

Credit Analysis of Small and Medium Sized Enterprises based on Logistic Regression

Zhao Qiuya¹, Cheng Chi², Xu Hailin³

¹Shenzhen University School of Mathematics and Statistics, Shenzhen, Guangdong, 518000

²School of Information, Xi'an University of Finance and Economics, Xi'an, Shaanxi, 710100

³International Business Faculty, Beijing Normal University, Zhuhai, Guangdong, 519000

Keywords: Quantitative analysis; Logistic regression; TOPSIS model

Abstract: Credit is one of the indispensable economic activities in modern society. Its emergence helps a large number of enterprises, especially small and medium-sized enterprises out of the predicament, and plays an important role in the economic development of enterprises. Based on credit, this paper gives the risk quantification and bank lending strategy for small and medium-sized enterprises. This paper analyzes the risk of 123 credit enterprises. In this paper, we use the trade notes and their names to formulate appropriate data and extract the information of the upstream and downstream industries and the strength of the company. Then, we use TOPSIS model to quantify the strength and reputation of the industry and the enterprise itself. At the same time, through the analysis, we found that the breach of contract is closely related to its own factors, but not to the industry. Therefore, we use the established enterprise strength and reputation index to carry out logistic regression, and get the expression of the enterprise's own index and its forecast default probability. After forecasting the default probability of an enterprise, the default probability is taken as the result of quantitative risk. Then, using single objective programming, taking the interest rate of enterprise loan as the decision variable, the maximization of bank loan income is taken as the goal.

1. Introduction

Credit is an indispensable economic activity in modern society. It refers to the form of value movement under the condition of repayment and interest payment. It is an important form of mobilization and distribution of funds by means of compensation in socialist countries, represented by China. It is also a powerful lever for economic development. Good credit mechanism has brought "living water" to enterprises, especially small and medium-sized enterprises. In today's world economy, small and medium-sized micro enterprises are one of the important pillars of social stability and economic development for a country, and also the most dynamic and potential part of the national economy. However, as we all know, small and medium-sized enterprises have small cash flow and insufficient anti risk ability. Therefore, bank loans are very important. Credit provides a solution to the short-term shortage of funds.

However, under the background of China's economic slowdown, there are also many non-performing loans. There are many reasons for non-performing loans, such as the omission of review content, the lack of due diligence and the bank team's misjudgment. In the guarantee bank's review team, and the bank team needs to review the background of SMEs applying for loans. During the review, the bank assesses the credit risk according to its enterprise strength and reputation, and then decides whether to lend, the amount of loan, interest rate and term.

2. Problem analysis

First of all, we need to make a quantitative analysis of the loan risk of 123 enterprises, and give the credit strategy of banks in the face of these small and medium-sized enterprises. Therefore, we need to extract information from the strength of the upstream and downstream companies.

We can understand the upstream and downstream information of an enterprise as the prospect of an industry. We can learn from economics that any enterprise has an upstream enterprise or a downstream enterprise. Therefore, the upstream and downstream information of an enterprise can be obtained from its industry. That is, whether the prospect of the industry is good or not is highly related to the stability of the upstream and downstream.

Therefore, we first extract the industry of each company according to its name. After that, we can search the latest annual indicators of these industries through the Internet. For these indicators, we can use TOPSIS model to score these industries, and then map the scores to the interval of [0,1] to get the score m_1 of the industry.

After that, the TOPSIS model is also used as the scoring model. As shown above, these companies are scored as objects, and the score is mapped to the interval of [0,1], and the score value m_2 of the company's individual strength is obtained.

After getting m_1 and m_2 , we find that the default of these enterprises is obviously not related to m_1 . We use logistic

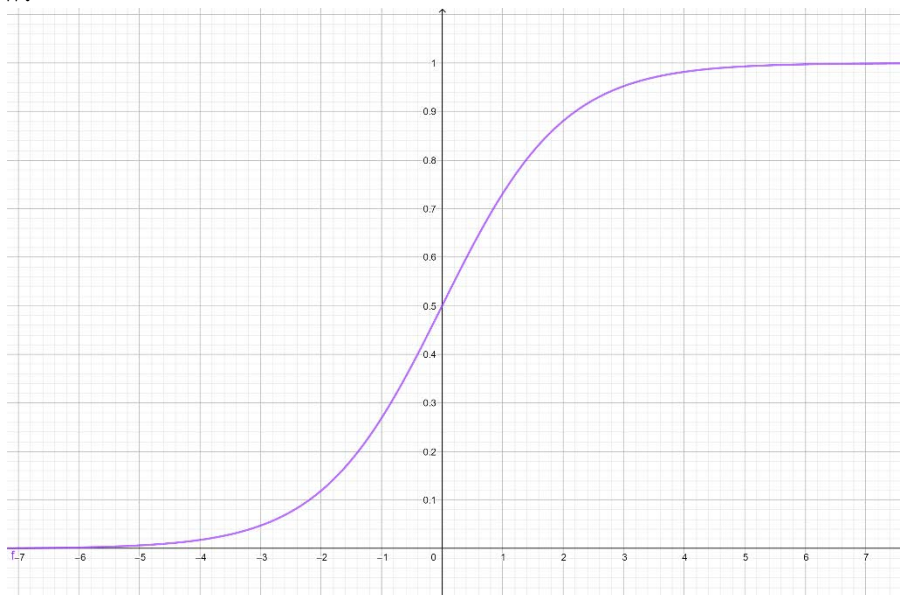
The regression model predicts the default probability p of the company, so the loan risk of 123 enterprises is quantified. After measuring the risk of enterprise loan, we use single objective linear programming to consider the maximization of enterprise loan income, and finally traverse the solution.

