

Empirical Study on Bond Default Early Warning Model Based on Logistic Modeling

Zheng Yu *

Department of management, Xiamen University, Xiamen, China

* Corresponding Author Email: 895741260@qq.com

Abstract. The objective of this paper is to develop a precise bond default early warning model for China's capital market by screening indicator variables and constructing a Logistic model based on existing literature on corporate default, bankruptcy, or financial distress. This paper uses financial data of all the bond-issuing enterprises that meet the research criteria during the period from 2014 to June 2019 as a training sample to construct the model. It examines financial indicators from one to three years before bond defaults occur and uses them to create an early warning model. Additionally, it uses financial data of all eligible bond-issuing enterprises from July 2019 to December 2019 as a test sample to evaluate the predictive accuracy of the developed model.

Keywords: Logistic models; bond default; early warning models.

1. Literature review

From the perspective of the influencing factors of default, the early studies were mainly based on corporate bankruptcy or debt default, and mainly based on financial and operational factors, such as Wu Shinong and Lu Xianyi (2001)[1]. In terms of the research methods of default, discriminant models[2] and structure models[3] have been applied. In recent years, methods based on machine learning have been used, such as artificial neural network (ANN)[4], support vector machine model (SVM)[5] and so on. At present, these early warning models based on machine learning methods have advantages in terms of accuracy (Wu Shinanong et al., 2021)[6], but it is difficult to clearly reflect the corresponding causes of bond defaults, and thus it is difficult to propose countermeasure suggestions.

2. Selection of data sources, study period, study sample and study indicators selection

The data in this paper is from Wind database. This paper takes the enterprises that issue credit debt in the bond market of our country as the object of empirical research.

The research time interval selected for training samples in this paper is from 2014 to June 30, 2019. In this paper, the year in which the first bond default of the enterprise occurs is noted as year t , 1 year before the default is recorded as $t-1$ year, 2 years before the default is recorded as $t-2$ year, and 3 years before the default is recorded as $t-3$ year, and then the eligible financial indicators in year $t-1$, year $t-2$, and year $t-3$ are introduced into the model as the independent variables. Three early-warning models of $t-1$ year, $t-2$ year and $t-3$ year bond default were established respectively, and the accuracy and effectiveness of the models were evaluated. After that, a unified model is sorted out. Finally, the data from July 2019 to December 2019 are utilized as test samples to test the prediction accuracy of the model.

According to the past research literature, this paper selects indicators reflecting profitability such as ROA, ROE, etc., reflecting cash generating ability such as cash assets/total assets, etc., reflecting debt ratio such as asset-liability ratio, etc., reflecting debt servicing capacity such as EBIT principal and interest cover times, etc., reflecting liquidity of assets such as quick ratio, current ratio, etc., reflecting asset turnover speed such as the total asset turnover rate, etc., reflecting the ability to grow such as capital accumulation ratio, reflecting the investment and financing situation, such as NCF

from investment activities / average total assets, NCF from financing activities/ average total assets, etc., totaling 48 indicators.

3. Empirical study of bond default early warning models

3.1. t-1 year Logistic model construction and test

3.1.1 Model construction

After doing significance test for 48 indicators in year t-1 as indicators to be selected, the Logistic model is constructed by forward stepwise regression method. Finally, 7 index variables including return on equity (ROE) (x_2), financial expense ratio (x_3), current asset ratio (x_{37}), capital accumulation ratio (x_{46}), NCF from financing activities/ average total assets (x_{47}), rigid liability ratio (x_{18}), and increase in cash and its equivalents/total assets (x_{14}) are obtained.

The Logistic regression equation for year t-1 is derived from the above indicators combined with the constant term:

$$\ln \frac{P}{1-P} = -7.526 - 0.034 x_2 + 3.356 x_3 + 1.125 x_{37} - 0.434 x_{46} - 7.079 x_{47} + 3.632 x_{18} - 7.144 x_{14} \quad (1)$$

3.1.2 Testing of the model

The first is the multicollinearity test. In this paper, tolerance (TOL) and variance inflation factor (VIF) are used to conduct multicollinearity tests on the above 7 indicator variables. Tolerance is relatively close to 1, while VIF is far less than 10, and the multicollinearity test is passed.

The second is the chi-square test. Chi-square test was performed on formula (1), and the chi-square value of the model was 199.442, and the significance level of the model was 0.000, indicating that the model was significant.

3.2. t-2 year Logistic model construction and test

3.2.1 Model construction

After significance test of 48 indicators in t-2 year, the Logistic model is constructed by forward stepwise regression method. Finally, 5 index variables including financial expense ratio (x_3), accounts payable turnover ratio (x_{45}), current asset ratio (x_{37}), NCF from financing activities/ average total assets (x_{47}) and rigid liability ratio (x_{18}) are obtained into the model.

From the above indicators combined with the constant term, a logistic regression equation is derived for year t-2:

$$\ln \frac{P}{1-P} = -7.945 + 3.154 x_3 + 0.006 x_{45} + 1.192 x_{37} - 6.109 x_{47} + 3.992 x_{18} \quad (2)$$

3.2.2 Testing of the model

The first is the multicollinearity test. The tolerance (TOL) and variance inflation factor (VIF) are used to conduct a multicollinearity test on the above 5 indicator variables, and the tolerance is relatively close to 1, while the VIF is far less than 10, and the multicollinearity test is passed..

The second is the chi-square test. Chi-square test is performed on formula (2), and the chi-square value of the model is 107.799, and the significance level of the model is 0.000, indicating that the model is significant.

