# Credit Decision of Small and Medium Sized Enterprises

**Haisen Wang[1], Yixuan Ma[2], Tiewen Chen[1]**

[1]School of Business, Central South University, Changsha, Hunan, 410000

[2]School of Economics, Beijing International Studies University, Beijing, 100024

**Keywords:** Logistic regression model, multivariate programming model, random forest algorithm

**Abstract:** This paper studies the credit strategy of banks to different enterprises, establishes a scoring system to quantitatively analyze the credit risk of enterprises, and establishes a credit strategy decision-making model for enterprises with different credit grades. Then, taking the data as training set, the credit rating and default are classified by random forest algorithm, and the planning model is established to calculate the specific loans. Finally, considering the impact of unexpected factors on the credit strategy in the credit decision-making model, the credit strategy is adjusted by industry. Firstly, we select a series of indicators that affect the credit risk of enterprises, and select the indicators that are completely multicollinearity to establish the disordered multivariate logistic regression model. Considering the probability that an enterprise belongs to different levels, the credit risk of the enterprise is quantified. The descriptive analysis of the purchase and sales invoice is convenient for the subsequent selection of indicators and the formulation of credit strategy. Taking Xinguan epidemic as a typical emergency factor, this paper analyzes its economic impact on different industries and industries.

## 1. Introduction

"Difficult financing" and "expensive financing" have always been the prominent problems in the development of small and medium-sized enterprises. It is difficult for an enterprise to meet all the capital needs only by internal accumulation. When the internal fund can not meet the demand, it must choose external financing. According to the empirical analysis results of different scale enterprises, Zhang Jie and Wang Xiao revealed that the external restriction of SMEs' financing structure mainly comes from the scale discrimination of banks, thus providing empirical evidence for the particularity of financing structure.

Generally speaking, banks will assess the credit risk of small and medium-sized enterprises according to their strength and reputation, and then determine whether to make loans and credit strategies such as loan amount, interest rate and term according to credit risk and other factors. However, limited by the weak strength of small and medium-sized enterprises, high credit risk, loan interest rate, customer churn rate and other factors, banks need to comprehensively consider the above factors when granting loans, establish clear and appropriate credit strategies, reduce bad debt risk as far as possible under the premise of protecting their own interests, avoid excessive financial risk and operational risk, and ensure the long-term and healthy development of banks Exhibition.

## 2. Credit Strategy

Based on the analysis of problems and data, and considering the influence of factors on corporate reputation, a multi classification logistic model is established. The dependent variables are four ordered multi classification variables. However, the parallel test of the dependent variables of ordinal classification failed, so the disordered multivariate logistic model was established.

For the disordered multi classification logistic regression, we first define the reputation level as a reference level, and compare the other three credit levels with it, and establish three generalized logistic models.

$$f_j(x) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon, (j = 1,2,3)$$

Three models were fitted to four independent variables

$$\ln\frac{P_1}{P_4}=\alpha_1+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4$$

$$\ln\frac{P_2}{P_4}=\alpha_2+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4$$

$$\ln\frac{P_3}{P_4}=\alpha_3+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4$$

At this time, the cumulative probability sum of the ordered values of dependent variables is 1:1

$$P_1+P_2+P_3+P_4=1$$

Then they are as follows:

$$P_1=\frac{e^{f_1(x)}}{e^{f_1(x)}+e^{f_2(x)}+e^{f_3(x)}+1}$$

$$P_2=\frac{e^{f_2(x)}}{e^{f_1(x)}+e^{f_2(x)}+e^{f_3(x)}+1}$$

$$P_3=\frac{e^{f_3(x)}}{e^{f_1(x)}+e^{f_2(x)}+e^{f_3(x)}+1}$$

$$P_4=\frac{1}{e^{f_1(x)}+e^{f_2(x)}+e^{f_3(x)}+1}$$

Using MATLAB to fit the established model, the following fitting information is obtained.

Table 1 model fitting information

| Model | Model fitting conditions | Likelihood ratio test | | |
|---|---|---|---|---|
| | -2 Log likelihood | Chi square | freedom | Significance |
| Intercept only | 337.027 | | | |
| final | 199.184 | 137.842 | 12 | .000 |

From the above fitting information, it can be seen that the model has statistical significance if the partial regression coefficients of all independent variables in the model are all 0 (significance < 0.1).

