The Application of the Logistic Regression Model in Credit Risk Assessment of Listed Companies

Kexuan Xu

School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, 215123 Suzhou, China
Kexuan.Xu24@student.xjtlu.edu.cn

Abstract. Nowadays, since the capital market is expanding increasingly, more and more listed companies are coming to show their promise. However, the number of their credit risk exposures has surged; this phenomenon not only influences the reputation and financial performance of these listed companies but also affects the stability and credibility of the whole trading market. This article used a logistic regression model as a tool to conduct research about listed companies' credit risk assessment. This model is suitable for binary classification evaluation of credit risk. The study results show that the ratios of asset liability, return on equity, and total asset turnover are the most influential factors for the companies' credit risk. Among these indicators, the asset liability ratio is directly proportional to risk, while the ratios of return on equity and total asset turnover are inversely proportional to risk. The logistic regression model built in this article can offer more proper and trustworthy suggestions for loan approval of a financial institution and investment in an enterprise or personally. By using this model, it can discover and avoid problems easily, which makes up for the objective and subjective disadvantages of traditional risk assessment.

Keywords: Credit Risk, Logistic Regression Model, Financial Indicator.

1 Introduction

As an important part of the capital market, the operation status and credit rating play a significant role in the stability of the whole capital trading market. In recent years, the fluctuations of the financial market have increased, and enterprise transformation and upgrading, as well as optimization of company results, are urgent. When the listed companies are developing and expanding continuously to seize the market share, a variety of credit risk problems have increased, such as poor management or default [1]. Implementing the credit risk assessment is important for financial institutions' credit decisions and investor risk aversion. It is better to estimate the potential default and investment losses in advance, which could promote efficient decision-making by all parties, profit maximization, and transparency and standardization of the trading market. As a fast and stable linear classification model, the logistic regression model has a strong explanatory power. Its output results are probabilities that are suitable for financial risk control and credit scoring. From these results, the direction and

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_129

magnitude of the impact of each variable can be observed, thereby effectively analyzing the influence of various indicators on the credit risk of listed companies.

To evaluate the credit risk of each listed company more correctly, the existing documents conducted a series of studies centered on a credit risk assessment model. Machine learning can obtain a good application in the field of credit risk assessment, whose ability to strengthen risk management is a key advantage [2]. Machine learning algorithms, neural networks, and natural language processing provide some innovative approaches to credit assessment, which are capable of selecting large data sets, identifying complex models, and adapting to the changing market dynamically [3]. Although financial institutions will classify consumers according to creditworthiness, there has not been any set of attributes or metrics that can be widely accepted up to now [4]. Smaller financial institutions can use a collaborative credit risk assessment model to raise expectations significantly [5]. Combining the quantum deep learning technology and an adaptive model based on industry type forecast and analysis can effectively improve the risk assessment level [6].

The existing literature analyzed the evaluation method of enterprises from various types of machine learning and deep learning. However, there's still a research gap based on financial information in the context of China. This article treats listed companies as research objects, uses a logistic regression model, analyzes data from the balance sheet, Income statement, and Cash flow statement to get the result that the ratios of asset liability, return on equity, and total asset turnover are the most influential factors for the companies' credit risk. In theoretical terms, different from traditional forecasting, this research applied the logistic regression model in the risk assessment of listed companies, which can produce results more quickly and accurately. In practical terms, this method is easy to operate and popularize, which can help crowds of investors avoid high-risk companies.

2 Related Theories and Model Foundation

Credit risk, which is the main risk for banks, means that listed companies, as debtors, cannot repay the principal and interest because of personal business, financial, and regulatory issues, and this may be attributed to the economic losses of creditors, investors, and shareholders [7]. The core reason for enterprises' credit problems is information asymmetry, decline, or even loss of solvency, willingness, and responsibility is its essence. There are five main dimensions to measure a company's credit metrics: profitability, solvency, operational capability, growth ability, and innovation ability. This article carries out an assessment from two directions: the quantitative analysis with various indicators as the main body, and the qualitative analysis based on corporate brand image as well as regulation structure. Then, choose the three most typical dimensions from the five: profitability, solvency, and operational capability. Based on the database collected during the research on the regression model, this paper uses the asset-liability ratio (x1) and current ratio (x2) to explain solvency, the total asset turnover rate (x3) to represent operational capability, and return on equity (x4), operating profit margin (x5), and net interest rate (x6) to

interpret profitability. Furthermore, the credit risk status of listed companies is taken as the explained variable and measured by dichotomy: listed companies marked as ST/*ST or in debt default are defined as default samples and assigned a value of 1, while normally operating companies without ST labels are defined as non-default samples and assigned a value of 0. Credit risk assessment is a fundamental process for finance to maintain operational stability and earnings sustainability. It is significant for identifying existing risks in the business activities and making wise and profitable decisions [8].

