

The Evaluation Model of Enterprise Credit Default Based on Logistic Regression

Guoqing Wang¹, Xialing Wu², Chenxi Xu³

¹*School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, 233000*

²*School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, Anhui, 233000*

³*School of International Economics and Trade, Anhui University of Finance and Economics, Bengbu, Anhui, 233000*

Keywords: Logistic regression, structural equation, decision tree, AHP under cloud model, cointegration neural network

Abstract: Since entering the new economic normal, my country's small, medium and micro enterprises have continued to reform and develop and become an important part of the real economy. The bank will conduct an overall assessment of SMEs and formulate corresponding credit strategies based on the business strength, development potential, creditworthiness, and default risks of the enterprise. While providing financial support to SMEs as much as possible, the bank will achieve interest rates on bank credit. Maximize. This paper establishes a Logistic regression model to quantify the default records and credit ratings of small and medium-sized enterprises, and builds a credit default evaluation model for enterprises; and then uses structural equations to reflect the relationship between corporate strength, default risk, bank credit lines, and bank credit interest rates. Formulate specific credit strategies. Use AHP improved by the cloud model to screen the industry to which the company belongs; then establish a change-point model to analyze the volatility of the company's operating conditions and default risks when affected by emergencies, and to fit the evolution path of different companies in the future ; Finally, through the establishment of a neural network model under the co-integration theory, the credit default risk is predicted, and the degree of impact of different industries is ranked, thereby adjusting the bank's credit strategy.

1. Background of the problem

After more than 40 years of reform and opening up, my country's economy has undergone tremendous changes and development, and small, medium and micro enterprises have fully integrated into the wave of market economy. Since entering the new economic normal, my country's economy has been generally stable and the drive for innovation has increased, but it has also faced challenges brought by many factors such as Sino-US trade frictions and the new crown pneumonia epidemic. Due to the characteristics of small, medium-sized and micro-enterprises, single operation content and low credit rating, the problem of "financing difficult and expensive" is particularly

obvious. The reasons for the credit problems of small, medium and micro enterprises are that, on the one hand, the structure and organization of small and micro enterprises are relatively chaotic, the management and operation are flawed, the financial system is not sound, and the awareness of credit enhancement is not strong [1]; on the other hand, it is attributed to the development of commercial banks Credit discrimination problems caused by factors such as lack of risk control awareness and imperfect credit system in credit strategy [2]. In order to solve the financing and loan problems of small, medium and micro enterprises, the existing empirical research has discussed from multiple perspectives such as optimization of corporate loan schemes, construction of bank loan credit evaluation systems, and policy-based financial support. The problem of this article is to construct an optimized plan of bank credit strategy from the perspective of credit risk.

This article considers a bank's credit risk assessment for small, medium and micro enterprises, and determines credit strategies accordingly.

2. Risk quantification based on logistic regression

The logistic regression model mainly solves the regression problem of binary dependent variables, quantitatively analyzes multiple variables, and finally obtains the weight of the explained variable. Through Logistic regression analysis, the correlation coefficient of the variable can be calculated, and the probability of the occurrence of the event can be predicted based on the influencing factors. The formula is as follows:

$$P(Y_i = 1|X_i) = \frac{e^{\alpha + \sum_{k=1}^k \beta_k x_{ki}}}{1 + e^{\alpha + \sum_{k=1}^k \beta_k x_{ki}}}$$

Where Y represents the dependent variable (0-1 variable, 0 represents the event does not occur, 1 represents the event occurs), $X = [x_1, x_2, \dots, x_k]$ is a set of dependent variables corresponding to Y. $P(Y_i = 1|X)$ represents the probability of an event occurring after a given set of values $X = [x_1, x_2, \dots, x_k]$. Determine the optimal parameter $[\alpha, \beta_1, \beta_2, \dots, \beta_n]$ by establishing a likelihood function for a given set of dependent variables Y and a set of independent variables X corresponding to it.

Assume that the probability of an event is:

$$P(Y_i = 1|X_i) = p_i$$

Then the ratio of the probability of occurrence of an event to the probability of not occurring is:

$$\frac{p_i}{1 - p_i} = e^{\alpha + \sum_{k=1}^k \beta_k x_{ki}}$$

Odds of events (odds). Odds must be positive, because $0 < p_i < 1$, we can get a linear function by taking the natural logarithm of odds:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \alpha + \sum_{k=1}^k \beta_k x_{ki}$$

Process the data and related variables, set non-default to 0 and default to 1; set corporate reputation levels A, B, C, and D to 1-4 respectively. The larger the value, the higher the credit risk; The total price and tax of the invoice minus the total price and tax of the input invoice get the capital gains and losses. The higher the value, the lower the credit risk; the customer churn rate, the annual interest rate of the loan and the loan line are set as control variables, which means the control conditions for the quantification of credit risk .

