ISSN 2616-5902 Vol. 3, Issue 6: 83-87, DOI: 10.25236/AJBM.2021.030612

Research on credit risk of commercial banks based on multiple logistic model

Jiexin Lu[1], Yongzhen Tong[2]

[1]*School of Electronic Science and Engineering, Xiamen University, Xiamen, Fujian, 361005, China*
[2]*School of Architecture and Civil Engineering, Xiamen University, Xiamen, Fujian, 361005, China*

Abstract: *This paper establishes a bank credit decision-making system for small and medium-sized enterprises. This paper quantifies the risk of enterprises with existing credit records and reputation ratings and establishes a risk rating model. The specific financial situation of 123 enterprises with credit records and credit rating is quantified and a model is established. In this paper, a large number of original data are processed and integrated by EXCEL and MATLAB software to get six indexes, and then three representative principal component factors are extracted by principal component analysis, which are used as independent variables for binary Logistic regression analysis, the evaluation of whether the enterprise is in breach of contract is obtained, the default screening model is established, the grade of the enterprise in breach of contract is rated as IV, and then the enterprise without default is analyzed by multivariate ordered Logistic regression analysis. Establish a risk level refinement model to further refine the non-default corporate rating, that is, I-level, II-level and III-level.*

Keywords: *Bank credit system, Risk level, Principal component analysis, Logistic regression*

1. Introduction

At present, China's economy is developing steadily, productivity has been greatly improved, and capital forms are more diversified. The number of small, medium and micro enterprises is increasing with the development of China's economy. However, most small and micro enterprises are in the initial stage of small scale, so they need financial support from commercial banks. However, due to incomplete financial data of SMEs and the lack of collateral assets recognized by the bank, the existing risk assessment and credit decision-making model of the bank is not perfect. Therefore, the bank will also provide loans to enterprises with stable transactions and strong strength according to the national credit policy, corporate financial data and the influence of relevant enterprises. And to high credit, credit risk of small enterprises to give preferential interest rates.

2. Principal component analysis

As a statistical factor analysis method used for dimensionality reduction, a group of correlated variables can be obtained through orthogonal transformation to obtain a group of mutually independent variables. This group of mutually independent variables is the principal component, which is a comprehensive index used to explain the original data [1].

$x_1, x_2, x_3, x_4, x_5, x_6$ respectively represent the six indicators of cost of sales, sales revenue, total value added tax, business scale, transaction failure rate and default. $c_1, c_2, c_3, c_4, c_5, c_6$ represent the weight of each variable, then the weighted sum is:

$$S = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5 + c_6x_6 \tag{1}$$

Each enterprise has a comprehensive score, which is denoted as $S_1, S_2, \cdots, S_{123}$. It is necessary to find an appropriate weight to spread $S_1, S_2, \cdots, S_{123}$ as far as possible.

Let $X_1, X_2, X_3, X_4, X_5, X_6$ represent random variables with $x_1, x_2, x_3, x_4, x_5, x_6$ as the observed values of the sample, and find $c_1, c_2, c_3, c_4, c_5, c_6$:

ISSN 2616-5902 Vol. 3, Issue 6: 83-87, DOI: 10.25236/AJBM.2021.030612

$$Var\left(c_1X_1+c_2X_2+c_3X_3+c_4X_4+c_5X_5+c_6X_6\right) \tag{2}$$

The value reaches the maximum. Usually rules:

$$c_1^2+c_2^2+c_3^2+c_4^2+c_5^2+c_6^2=1 \tag{3}$$

Under this constraint, find the optimal solution of Equation (2).