There are many kinds of credit risk formative factors for listed companies, this article mainly divided into financial indexes and industry indicators two aspects to study and discuss. Financial indexes, as the primary elements, are the subjective factors for companies' credit issues, and these kinds of indexes are used to assess the financial performance and credit standing of enterprises [9]. The solvency, profitability, and operational capability will directly impact companies' financial indexes. The study uses the asset liability ratio to judge the long-term debt capability of enterprises; the higher the numerical value, the higher the credit risk. For the current ratio, if the value is relatively low, then the debt-paying ability is insufficient in the short term, and short-term debt default is more likely to occur. The total asset turnover ratio, it reflects the operational efficiency of companies; the higher the value, the stronger the operational capability. The high value of return on equity indicates strong profitability of shareholder equity and also good profitability. Operating profit margin and net income represent the profitability of a company; the higher the value, the stronger the stability. Moreover, some industry indicators, which act as objective factors, are also important reasons that influence the credit risk of companies, such as the strong periodicity of the trading market, competitive landscape due to industry density, and even some sudden and impactful industry events.

Credit risk mainly comes from information asymmetry, whose problems between the buyers and the sellers will lead to adverse selection or moral hazard indirectly, resulting in a reduction in the credibility and profit of all parties. Taxonomy is usually used to evaluate credit, that is, dividing the listed companies into default and non-default samples, then measuring the feasibility of cooperating with these companies. This method makes the information more transparent and reduces the information asymmetry, which is convenient for listed companies to monitor their own operational issues and financial institutions to evaluate cooperation income. Logistic regression model is the classic statistical modeling method for classification issues, mainly used to analyze the impact of independent variables on binary dependent variables and forecast the probability of the occurrence of the dependent variable when sorting out one of the independent variables. When analyzing credit issues of listed companies, the study defines dependent variables as default (1) or nondefault (0), and defines various types of financial indicators as independent variables. Then the method calculates the default probability and analyzes the core factors of credit problems by using a regression model to fit the relationship between explanatory variables and non-explanatory variables. If the default probability exceeds the critical value of 0.5, the company will be judged as a high probability of default enterprise, and this will play a warning and reminder role in the following financial trading.

The estimation of linear regression is continuous and has no fixed range. There will appear some problems if it is being used in the listed companies' credit risk analysis, for example, the estimation exceeds the range of [0,1] or the homogeneity of variance does not hold. Furthermore, accurate conclusions cannot be drawn [10]. To avoid these kinds of problems, the study introduces a logistic regression model and uses a binary dependent variable to analyze whether various independent variables have an impact on default probability. Logistic regression model maps the linear regression results within the interval [0,1] through the sigmoid function. Suppose Y represents the credit status of listed companies, among which, Y=1 represents default and Y=0 represents nondefault. Assume X_1, X_2, ..., X_6 as the credit risk explanatory variables, P=P (Y=1|X_1, X_2, ..., X_6) as the default probability under given risk factors. When the probability of Y=1 exceeds the critical value 0.5, the company will be judged as a high probability of default enterprise, and needs to consider the investment issues again.

3 Construction and Application of the Logistic Regression Model

The data used for analysis is from the Balance Sheet, Income Statement, and Cash Flow Statement in the CSMAR database. Before putting into study, the study does some preprocessing, including associating each enterprise's three financial statements by securities code and deadline, and organizing the Balance sheet, Income statement, and Cash flow statement from the same company and the same reporting period. Then label the default samples according to the ST/*ST identification in securities abbreviation. If one company is a default sample or has ST/*ST identification, then recorded as 1; if it is nondefault and has no ST/*ST sign, recorded as 0. After cleaning the data, the study extracted the financial indicators as explanatory variables.

In model settings, in order to construct the Logistic regression model, the study uses the binary results of default and non-default as explanatory variables, and uses the core financial indicators as explanatory variables. The explained variables stand for the credit risk status of listed companies, which belong to binary discrete variables. Meanwhile, the explanatory variables are characteristic parameters, as the input of the model, they fall into three categories: solvency includes asset liability ratio (X_1) and current ratio (X_2), operational capability includes total asset turnover (X_3), profitability includes return on equity (X_4), Operating profit margin (X_5) and net interest rate (X_6).

