

# Research on Bond Investment Decision-Making Based on Logistic Regression Model

Dengsheng Liu, Chunyi Huang, Yilin Guo, Ying Li

*College of Tourism Date, Guilin Tourism University, Guilin, GuangXi 541006, China*

**Abstract:** At present, the bank's investment decision in a company directly determines its profitability. Appropriate usage of a multivariate logistical regression model can estimate the probability of default, which can enable the bank to make the best option indirectly. Firstly, the credit risk of many listed companies in different fields is used as a quantitative criterion and characteristic coefficients. Besides, it used to make a standard credit rating. Using this as a rule, it can count the credit ratings of different companies in emergencies. Then it can be used to help bank's investment decision. It is found that the results are reasonable and accurate.

**Keywords:** Logistical Regression Model; Probability of Default; Credit Evaluation; SPSS

## 1. Introduction

In the bank lending decisions, the unstable credit guarantees of micro enterprises and high financing risks are the main reasons why banks and other commercial institutions are unwilling to lend. Micro enterprises occupy a very important position in the national economy, which is a strong growth point for my country's economic growth. However, the current financial support of micro enterprises is far from meeting their development needs.[1]. In the conventional expert evaluation method, analytic hierarchy process and stepwise discrimination method need a large number of expert evaluations or important detailed index data[2]. The high cost of obtaining information and ensuring the validity of the information are the main difficulty that hinders the common credit evaluation of small and medium-sized enterprises. It results that banks are unable to credit micro enterprises well. At present, most commercial banks have realized that the business of micro enterprises has become a new profit growth point. But most of their existing

corporate credit rating models are applicable to large enterprises. Research has shown that there is significant difference in credit performance between large enterprises and small enterprises. It is important for the establishment of credit evaluation criteria for micro enterprises.

Secondly, in today's primary form of credit bonds, the bond credit ratings of many companies are also mere formalities, which hasn't much practical reference significance for banks. Often the error in credit ratings is a factor that the previous level did not consider[3]. Therefore, if we want to improve the level of financial services and enhance the level of risk prevention[4], we need to find a convenient, effective and economical method of obtaining information and establish a supporting credit evaluation system. It will help banks make loan decisions for SMEs[5]. The data mining mentioned by Lu Chen[6]is now a popular analysis method for commercial bank bond investment. It uses historical data to analyze and simulate future numbers. Zhou Hong and Lin Wanfa[7] built a related model, which found that the information asymmetry of bond issuers and investors is positively correlated with their corporate credit spreads. Wang Yong[8]used the KMV model to study that investment bonds also need to be based on local fiscal policies and conditions, the financial capabilities of enterprises, and local support for enterprises. This also confirms that market risk can affect the credit rating of a company. Experience has shown that the larger the scale of the enterprise, the more applicable the judgment standards applied in the industry will be suitable[9]. Because the retail loans of micro enterprises have been increasing year by year, the profits of commercial banks have steadily increased. The vigorous development of retail loans for micro enterprises has significant influence on the future. However, with retail loans increasing, the number and probability of corporate defaults have also increased. Compared with large enterprises, micro enterprises are more

susceptible to random economic fluctuations. Therefore, their relative default probability is also higher[10].

Aiming at the problem of commercial banks' profit, solving the credit guarantee problem of micro enterprises, the research methods and models have changed from linear regression to nonlinear regression. The Logistic model was firstly used by Martin to predict the company's bankruptcy and default probabilities. The data shows that whether foreign or domestic, it must judge the default risk probability of many commercial banks. Multi-objective decision-making solves the hindering influence of multiple decision factors on the Logistic model. Besides, the influence of covariance between each decision factor can be solved. Logisitic models have higher prediction accuracy than many neural network models, which makes up for the defects of least squares. All only needs to verify the correlation analysis of the factors that have an impact on the probability of default[11]. A variety of data has verified that Logisitic model does not need to meet the law of normal distribution. It can also achieves the same accuracy.[12] For the bank's evaluation of micro enterprises, only good and bad enterprises are classified. The result can be between 0 and 1. An accurate score is given between them so that banks can easily make bond investment decisions[13]. Therefore, the Logisitic regression model has been widely used and valued in economic decision-making and management.

## 2. The Solution and Testing of Logistical Model

### 2.1 The Solution of Logistical Model

#### 2.1.1 Research Ideas

Data collection and analysis were carried out, and it is for 302 listed companies that have been evaluated by experts. Logistical model and SPSS are used to calculate the correlation between the main factors of enterprises of different sizes and fields. Iterative algorithms were used to iterate, and it can dismiss the algebra. The regression value is obtained by the covariance test. The obtained results are compared with the actual analysis. The correctness and rationality of the model are tested. For companies of different nature, the expert evaluation method and the Logistical model are used for comprehensive comparison. A convincing model explanation is

obtained. A better decision-making plan is simulated on the bank's bond investment amount.