Let F_n represent the nth principal component, $n=1,2,\cdots,6$ which can be set

$$\begin{cases} F_1=c_{11}X_1+c_{12}X_2+c_{13}X_3+c_{14}X_4+c_{15}X_5+c_{16}X_6 \\ F_2=c_{21}X_1+c_{22}X_2+c_{23}X_3+c_{24}X_4+c_{25}X_5+c_{26}X_6 \\ \vdots \\ F_6=c_{61}X_1+c_{62}X_2+c_{63}X_3+c_{64}X_4+c_{65}X_5+c_{66}X_6 \end{cases}. \tag{4}$$

In the formula, $c_{n1}^2+c_{n2}^2+c_{n3}^2+c_{n4}^2+c_{n5}^2+c_{n6}^2=1$ for each n, and $\left[c_{11},c_{12},c_{13},c_{14},c_{15},c_{16}\right]$ maximized the value of $Var\left(F_1\right)$; $\left[c_{21},c_{22},c_{23},c_{24},c_{25},c_{26}\right]$ is not only perpendicular to $\left[c_{11},c_{12},c_{13},c_{14},c_{15},c_{16}\right]$ but also maximizes the value of $Var\left(F_2\right)$; $\left[c_{31},c_{32},c_{33},c_{34},c_{35},c_{36}\right]$ is simultaneously perpendicular to $\left[c_{11},c_{12},c_{13},c_{14},c_{15},c_{16}\right]$ and $\left[c_{21},c_{22},c_{23},c_{24},c_{25},c_{26}\right]$ and maximizes the value of $Var\left(F_3\right)$.And so on for all the principal components.

Using SPSS software to realize the above discussion, the following analysis results are obtained.

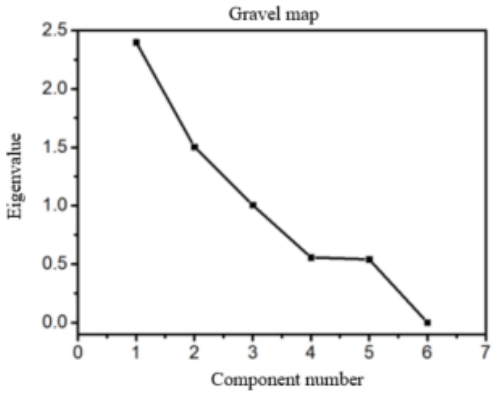

Figure 1: SPSS analysis results - lithographs

According to the gravel diagram, the eigenvalues of F1, F2 and F3 are greater than 1, so they can be used as main components.

3. Binary Logistic regression analysis - default screening model

Logistic regression is a generalized linear regression analysis model, which is often used in the field of economic forecasting. Its independent variables can be continuous or classified. The dependent variables of binary Logistic model are classified data, such as default (represented by 0) and non-default (represented by 1) in this paper [2] [3].

The basic form of Logistic model function regression is as follows:

$$P=\frac{1}{1+e^{-(Q)}} \tag{5}$$

Where, P value is the probability of whether the enterprise output has defaulted after the Logistic model was run, and P=0.5 is selected as the threshold value of whether the enterprise has defaulted. If the output P value of the model is less than 0.5, the enterprise is considered to have defaulted; if the

ISSN 2616-5902 Vol. 3, Issue 6: 83-87, DOI: 10.25236/AJBM.2021.030612

output P value of the model is greater than 0.5, the enterprise will not default. The expression of Q is as follows:

$$Q = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 \tag{6}$$

In the above formula, F1,F2 and F3 are the three principal component factors extracted by principal component analysis in (1), which are taken as the input variables of the Logistic regression model. The actual default situation of 123 small, small and micro enterprises in Attachment 1 is taken as Q in the equation for regression, and the regression results using the binary Logistic model are as follows:

Table 1: The result of binary Logistic regression Analysis-- Omnibus Test

Omnibus Test of Simulation coefficient				
		Chi-square	Degree of freedom	Significance
Step1	Step	108.438	3	.000
	Block	108.438	3	.000
	Model	108.438	3	.000

The fitting method in this paper was input type. The Omnibus test based on the results of binary Logistic regression analysis showed that before the model fitting, the overall significance was less than 0.05, indicating the existence of correlation. Therefore, it was reasonable to use this model.