3. Model analysis

Logistic model, also known as logistic regression model, is a classical binary classification method in statistical learning. It depends on the maximum entropy as a criterion of the probability. For a continuous random variable x , X is said to obey the logistic distribution when x has the following distribution function:

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-(x-\mu)/\gamma}}$$

Where, μ is the position parameter and $\gamma > 0$ is the shape parameter. And its distribution image is centrosymmetric image: it grows slowly at both ends and grows faster near the center, as shown in the figure below.



For binomial logistic regression model, the model calculates two conditional probabilities of the following formula:

we mark $w = (w^{(1)}, w^{(2)}, \dots, w^{(n)}, b)^T$ are weight vectors, and $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T$

Logistic classification is to use the comparison of the two conditional probability values to classify case x into the one with higher probability value. In order to get the weight vector w , we

need to use the maximum likelihood method.

The following formula

$$\ln \frac{y}{1-y} = w^T x$$

4. Model establishment

In this paper, small and medium-sized micro enterprises are classified into 22 categories according to their names. Therefore, we can get the indicators of various industries by consulting. We detected the correlation between the indicators. Among them, we find that there is no strong correlation between variables. Therefore, we do not remove these indicators, but use all indicators in the subsequent operation.

For individual companies, we can only get their trade notes information. Therefore, we need to consider from different angles and construct indicators. The starting time of invoices and bills of different companies is different, and the length of the time period is also different. Even the longest time length is half a year longer than the shortest time length. Therefore, it is not advisable to measure the strength of a company by simply seeking sum.

We calculate the difference between the starting time of each company's invoice and define it as the days interval t . We use the sum of sales revenue divided by the days interval to get the average daily sales revenue s ; we use the sum of sales costs divided by the days interval to get the average daily sales cost C . Therefore, we can use the following formula to calculate the annual average gross profit ψ_1 , sales gross profit rate $_2$ and cost gross profit margin $_3$ (the default year is 365 days)

$$\psi_1 = 365(S - C)$$

$$\psi_2 = \frac{S - C}{S}$$

$$\psi_3 = \frac{S - C}{C}$$

After testing the correlation, we need to ensure that the distribution of the selected indicators is different in the two samples (i.e. different default situations). Therefore, we use the reputation distribution of 27 companies to test the significant difference between them. Therefore, the final index evaluation system is determined. The principle of kW (Kruskal - Wallis) test is to mix multiple samples and calculate the rank. If knotting occurs, the average rank is used. Then, the rank sum is calculated according to the sample group. We construct the following hypothesis testing questions

H_0 : The two populations are in the same position H_1 Two different overall locations

Table 1 P value table of each index

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6
P	0.00016	0.96550	0.96550	0.00397	7.71630e-05	4.36419e-08

We use SPSS software to test the hypothesis. Among them, we use SPSS to calculate the p value of the hypothesis problem of each index, and its p value represents the probability of error in rejecting the original hypothesis. We tested the hypotheses of the above six indicators, and found that the P values of ψ_2 and ψ_3 were far greater than 0.05 (i.e. at the significance level of 5%), we accepted the original hypothesis,

5. Model solving

After getting the industry score vector M^1 and the company individual score vector M^2 , we fully consider the upstream and downstream of the industry and individual companies, and use the logistic regression model to obtain the probability of default risk. Therefore, for each company, we can get the scores of the upstream and downstream industries and the individual strength of the

company.

The number of cases in which $x_i = [M^1, M^2]$ is closely related to the company's breach of contract is 10. But the value of M_1 is not related. Therefore, we conclude that it is not advisable to use M_1 to predict the probability of default.

The following results can be obtained

$$w = [-0.896, -2.82e - 08, 5.86, -0.005, -0.001]$$

Table 2 Corresponding rating and default probability of a few enterprises

Enterprise code	Credit rating	Probability of default
E1	A	0.00830
E2	A	2.22E-16
E30	B	0.03002
E41	C	0.04803
E123	D	0.87202

6. Conclusion

In this paper, six indicators reflecting the strength of enterprises are extracted from the data, and four representative indicators are selected by means of kW test, which are used as independent variables, and the default probability of small and medium-sized enterprises is analyzed by logistic model. The correct rate of regression was 74.8%. The results showed that the regression model performed well. Then, the paper points out the loan strategy of the bank when the total amount of fixed loans is fixed by the linear programming model.

References

- [1] Ying Peipei. The impact of bank credit on the development of SMEs [J]. Modern economic information, 2015
- [2] Li Yang, Chen Zhifang. Causes and control measures of credit risk of commercial banks [J]. Special Economic Zone, 2016 (11): 65 - 66
- [3] Tang Xueting. Research on credit risk -- Research on credit risk assessment of China's commercial banks for small and medium-sized enterprises [D]. 2015