#### 2.1.2 Logistical Regression Model

In the bond field, Logisitical model has linear relationship between the natural logarithm of the ratio of corporate bond default, non-default and each characteristic factor. Compared with the Z-value model, the Logisitical model does not require too large historical data support. It does not need to fully conform to the normal distribution. Compared to the neural network model, the Logisitical model is easier to interpret. It does not require huge training samples. Compared with the KMV model, the Logisitical model does not require too large default samples to simulate the corporate default probability in the future, which is in line with the current domestic market. Therefore, the Logisitical model is more suitable for the domestic credit rating market in terms of sample and accuracy.

For SMEs, banks need to have a credit risk quantification standard in terms of strength and reputation as a lending strategy. In the economic field, Logisitical model uses the probability of default as the criterion for judging default. Its value is distributed from 0 to 1, and most companies cannot completely make the probability of default close to 1 or 0. So its value can be used to show the good or bad degree of the company with data. In addition, Logisitical model can also appraise the credit rating of enterprises considerably. It solves the difficulty of disorder of multi-categorical variables, which brings a lot of convenience in solving practical

applications. According to the model:  $p = \frac{1}{1+e^{-s}}$ , among then

$$s = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \dots\dots\dots(1)$$

$$p(y_i) = p_i^{y_i} (1 - p_i)^{1-y_i} \dots\dots\dots(2)$$

$x_k$  is characteristic factor variable when calculating the probability of default and evaluating the credit risk.  $\beta_n$  is the index coefficient before each feature factor. Through SPSS, it can use maximum likelihood regression calculation, and perform an iterative algorithm, which obtains various constraint results under restricted conditions. Therefore, the index coefficient of each feature factor can be obtained.  $P$  is the probability of default. The larger the value, the more it can indicate a type

of enterprise with low default probability and high credit credibility. When the number of feature factors exceeds 3 or more, the accuracy of the model will be very close to the true value. Calculating the corresponding characteristic factors for 302 listed companies, it can be found

that the three characteristic factors, the proportion of invalid invoices, the proportion of default invoices, and the proportion of negative invoices are the most prominent factors. The average value of multiple companies is fitted, therefore it can:

**Table 1. SPSS Preliminary Results**

Variable	B	Standard Error	Wald	Degree of Freedom	Significance	Exp(B)
Proportion of voided invoices	-1.65	2.591	0.408	1	0.523	0.191
Proportion of default invoices	11.549	28.902	0.160	1	0.689	10.367
Proportion of negative invoices	-3.105	3.299	0.885	1	0.347	0.045
Constant	1.520	0.327	21.541	1	0.001	4.572

In this way, the model formula of the bank’s Logistical model for hundreds of randomly selected companies is as follows:

$$P = \frac{1}{1 + e^{-1.520 + 1.656 x_1 - 11.549 x_2 + 3.105 x_3}}$$

When fitting and calculating the Logistical model in different economic fields, too large sample size will also cause large errors in the value. It results in unsatisfactory simulation results. At this point, further testing of the model is required.

For the test of regression coefficient:

$$w = \left( \frac{\bar{\beta}_j}{SE \bar{\beta}_j} \right)^2 \dots \dots \dots (3)$$

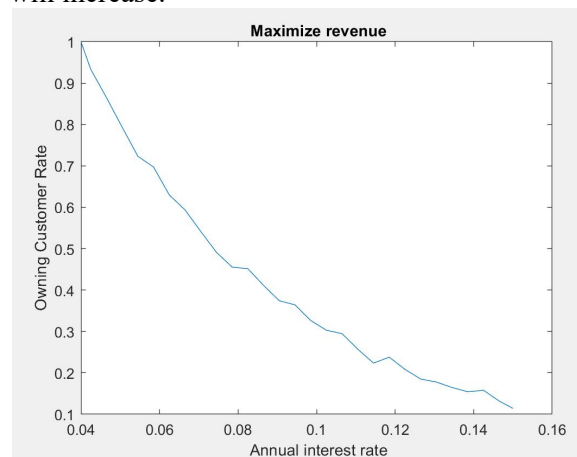
Among them,  $SE \bar{\beta}_j$  is the standard error. The obtained value of  $w$  is equal to or less than the critical value  $w_\alpha$ . That is, it can be considered that the independent variable has an influence on the occurrence of the event. The acceptable test error of a commercial bank is almost 0.05, so the bank can measure and evaluate the three characteristic factors.