Table 2 goodness of fit

| | Chi square | freedom | Significance |
|---|---|---|---|
| Pearson | 227.617 | 354 | 1.000 |
| deviation | 199.184 | 354 | 1.000 |

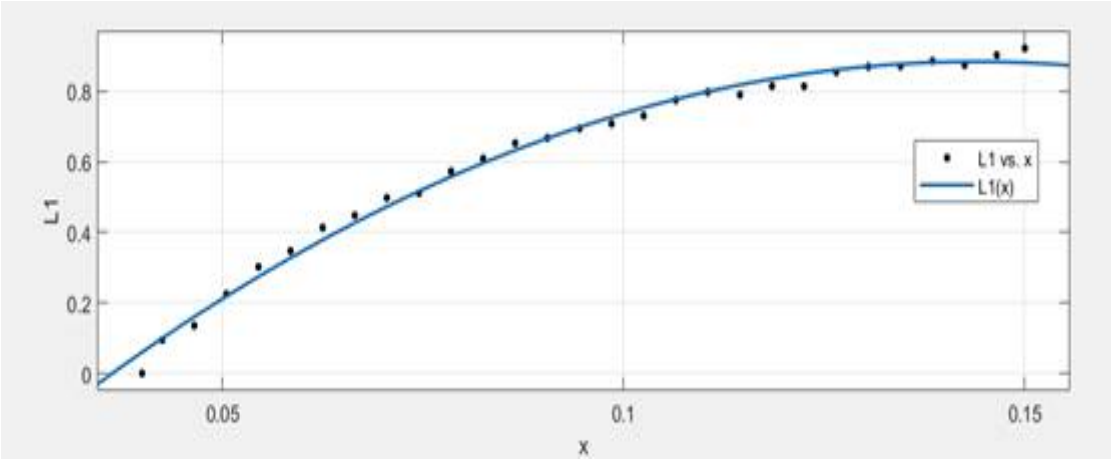


Figure 1 Fitting curve of relationship function between loan interest rate and customer churn rate of enterprise a with credit rating

$$L_1(x) = -76.41 * x^2 + 21.98 * x - 0.70$$

After verification, the fitting degree is high and the fitting effect is good.

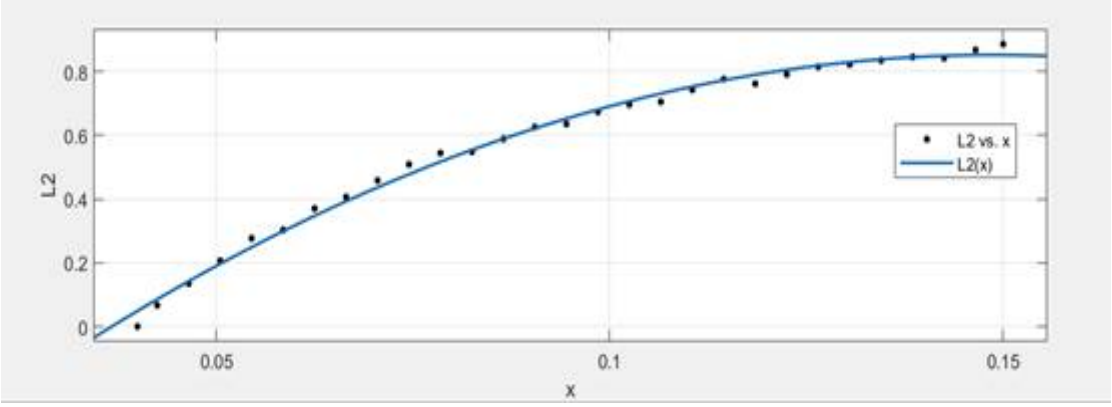


Figure 2 Fitting curve of relationship function between loan interest rate and customer churn rate of enterprise B with credit rating

$$L_2(x) = -67.93 * x^2 + 20.21 * x - 0.65$$

After verification, the fitting degree is high and the fitting effect is good.

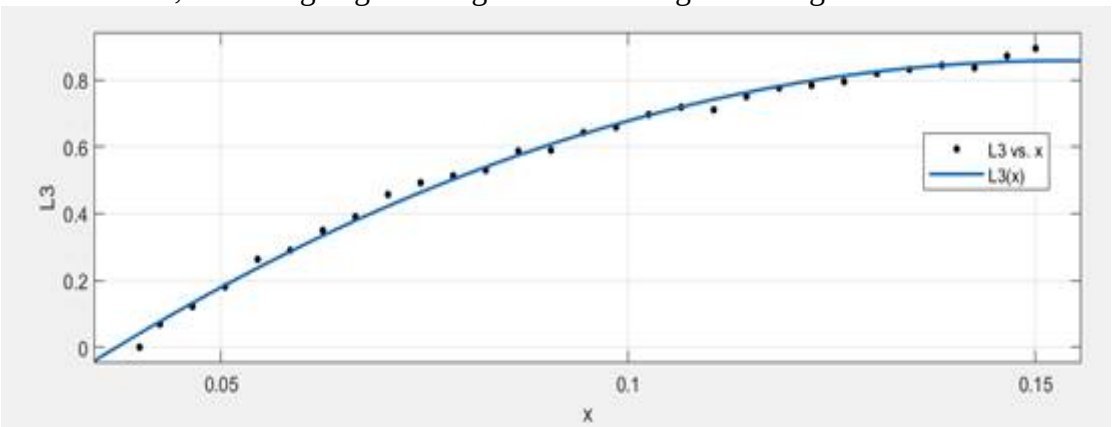


Figure 3 fitting curve of relationship function between loan interest rate and customer churn rate of enterprise with credit rating of C

$$L_3(x) = -63.94 * x^2 + 19.57 * x - 0.64$$

After verification, the fitting degree is high and the fitting effect is good.