The Logistic regression maps the linear regression results in the [0,1] interval through the sigmoid function; its main parameters are the regression coefficient $\beta = (\beta_0, \beta_1, \beta_2 \ldots \beta_6)$ and the critical value of default (0.5). Firstly, the linear combination is as model (1):

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_6 X_6 \quad (1)$$

Among these, β_0 is the model intercept term, reflecting the basic default probability if there is no explanatory variable effect. The regression coefficient of each

explanatory variable reflects the impact and direction of the corresponding financial indicators on credit risk. A positive coefficient means that this indicator will increase the default probability; a negative coefficient means a decrease in default probability. Secondly, the sigmoid function is used to calculate default probability. When $P \geq 0.5$, the company will be convicted of default; similarly, $P < 0.5$ will be convicted as nondefault. Thirdly, about the model fitting parameters, the larger the log likelihood, the closer the goodness of fit to 1, the better the fitting effect.

4 Effect Analysis of Model Assessment

According to regression results, when the observed value is equal to 1, residual value=1-predicted probability and the interval is between (0,1); when the observed value is equal to 0, residual value=0-predicted probability and the interval is between (-1,0). The residual values show a linear downward trend from positive to negative, spread regularly above and below zero. In the range of $P<0.2$, they have a low probability for actual defaulting companies to be forecasted as default, illustrating that this model underestimates the risk of default of these enterprises. In the range of 0.2 to 0.8, influenced by asset liability ratio ($\beta_1=0.788$) and net interest rate ($\beta_6=0.203$), the model overestimates the probability of default of some enterprises. In the range of $P>0.8$, the model overestimates nondefault companies and regards them as credit risk. In conclusion, the residual values distribute around zero decently, showing there exists systemic error but still within control.

From the results of model analysis, the predicted probability of the nondefault group is centered around zero, showing that the model can identify low-risk enterprises correctly for financial institutions to cooperate with; the predicted probability of the default group is centered around 1, proving that the model can confirm and catch high-risk enterprises for timely evasion. Two groups of data are distributed along the reference line Y=X, which represents the effectiveness of the logistic model. Then this article came out, the core fit index according to the observed value and predicted value: RMSE=0.387; MAE=0.274. The smaller the two values, mean a smaller the average error. They are within control, have no extreme bias, and are in a good fit. Another index, R^2, is 0.286, which shows that the model can explain 28.64% default status, and the evaluation ability tends towards a moderate level.

Table 1 presents the coefficient of β, reflecting the impact of each index on companies. In the aspect of error, the residual shows a linear decreasing trend and is distributed above and below zero regularly, which reveals the instability of the model in different forecasting intervals. However, for the aspect of residual, it is distributed evenly and has no obvious systemic bias. In terms of goodness of fit, the observed value and predicted value mainly distribute around the reference line Y=X, showing that the model can catch the overall trend of the data. Moreover, for indicators, the overall accuracy is 71.68%, showing the overall prediction effect is good; the accuracy of nondefault recognition is relatively high, with a figure of 83%, and the default recognition can cover most defaulting enterprises with a probability of 66%. In summary, the model obtains good fitting results.

Table 1. Model coefficient.

Parameter symbol	Corresponding variable	Parameter value	Economic significance
β_0	Intercept term	-0.108	The basic default trend of the model
β_1	Asset liability ratio	0.788	Positive correlation: the higher the ratio, the greater the default risk (significant impact)
β_2	Current ratio	0.018	Weak positive correlation: less impact on risk
β_3	Total asset turnover	-0.322	Negative correlation: the higher the ratio, the lower the risk
β_4	Return on equity	-3.865	Strong negative correlation: the stronger the ability, the smaller the default risk (the greatest impact)
β_5	Operating profit margin	-0.226	Negative correlation: the better the profitability of the main business, the lower the risk
β_6	Net interest rate	0.203	Positive correlation: should be combined with business to conduct further analysis

Based on the analysis above, this article proposes the following opinions: Firstly, regarding the listed companies, they should operate and manage companies properly; supervise and control the cash flow strictly; formulate strategies focused on urgent economic dilemmas to strengthen solvency; and ensure the rightful rights and interests are not infringed upon, maintain the bond between shareholders and the company. Secondly, regarding the financial institutions such as banks, they have better to use different regression model to forecast the credit risk of the companies they want to invest in, making the loan risk lower; and develop more reciprocal financial products and lending policies, help small enterprises grow quickly, and maintain the listed companies' capital. Thirdly, regarding the regulation departments, it is suggested that they use a quantitative model to monitor default or near-default companies, prevent default issues promptly, and protect the stability of the capital market trading platform.