Under different regression coefficients, the basic and critical values are the same. The three characteristic factors also have a greater impact on the enterprise’s default probability. Substituting it into equation (1), it gives a good effect and a high degree of simulation. At the same time, the enterprises are divided into three categories: large, medium, and small. SPSS is used to simulate and calculate the results of three sizes of enterprises. It can analyze the three dimensions.

**Table 2. Validation Regression Coefficient**

	$W$	$W_\alpha$
Proportion of obsolete	0.408	0.408
Proportion of Default	0.159	0.160
Proportion of negative invoices	0.885	0.885

When the bank’s annual limit is fixed every year, the repayment ability and credit evaluation ability of the enterprise are integrated to divide the amount that different enterprises can repay. When the amount of bonds invested in a company is limited, the churn rate of customers will increase.



**Figure 1. Owning Customer Rate**

With the ever-increasing annual interest rate of banks, the number of customers owned by corporate customers will gradually decline. The default cut-off value of SPSS is 0.5. Banks can roughly divide the companies into two types of companies with large credit gaps, which is by the way of below and above the cut-off value. They also have more accurate and reasonable grasp of bond investment, in order to avoid excessive investment and small return. Removing the larger error extremes on both sides of the default probability, it can be made into a commercial bank’s own credit rating setting range.

**2.2 Analysis of Bank Profit Strategy**

**2.2.1 Research Ideas**

The return value based on the logistical model is the probability of default. The default

probability is calculated to quantify the credit evaluation of each enterprise. Using SPSS can carry out credit evaluation standards for credit risk factors. Classify enterprises into four grades: good, medium and bad. From this, we can get the number of good and bad enterprises among the enterprises.

Calculate the rate of return and annual profit based on the amount of bank loans and the actual profit. The value calculated when LINGO is used for linear programming (assuming that all enterprises above the difference category can borrow from the bank). It shows that companies can borrow more to achieve higher returns.

### 2.2.2 Banks Developing Strategies

When the bank can already lend to the determined enterprise limit, and the annual interest rate is determined. The customer churn rate is calculated, and the return is the highest. Judge the credit risk of premium enterprises. The companies whose evaluation criteria fluctuate within 0.05 are removed. Those companies that are significantly higher than the evaluation criteria are selected to allow banks to have an accurate grasp of corporate lending.

Logistic regression is a model based on linear regression. For variables such as enterprise size and risk value, the characteristic variables need to be standardized when the logistic regression value is adjusted. For discrete variables, they are processed by grouping. For continuous variables, Min-Max is used for standardization.

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Among them,  $x^*$  represents the standardized value,  $x$  represents the original value,  $x_{\min}$  represents the minimum value of the variable,  $x_{\max}$  and represents the maximum value of the variable.

It can obtain in the credit rating of excellent enterprises. According to the proportion of negative invoices, it can be seen that the proportion of negative invoices of some companies is less than 1%. It can be seen that large companies with good credit still have insufficient working capital. It is also has set a fixed under the premise of the amount of money. The choice for such enterprises is not to give them loans. Count the remaining large companies in the superior category.

So set  $A_1$  as the maximum loan amount for each enterprise.

For the remaining small enterprises after the decision on credit risk of the premium category, if the probability of default does not exceed the evaluation criteria, they are discarded. We set the maximum loan amount for each company as  $A_2$ .

In the same way, the good and medium-level large enterprises use the annual interest rate-the highest value of the customer churn rate. In this way, it can select the good and medium enterprises. It use  $B_1$ 、 $C_1$  to represent the loanable amount.

The remaining small and medium-sized enterprises are also screened. The number of good-quality small and medium-sized enterprises is obtained. The number of medium and small-sized enterprises is set as  $B_2$ 、 $C_2$ .

It is required that the proportion of large enterprises in each category must be greater than that of small and medium-sized enterprises. In this way, it can ensure that the bank's customer default rate will not be too large. Therefore, LINGO is used to perform linear programming for some of the selected enterprises. The calculated values are as follows:

In response to this, according to these 6 different risks and different sizes of enterprises, we use SPSS to calculate the mean and standard deviation of the default probabilities. So that we can see that better comparisons and verifications can be found among enterprises of different sizes.

**Table 3. Type Distribution and Total**

$A_1$	0.1845
$A_2$	0.1845
$B_1$	0.3309
$B_2$	0
$C_1$	0.2999
$C_2$	0

It can be seen that among the credit risk companies, the average value is only slightly higher than the evaluation standard. Therefore, banks need to be cautious in borrowing among companies. It increases the number of investments and reduces the amount of investment. The profit cycle is as long as possible.

