on the other hand, it also suggests that lending rate itself have influence on customer's repayment ability and willingness.

Formula (4) also shows that our CSM's predictive variables reflect the unique characteristic of China's microfinance market. Compared with other countries' microfinance models, there are similarities and differences. The most important predictive variable in Dinh and kleimeier (2007) are Time with bank, Gender, Number of loans, Loan duration and Savings account. What the two models have in common is that the relationship between customers and company is very important. Time with bank and Number of loans in Dinh and kleimeier (2007) model, Old client in our model all reflect the previous relationship between customers and lending institutions. This to some extent confirms the relatively popular view in the field of microfinance that microfinance is of some particularity and its information is opaque and proprietary that needs to be obtained through close contact between institutions and customers. A significant difference between our model and foreign models is gender variable. Gender in both Dinh and kleimeier (2007) and Schreiner (2004) is an important variable, and female lenders have lower default rates. However, the gender variable in our model is not very important, ranking tenth in 16 predictive variables, and it is less likely for male borrowers to default.

Model's discriminant and classifying ability are shown in Table 9. Indexes of PCC_{Bad} and PCC_{Good} are 75.00% and 98.59% respectively, and compared with the same international models, the classification effect is good. The PCC_{Bad} in Viganò (1993) was 91.84%, higher than 75.00% in the present model, but the PCC_{Good} was only 62.75%, lower than 98.59% in the present model. PCC in Dinh and Kleimeier (2007) were 97.74% and 75.06%, basically in line with our model. The highest PCC attainable in Schreiner (2004) were that: PCC_{good} was 99% and PCC_{bad} was 71%, also close to our model.

Table 9. Discriminant effect of final model

		Default or not	Member of prediction group		footing
			0(normal)	1(default)	
observation	count	0	70	1	71
		1	6	18	24
	%	0	98.59	1.41	100.00
		1	25.00	75.00	100.00

5. Conclusion

In microfinance institutions, credit risk is the major risk. Regardless of whether microfinance institutions is willing to accept, measuring and managing credit risk in a quantitative way will become a trend in this sector. As early as in 1996, almost 97% of American banks used CSM to process credit card loan applications, 70% used CSM to assess small enterprise loans. As mentioned above, the practices in developing countries have also proved that the use of CSM greatly improved the judgment and also provided a guideline to assess default risk. CSM results reflect a true picture and provide more reliable, effective and efficient clear situation for management. Introducing CSM, measuring and managing the credit risk in a quantitative way can significantly reduce costs, loan evaluation time and loan officer's effort, and in the long run, it will certainly become a

powerful tool for microfinance institutions to improve efficiency and competitiveness.

However there are still some limitations of this study. Firstly, similarities and differences of CSM modeling techniques need to be identified for MFIs with different target customers. MFIs usually aim at three types of customers: peasant households, small enterprises as in our example, and some special customers like Taobao online merchants of Ali MFI. How should the CSMs be modified and revised depending on different customers have not been examined yet in our study. Secondly, the robustness of the result still has not been done. The reliability of conclusion that Bank credit, Education, Old client and Rate variables have the greatest impact on credit risk still needs to be consolidated by more MFI data in China. The limitations have pointed out the directions for our future work.

Acknowledgments

This research was financially supported by Key Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province under Grant No.2017ZDIXM085, Jiangsu Province Social Science Foundation under Grant No.14EYB006 and Youth Science and Technology Innovation Fund of NUAA under Grant No.1009-XAB16002.

References

- Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*, 40, 356-364. <https://doi.org/10.1016/j.eswa.2012.07.051>
- Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, 16(5), 471-495. <https://doi.org/10.1016/j.irfa.2007.06.001>
- Mester, L. J. (1997). What's the point of credit scoring? *Business Review*, Sep/Oct(Sep), 3-16.
- Rhyne, E., & Christen, R. (1999). *Microfinance enters the marketplace*. USAID Microenterprise Publications.
- Schreiner, M. (2004). Scoring arrears at a microlender in Bolivia. *Journal of Microfinance*, 6, 65-88.
- Shusong, B., Yongfeng, W., & Xingliang, S. (2012). The status quo and reform trend of China's microfinance institutions. *Finance Forum*, 198(6), 18-25. <http://dx.doi.org/10.16529/j.cnki.11-4613/f.2012.06.009>
- Sung, T. K., Chang, N., & Lee, G. (1999). Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, 16(1), 63-85. <https://doi.org/10.1080/07421222.1999.11518234>
- Viganò, L. (1993). A credit scoring model for development banks: an African case study. *Savings & Development*, 17(4), 441-482.

Notes

Note 1. The data come from the PBOC website.

Note 2. It should be pointed out that as described in Section 2, there are diverse methods to reduce variable dimensions. We have separately tried several methods and the results show that it is most suitable to reduce dimensions by step-by-step method. When using correlation method for dimension reduction, too few variables pass the partial correlation test, in most cases only 2-4 variables, so the method is not suitable for the establishment of the scoring model. When attempting to use principal component analysis or factor analysis, none of the 6 data tables can pass the KMO and Bartlett tests, so these methods cannot be used. Therefore, we eventually chose the step-by-step method for dimension reduction.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).