Table 2: Results of Binary Logistic Regression Analysis -- Accuracy Test

Classification table					
	Actual measurement		Forecast		
			Rating		Correct percentage
			0	1	
Step1	Rating	0	23	1	95.8
		1	2	97	98.0
	Overall percentage				97.6
Demarcation value 0.500					

According to the above table, in the process of model fitting prediction, the prediction accuracy of default (0) is 95.8%, the prediction accuracy of non-default (1) is 98.0%, and the accuracy of overall rating is 97.6%, indicating that the prediction effect of this model is good.

Table 3: Results of Binary Logistic Regression Analysis -- Coefficients in the Regression Equation

Variable									
		B	Standard deviation	Ward.	Degree of freedom	Significance	EXP(B)	95% confidence interval of EXP (B)	
								Lower limit	Upper limit
Step	F1	20.855	23.133	0.813	1	0.067	1141248789.938	0.000	5.59E+28
	F2	7.607	7.518	1.024	1	0.012	2012.381	0.001	5044682607.066
	F3	4.517	3.586	1.587	1	0.108	91.582	0.081	103259.116
	F4	10.993	4.899	5.035	1	0.025	59456.390		

The above table gives the results of Logistic regression coefficients of the three principal component factors. The coefficients of principal component factors F1, F2 and F3 are all greater than zero and significant at the 95% confidence interval, indicating that these factors are negatively correlated with corporate credit default risk. According to the above table, the default screening model of small, medium and micro enterprises can be expressed as:

$$P = \frac{1}{1 + e^{-(10.993 + 20.885F_1 + 7.607F_2 + 4.517F_3)}} \tag{7}$$

Substituting the principal component factors F1, F2 and F3 of 123 enterprises into the default screening model, the default probability of 123 enterprises and the predicted default situation of 123 enterprises are obtained

4. Binary Logistic regression analysis - default screening model

The 26 defaulting enterprises were eliminated, and the risk level of these 26 defaulting enterprises

ISSN 2616-5902 Vol. 3, Issue 6: 83-87, DOI: 10.25236/AJBM.2021.030612

was rated as Level IV. Multiple ordered Logistic regression analysis was carried out on the remaining 97 enterprises to further evaluate the level of non-defaulting enterprises and establish a detailed risk level model [4] [5].

In this multivariate ordered Logistic regression analysis, the input variables are F1, F2 and F3, the three principal component factors extracted by principal component analysis in (1), and the output variables are the original rating (i.e., A, B, C) of 97 non-defaulting companies in Annex 1.The following are the results of multiple ordered Logistic regression analysis of the 97 non-defaulting companies using SPSS:

Table 4: Multivariate ordered Logistic regression analysis results--model goodness of fit

	Chi-square	Degree of freedom	Significance
Pearson	183.299	189	0.603
Deviation	191.003	189	0.446
Correlation function: negative double logarithm			

Table 5: Multivariate ordered Logistic regression analysis results -- parallel line test of the model

Parallel line test				
Model	-2 logarithmic likelihood	Chi-square	Degree of freedom	Significance
Original hypothesis	191.003			
Routine	134.052b	56.950c	3	0.062
The location parameters are the same in each response category				
Correlation function-negative double logarithm				
After reaching the maximum step-by-step dichotomy, the logarithmic likelihood cannot be further increased.				
The calculation of chi-square statistics is based on the logarithmic likelihood obtained by the last iteration of the general model, and the validity of this test is uncertain.				

In Table 5, the P value of Pearson and Deviance, the statistics representing goodness fit of the model, is 0.603 and 0.446, both greater than 0.05, which proves that the model fits well. In addition, the significance of parallel line test of the model is 0.062, greater than 0.05, which proves that it can pass the parallel line test.

Since the multiple ordered Logistic regression analysis is a cumulative regression model, the interpretation of its coefficients can only be based on the statistical significance and statistical direction. This model estimates the cumulative probability, and its estimated parameter is the cumulative probability ratio, so the effect size of variables cannot be directly expressed by probability, but only the cumulative probability size can be obtained [6].