It can be seen that in the evaluation of premium credit, regardless of whether it is a large or small enterprise, banks try to give them loans as much as possible. They can fully drive economic development. In the good and medium credit evaluations, since the averages are not much different, we need to consider giving large companies more loans. This confirms that the calculated ratios are reasonable and beneficial to the bank's decision-making.

Regardless of whether it is excellent, good, or medium, the total proportion of large enterprises is basically more than 0.5. Excellent is 0.576, good is 0.512, and medium is 0.351. Therefore, the bank's investment decision depends not only on default. The probability also depends on whether the company is a large or small company.

### 3. Conclusion

In view of the bank's bond issuance strategy for micro enterprises and the data results, the bond quota issued by commercial banks to micro enterprises depends largely on the creditworthiness of the enterprises. Calculate the probability of default using the company's percentage of invalid invoices, default invoices, and negative invoices. Based on historical data and Logistical model, SPSS software is used to calculate the default probabilities of more than 600 listed companies. After verifying them, it is concluded that the model fits well. It can be known that the accuracy rate of companies that are not affected by major fluctuations in the Logistical model research, which is higher. The main goal of this paper is to calculate the value of default probabilities for various types of enterprises. Besides, it formulates most types of micro enterprises issuing bonds, which can be expanded to larger bond areas in the future.

### Acknowledgement

The research is supported by: 2019 project of foundational research ability enhancement for Young and Middle-aged University Faculties of Guangxi (2019KY1046), 2019 project of foundational research ability enhancement for Young and Middle-aged University Faculties of Guangxi (2019KY1037), Nature and Science Foundation of Anhui(2008085QA08), Scientific Research Projects of Institute of Information Technology of GUET.(B201911), Science and technology research project of Heilongjiang Provincial Department of Education12543079,

2020 Project of Three-wide education of Institute of Information Technology of GUET(2020SQ03).

### References

- [1] Li Riyang. Research on the Credit Evaluation of Small, Medium and Micro Enterprises Based on Logistics[J]. Market Weekly, 2019(11): 133-13
- [2] Zhang Wen. Credit evaluation of small and medium-sized enterprises based on logistic regression [D]. Shandong University, 2020.
- [3] Wang Qiulong. Research on the credit risk of my country's credit bond market based on the logit model[D].Southwestern University of Finance and Economics, 2018.
- [4] Guan Duo. Research on GD Bank's SME Credit Evaluation System Optimization[D].Yanshan University, 2018.
- [5] Chu Zehao. Research on the Credit Evaluation System of Small and Medium-sized Enterprises Based on Tax Information [D].Henan University of Economics and Law, 2020.
- [6] Lu Chen. Analysis of bank's corporate loan default risk prediction[D].University of Electronic Science and Technology of China, 2012.
- [7] Cheng Mengting, Yang Huawei. The financial status of listed companies predicts bond credit risk-based on logistic model [J]. Industry and Technology Forum, 2019, 18(16): 97-99.
- [8] Zeng Qimin. Research on the credit risk of water-related urban investment bonds based on the multivariate logistic model[D]. Chongqing University, 2019.
- [9] Lei Youxi. Research on the Credit Evaluation of Small, Medium and Micro Enterprises Based on Logistics Regression[D]. Xiangtan University, 2016.
- [10] Gao Fei. Analysis of commercial bank retail loan customer default based on Logistic regression model—Taking MSLZ Bank as an example [J]. Regional Governance, 2019(38): 204-207
- [11] Pang Sulin. Application of Logistic Regression Model in Credit Risk Analysis[J]. Mathematics in Practice and Knowledge, 2006(09):129-13
- [12] Cai Wenxue, Luo Yonghao, Zhang Guanxiang, Zhong Huiling. Personal credit risk assessment model and empirical analysis based on the integration of GBDT

- and Logistic regression [J]. Management Modernization, 2017, 37(02): 1-4.
- [13] Shi Linli, Li Chenyu, Ji Chen, Liu Dingnan, Qian Runkang, Jin Jingliang. Credit evaluation of small and medium-sized enterprises based on Logistic regression model[J]. Journal of Science of Normal University, 2018, 38(05): 10-17.
- [14] Huang Jianwu. Research on the Influencing Factors of Manufacturing Corporate Bond Default Based on the Logit Model [D]. Lanzhou University, 2020.
- [15] Guo Qi, Zhang Yu. Development of my country's bond market credit rating industry and suggestions[J]. Bonds, 2012(04): 14-19.