The linear programming model is used to solve the credit strategy. The following functional

relationship between annual total credit and credit rating of banks is listed as follows:

a) Decision variables:

$$k_a, k_b, k_c$$

b) Objective function:

max

$$k_a * \sum_{i=1}^{27} (1 - L_1(x_1)) * (1 - p_i) * (1 + x_1) +$$

$$k_b * \sum_{j=1}^{38} (1 - L_2(x_2)) * (1 - p_j) * (1 + x_2) +$$

$$k_c * \sum_{n=1}^{34} (1 - L_3(x_3)) * (1 - p_n) * (1 + x_3)$$

c) Constraints:

$$k_a + k_b + k_c = 1$$

$$10 \le \frac{W * k_a}{N_a} \le 100$$

$$10 \le \frac{\mathrm{W} * \mathrm{k_b}}{N_b} \le 100$$

$$10 \le \frac{W * \mathrm{k_c}}{N_c} \le 100$$

$$0 \le k_a, k_b, k_c \le 1$$

d) The results are as follows

$$k_a = 0.28, k_b = 0.38, k_c = 0.34$$

Therefore, we conclude that the proportion of the loan line allocated by the bank to the enterprises with credit rating of a, B and C is 0.28, 0.38 and 0.34 respectively.

## 3. Impact of Unexpected Factors

(1) Analysis of the impact of unexpected factors on the three major industries

Sudden factors often have different impacts on enterprises belonging to different industries and different industries. This paper summarizes the general situation of the gross output value and output value of the three major industries during the epidemic period and the general situation of different industries affected by the epidemic situation from the statistical data published by the relevant national departments.

(2) Establishment of multivariate nonlinear programming model

Enterprises are divided according to the standard of industrial division, and the number of enterprises belonging to the primary industry is 0, and the epidemic situation has a great impact on the secondary and tertiary industries. Therefore, the degree of impact of the new crown on the secondary and tertiary industries should be considered. When emergencies occur, banks often reduce the lending interest rate or increase the amount of loans to enterprises to slow down the impact of emergencies on enterprises and make enterprises develop normally.

Based on the above analysis, a multivariate nonlinear programming model can be established

max

$$k_a * \sum_{i=1}^{72} (1 - L_1(x_1 + a_s)) * (1 - p_i) * (1 + x_1 + a_s) +$$

$$k_b * \sum_{j=1}^{95} (1 - L_2(x_2 + a_s)) * (1 - p_j) * (1 + x_2 + a_s) +$$

$$k_c * \sum_{n=1}^{74} (1 - L_3(x_3 + a_s)) * (1 - p_n + a_s) * (1 + x_3 + a_s)$$

$$W = 10000(\text{万元})$$

$$10 \le \frac{W * k_i}{N_i} \le 100, (i = a, b, c)$$

$$0 \le k_a, k_b, k_c \le 1$$

$$0.04 \le x_i + a_s \le 0.15 (i = 1,2,3; s = 2,3)$$

$$k_a + k_b + k_c = 1$$

(3) The solution of the model

Through the constraints of the model, matlab programming and SQP sequential quadratic programming method are used to explore the optimal solution

$$k_a = 0.791, k_b = 0.105, k_c = 0.104$$

$$a_2 = -0.0215, a_3 = -0.0215$$

Through the analysis of the solution results, it can be seen that when the emergency occurs, the bank can reduce the lending interest rate of 2.15% for the enterprises belonging to the secondary and tertiary industries to ensure the maximum bank income, and make the lending enterprises smoothly pass through the low development period caused by the sudden situation.

## 4. Analysis and Promotion

Logistic regression model can overcome the shortcomings of traditional mantel Haenszel analysis method and linear regression analysis. It can analyze multiple factors of discrete variables and continuous variables at one time, and can effectively analyze the confounding and interaction of external variables. It provides a quantitative description of the relationship between multiple risk factors and the probability of memorial hall.

This paper mainly uses the stochastic forest model to predict the reputation level and default of enterprises. Random forest is a classifier built by random way and contains multiple decision trees. The output category is determined by the mode of the categories output by each tree. Compared with other algorithms, random forest has many advantages, which is worth popularizing.

However, if the random forest model is popularized and used, there may be some problems. On the one hand, the random forest model has been proved to be over fitted in some noisy classification or regression problems. On the other hand, for the data with different values, the attributes with more values will have greater impact on the random forest, so the random forest is produced on this data. Attribute weight is not credible, so we need to pay attention to the related problems in the promotion and use.

## References

[1] Lin Yifu, Li Yongjun. Development of small and medium-sized financial institutions and financing of small and medium-sized enterprises [J]. Economic research, 2001 (01): 10-18 + 53-93

[2] Zhang Jie, Wang Xiao. Financial growth cycle and financing structure change of small and medium-sized enterprises [J]. World economy, 2002 (09): 63-70

[3] Liu Zhenhua. Research on credit management of commercial banks based on customer credit rating [D]. Hunan University, 2015