Establish a logistic regression equation based on the selection of the above variables:

$$\text{logistic}(\text{CR}) = \mu + \beta_1 P + \beta_2 \ln(1 + \text{CRAT}) + \sum \gamma_n \text{CGL} + \varepsilon_j$$

The result of the operation, with 0.5 as the threshold, according to the observation and prediction values in the following table, we can see that the prediction accuracy of the model can reach 79.36%, and the accuracy of the model is high.

Table 1 Comparison of observed and predicted values of Logistic regression

Observed		Predicted			
		Credit risk measurement			Percentage correction
		1	2	3	
Credit risk measurement	1	80	43	34	70.5
	2	36	103	46	78.8
	3	39	46	90	88.8
Percent of total		—			79.36

Through stepwise regression, the final fitting results are obtained. As shown in the following table, the significance level of the four control variables such as default, credit rating, capital gains and losses, and customer churn rate is less than 0.05, which has a significant impact on credit risk. The fitted equation is:

$$\text{logistic}(\text{CR}) = -0.318 + 0.285P - 0.587 \ln(1 + \text{CRAT}) - \sum 0.465\text{CGL} + 0.162\varepsilon_j$$

Table 2 Variables in the equation

variable	B	S.E	Wals	df	Sig.	Exp(B)	EXP(B)的 95% C.I.	
							Lower limit	upper limit
Default risk (P)	0.285	0.198	7.375	1	0.007	0.584	0.396	0.861
Credit Rating (CRAT)	-0.587	0.226	4.227	1	0.040	1.592	1.022	2.480
Capital gains and losses (CGL)	-0.465	0.130	4.804	1	0.028	1.330	1.031	1.717
Control variable (ε_j)	0.162	0.195	7.604	1	0.006	0.585	0.399	0.856
constant	-0.318	1.433	0.100	1	0.005	0.636	—	—

3. Credit decision under structural equation

Structural equation model (SEM) is a statistical analysis method that analyzes the relationship between variables based on the covariance matrix of variables. It is mainly used to deal with the relationship between multiple causes, multiple results, and problems containing latent variables [2]; Structural equation analysis can process multiple dependent variables at the same time, compare and evaluate different theoretical models, and test whether it is consistent with the data through various fitting indexes.

First, perform KMO test and Bartlett's sphere test analysis to verify whether the commonality between items is significant. The test results are as follows:

Table 3 KMO and Bartlett test results

Kaiser-Meyer-Olkin measure of sampling adequacy		.880
Bartlett's sphericity test	Approximate chi-square	3234.334
	Degree of freedom	36
	Significance Sig.	.000

First, perform KMO test and Bartlett's sphere test. The KMO test coefficient is 0.880, which is greater than the minimum test coefficient standard of 0.5. Bartlett's sphere test value is 3234.334, and the P value is less than 0.001. Therefore, the questionnaire has structural validity and there are common factors between the items. Perform factor analysis.

In order to test the structural validity of the scale, principal component analysis and maximum variance rotation method were used to perform exploratory factor analysis (EFA) on the obtained data. Through factor analysis, a total of 3 factors with feature values greater than 1 are extracted. The gravel diagram is as follows:

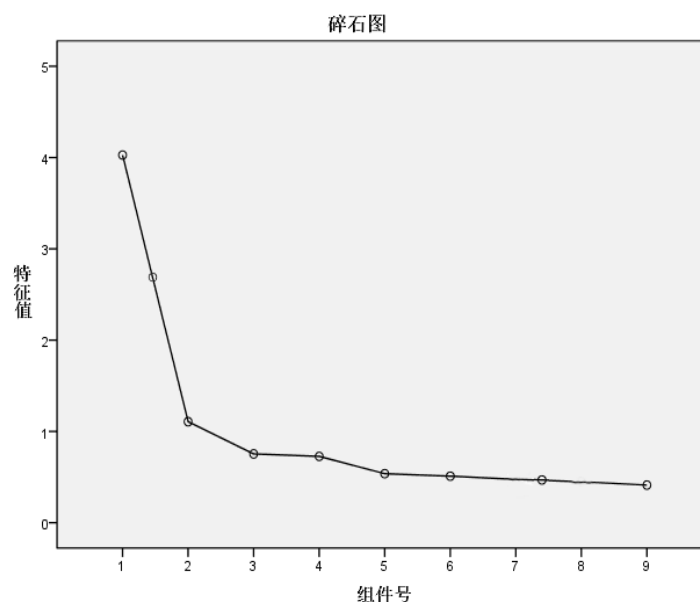


Figure. 1 Exploratory factor analysis factor characteristic root result graph (gravel graph)