Table 6: Multivariate ordered Logistic regression analysis results -- parameter estimates

Variable								
		Standard deviation	Ward.	Degree of freedom	Estimate	Significance	95% confidence interval of EXP (B)	
							Lower limit	Upper limit
Threshold value	1	0.215	2.283	1	-0.324	0.131	-0.745	0.096
	2	0.245	15.631	1	0.969	0.000	0.489	1.449
Step	F1	1.108	10.330	1	-3.561	0.001	-5.733	-1.389
	F2	0.372	0.166	1	0.151	0.684	-0.577	0.880
	F3	0.192	1.567	1	1.567	0.211	-0.136	0.617
Correlation function: negative double logarithm								

It can be seen from Table that the coefficient before the principal component factor F1 is -3.561, indicating that the higher the value of F1 is, the lower the risk level of the assessment result will be. The coefficient before the principal component factor F2 is 0.151, indicating that the higher the value of F2 is, the higher the risk level of the assessment results will be. The coefficient before the principal component factor F3 is 0.240, indicating that the higher the value of F3, the higher the risk level of the assessment results.

According to the above table, the risk grade refinement model is established as follows:

ISSN 2616-5902 Vol. 3, Issue 6: 83-87, DOI: 10.25236/AJBM.2021.030612

$$\begin{cases} p_1 = \dfrac{1}{1+\exp\left(-\alpha_1+\beta X\right)} \\ p_2 = \left(p_1+p_2\right)-p_1 = \dfrac{1}{1+\exp\left(-\alpha_2+\beta X\right)} - \dfrac{1}{1+\exp\left(-\alpha_1+\beta X\right)} \end{cases} \tag{8}$$

$$\beta X = -3.561F_1 + 0.151F_2 + 0.240F_3 \tag{9}$$

In this paper, since the risk assessment levels of non-defaulting enterprises analyzed by multiple ordered Logistic regression are divided into three levels: I, II and III, only two cumulative models are needed to be established. The probability of the third level is 1 minus the sum of the probabilities of the first two levels, namely:

$$p_3 = 1-p_1-p_2 = 1-\frac{1}{1+\exp\left(-\alpha_2+\beta X\right)} \tag{10}$$

Substituting the principal component factors F1, F2 and F3 of 97 non-defaulting enterprises into the above risk grade refinement model, the risk grade evaluation was obtained. By comparing with the original rating, the accuracy of the model was 62%. There were 30 enterprises rated as I, 52 enterprises rated as II and 15 enterprises rated as III.

5. Conclusion

In this paper, principal component analysis is used to effectively extract principal component factors to quantify the specific financial situation of enterprises with existing credit records and reputation ratings, and a model is established, which eliminates the correlation between independent variable indicators. Principal component analysis is used to extract three representative principal component factors, which are used as independent variables for binary Logistic regression analysis. Then carry on the multivariate ordered Logistic regression analysis to the non-default enterprises, establish the risk grade refinement model, and further refine the non-default enterprise rating, namely I level, II level and III level. In this paper, six useful indexes are extracted effectively, and then simplified to three indexes according to the principal component analysis method, which greatly simplifies the solution of the problem.

References

[1] Shoukui Si, Zhaoliang Sun, Mathematical Modeling Algorithms and Applications, Beijing, National Defense Industry Press, 2015.
[2] Wang Yue, A Comparative Study of Double Model Forecasting Method for Credit Risk Identification of Listed Private Enterprises, Master Thesis of Inner Mongolia University of Science and Technology, 2020.
[3] Ma Guangjun, Research on Credit Risk Evaluation of Urban Investment Bond in China Based on Multivariate Orded Logistic Model, Master Thesis of Tianjin University of Finance and Economics, 2012
[4] Guo Qian, A Study on Specific Risk Adjustment Coefficients in the Valuation of Unlisted Companies, Master Dissertation of Beijing Jiaotong University, 2015
[5] Bian Yun, A study on hierarchical diagnosis model of chronic pancreatitis based on factor analysis with multiple ordered Logistic regression, PhD Dissertation of Second Military Medical University, 2016
[6] Zheng Fangyun, A Study on Forecasting Financial Distress of Listed Companies in China -- Based on Ordered-Triclass Logistic Regression, Master Dissertation of Zhejiang University, 2013.