5 Conclusion

This article takes Chinese listed companies as the subject, chooses data from Balance sheet, Income statement, and Cash flow statement in the CSMAR database from 2022 to 2025 as study samples, and applies the Logistic regression model to analyze the credit risk of listed companies. Through a series of steps, including data cleaning and sorting, model constructing, and error and fitness analysis, to test the feasibility of this model and draw the research conclusion, as follows: Firstly, solvency, operational

capability, and profitability are the significant factors that influence listed companies' credit. The study calculated six financial indicators by filtering through database headers and determined their impact on companies' default, including asset liability ratio, current ratio, total asset turnover, return on equity, operating profit margin, and net interest rate. Secondly, the ratio of asset liability, return on equity, and total asset turnover are the most influential factors from the regression model. Return on equity is the most critical indicator of default issues, which needs to be given special attention. The more profitable a company, the lower the default risk. The asset liability ratio is positive to the risk of default; highly indebted enterprises need stronger management. The total asset turnover reflects the operational efficiency; the higher the efficiency, the lower the risk, so it is suggested to improve the utilization efficiency. This shows that solvency and operational capability play a decisive role in listed companies' credit problems. Thirdly, the Logistic regression model has an accuracy probability of 71.56%. This model has a good fit and practicality, so it can be used as the preliminary screening tool for listed companies, helping with investment decisions and risk management. Using this model can distinguish between the default and non-default companies quickly and correctly, playing a warning role for both investors.

The main contributions of this article are as follows: Firstly, this article optimizes the credit risk assessment model, making the analysis more convenient and accurate. Secondly, this article proposes the core factors that influence the companies' credit risk, and provides some research thoughts for other researchers, for example, deducing the important capabilities for companies according to the referred financial indexes. Moreover, there are also some shortcomings of this study. The construction of the logistic regression model has not covered all its features, such as not considering the dynamism of factors. The code model will be introduced to interpret temporal changes and explore the impact of dynamic changes on credit risk in further research. In the future, the study may be combined with AI technology, utilizing machine learning algorithms to identify trends and factors that traditional models fail to detect. Meanwhile, the scope of research samples will be expanded by selecting databases containing more indicators to analyze how non-financial factors influence credit risk and contribute to the improvement of China's capital market.

References

1. Roy, J.K., Vasa, L.: Transforming credit risk assessment: a systematic review of AI and machine learning applications. Journal of infrastructure, policy and development 9(1), 9652 (2025).
2. Bello, O.A.: Machine learning algorithms for credit risk assessment: an economic and financial analysis. International journal of management 10(1), 109–133 (2023).
3. Edunjobi, T.E., Odejide, O.A.: Theoretical frameworks in AI for credit risk assessment: towards banking efficiency and accuracy. International journal of scientific research updates 7(01), 092–102 (2024).

4. Amarnadh, V., Moparthi, N.R.: Range control-based class imbalance and optimized granular elastic net regression feature selection for credit risk assessment. Knowledge and information systems 66(9), 5281–5310 (2024).
5. Chen, Z.S., Zhou, J., Zhu, C.Y., et al.: Prioritizing real estate enterprises based on credit risk assessment: an integrated multi-criteria group decision support framework. Financial innovation 9(1), 120 (2023).
6. Rath, M., Date, H.: Quantum powered credit risk assessment: a novel approach using hybrid quantum-classical deep neural network for row-type dependent predictive analysis. EPJ quantum technology 12(1), 75 (2025).
7. Xu, S., Zhang, M.: Research on credit risk assessment of listed companies in technology sector based on support vector machine integration. Procedia computer science 214, 867–874 (2022).
8. Chakraborty, G.: Evolving profiles of financial risk management in the era of digitization: the tomorrow that began in the past. Journal of public affairs 20(2), e2034 (2020).
9. Attigeri, G.V., Manohara, P.M.M., Pai, R.M.: Credit risk assessment using machine learning algorithms. Advanced science letters 23(4), 3649–3653 (2017).
10. LaValley, M.P.: Logistic regression. Circulation 117(18), 2395–2399 (2008).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