According to the gravel graph, there are 3 characteristic roots greater than 1 in the correlation coefficient matrix, so three public factors are considered: credit risk, loan availability, and interest rate preference. After factor rotation, the factor loading matrix is obtained as follows:

Table 4 Factor loading matrix

Project information	ingredient		
	1	2	3
Credit Risk	.110	.826	0.159
Loan or not	.189	.789	0.356
Preferential interest rate	.300	.678	0.462

Combined with the above logistic regression analysis, the credit risk of the 123 companies in Annex 1 is quantified, and then the structural equation is used to comprehensively consider the credit risk, whether to provide loans, and interest rate concessions. Based on this, make the credit decision of small, medium and micro enterprises: Of companies give a fixed amount of 43.7% of loans, 31.8% of loans are allocated to B-level companies, and C-level companies are allocated 24.5% of loans. At the same time, loans are allocated preferentially according to ABC level and interest rate concessions are given.

4. Model evaluation and promotion

The logistic regression model can carry out a multivariate quantitative analysis on the two-category dependent variables of corporate reputation and default risk, and predict the probability of corporate credit default based on influencing factors. The improvement of the AHP method combined with the cloud model has eliminated a greater degree of subjective assumptions, and has strong representativeness and credibility for the selection of the seven major industries. The change point model uses the data of the economic development of various industries to analyze the trend of changes, which can more intuitively reflect the characteristics. The discriminant analysis function quantifies the corporate reputation levels A, B, C, and D as 4, 3, 2, 1, increasing the accuracy of the indicators, which is conducive to more rigorous judgments on the reputation of the company, making the results more reliable. The decision tree and forest algorithm can relatively simplify the calculation of the model, and get a clearer view of the importance of the various factors in the bank's credit strategy, which is conducive to formulating a reasonable and accurate credit strategy. The neural network model under the cointegration theory reveals the long-term stable non-linear equilibrium relationship between various variables, and makes corresponding loan decision adjustments for banks according to the future profitability and debt solvency of enterprises in various industries. The model has a high degree of fitting Degree and strong fit.

Based on the background of the financing difficulties of small, medium and micro enterprises and the imperfect bank credit system construction, this paper establishes a logistic regression model on the basis of discriminant analysis to quantitatively analyze the creditworthiness of small, medium and micro enterprises. At the same time, the decision tree, change point model and neural network model are used to discuss the specific adjustment plan of bank loan decision under the influence of sudden factors, so as to maximize the bank loan interest while giving as much financial support to SMEs as possible.

The model established in this paper can be widely used in real life commercial banks to evaluate the credit rating of small, medium and micro enterprises, formulate credit strategies, and solve the problems of credit strategy adjustment under the influence of sudden factors. At the same time, the analysis object of the model can also be expanded to listed companies with a certain scale, and further combined with Internet credit, the establishment of online + offline loan reputation evaluation and loan strategy formulation models.

References

- [1] Liu Ying. *Analysis of Credit Risk Management of Commercial Banks Loans to Small and Medium-sized Enterprises* [J]. *National Circulation Economy*, 2020 (10): 147-148.
- [2] Li Xinhang, Gao Yangge. *Analysis and Countermeasures of Difficulties in Financing of my country's Small and Medium-sized Enterprises* [J]. *Business Economics*, 2020 (03): 79-80+164.
- (Not found) [3] Cao Li, Lu Qiujuan, Huang Jingyu. *Application of support vector machine and logistic regression in predicting the risk of AIDS in MSM population* [J]. *China Health Education*, 2020, 36 (04): 332- 335.
- [4] Xu Lan, Fang Zhigeng. *Analysis of the influence degree of regional innovation system subsystems based on structural equations* [J]. *Science and Technology Management*, 2010 (05): 162-165.

- [5]Zhang Miaolin, Zhang Chengzhong. *Research on the Model Construction of Comprehensive Pipe Gallery Investment Operation Mode Selection [J]. Tunnel Construction*, 2019 (08): 1319-1325.
- [6] Xiao Yang. *Coal price prediction model based on cointegration theory and artificial neural network [D]. Xi'an University of Science and Technology*, 2016.