3.3. t-3 year Logistic model construction and test

3.3.1 Model construction

After significance test of 48 indicators in t-3 years, the Logistic model is constructed by forward stepwise regression method. Finally, 5 index variables including principal and interest cover ratio 2

(x_{24}), accounts payable turnover ratio (x_{45}), current asset ratio (x_{37}), equity generation ratio (x_{12}) and interest-bearing liability ratio (x_{17}) are obtained into the model.

The Logistic regression equation for year t-3 is derived from the above indicators combined with the constant term:

$$\ln \frac{P}{1-P} = -7.804 + 0.005 x_{24} + 0.004 x_{45} + 1.3 x_{37} - 1.215 x_{12} + 4.668 x_{17} \quad (3)$$

3.3.2 Testing of the model

The first is the multicollinearity test. The tolerance (TOL) and variance inflation factor (VIF) are used to conduct a multicollinearity test on the above 5 indicator variables, and the tolerance is relatively close to 1, while the VIF is far less than 10, and the multicollinearity test is passed.

The second is the chi-square test. Chi-square test is performed on formula (3), and the chi-square value of the model is 64.799, and the significance level of the model is 0.000, indicating that the model is significant.

4. Construction and testing of a unified model

4.1. Construction of a unified model

Based on the frequency of each indicator in each year and the significance level of each indicator in each year of the model, 7 indicators, namely ROE, financial expense ratio, current assets ratio, NCF from financing activities/average total assets, cash and its equivalents increase/total assets, rigid liabilities ratio, and accounts payable turnover ratio, are selected as the indicator variables of the unified model. Adding these 7 indicators, the following expression is obtained according to SPSS21.0 regression results:

$$\ln \frac{P}{1-P} = -7.407 - 0.038 x_2 + 3.533 x_3 + 1.081 x_{37} - 6.959 x_{47} - 7.387 x_{14} + 3.321 x_{18} + 0.006 x_{45} \quad (4)$$

Among them, x_2 is return on equity ROE, x_3 is finance expense ratio, x_{37} is current assets ratio, x_{47} is NCF from financing activities/average total assets, x_{14} is increase in cash and its equivalents/total assets, x_{18} is rigid debt ratio, x_{45} is accounts payable turnover ratio.

4.2. Tests of the unified model

The first is the multicollinearity test. The tolerance (TOL) and variance inflation factor (VIF) are used to conduct the multicollinearity test on the above seven indicator variables, and the tolerance is large and close to 1, while the VIF is much less than 10, passing the test of multicollinearity.

This is followed by the chi-square test. The chi-square test of Equation (4) shows that the chi-square value of the model is 203.164, and the significance level of the model is 0.000, indicating that the model is significant.

Finally, the robustness test. To test the robustness of the unified logistic model, we consider choosing the same kind of indicators that have already existed in the original model, replacing the indicators in the original model in turn, and testing whether they still have predictive ability or whether their predictive ability is basically unchanged. The empirical results show that the difference between the overall correct rate of the substituted model and the correct rate of the original model is between 1 percentage point and 3 percentage points, which proves that the predictive ability of the model is basically unchanged.

5. Predictive ability of the unified model and the comparison with other models

The test sample used in this paper is from July 1, 2019 to December 31, 2019.

The logistic model proposed by Wu Shinong and Lu Xianyi(2001) (hereafter referred to as "Logistic Model 2") uses the earnings growth index, return on assets, current ratio, long-term debt to shareholders' equity ratio, working capital to total assets ratio, and asset turnover ratio to predict corporate financial distress. This paper compares equation (4) with Logistic Model 2 to see the accuracy of their predictions for the test sample.

The results are as follows:

Table 1. Comparison of the predictions of the unified Logistic model and the Logistic Model 2

Model	Test Sample		Predictions			Overall Arithmetic Average correct rate
			Whether or not bond default occurs		Overall	
			0 (normal business)	1 (bond default)		
Unified Logistic Model	amount	0 (normal business)	874	346	1220	76.45%
		1 (bond default)	3	13	16	
	percentage	0 (normal business)	71.64%	28.36%	100%	
		1 (bond default)	18.75%	81.25%	100%	
Logistic Model 2	amount	0 (normal business)	1098	122	1220	63.75%
		1 (bond default)	10	6	71	
	percentage	0 (normal business)	90%	10%	100%	
		1 (bond default)	62.5%	37.5%	100%	

As can be seen from Table 1, the unified Logistic model proposed in this paper also has relatively good predictive ability for the test sample, and the prediction accuracy rate for the the issuer of defaulted bonds is high. It shows that the unified Logistic model still has good prediction ability in application field.

In contrast, the Logistic model proposed by Wu Shinong and Lu Xianyi (2001) has only about 60% prediction accuracy rate, and the prediction accuracy rate of the issuer of defaulted bonds is low. Compared with the two models, the unified Logistic model proposed in this paper is better.

6. Limitations and future prospects

The model is based on the existing sample data. If the model is really applied to the actual prediction, it may face many uncertainties in the future. In addition, there may be some limitations in the choice of model and indicator variables selections.

In the future, we can consider including more non-financial indicators and financial indicators, and attempts could be made to compare more other models and select the one with the best predictive ability. In addition, there is still a need to continuously improve the model to enhance its predictive ability, stability and flexibility.

References

- [1] Wu Shinong, Lu Xianyi. Research on the prediction model of financial distress of listed companies in China[J]. *Journal of Economic Research*, 2001, 6: 46-53.
- [2] Altman E I. Financial Ratio, Discriminant Analysis and the Prdicion of Corporate Bankruptcy [J]. *Journal of Finance*, 1968, September: 589-609.
- [3] Merton R C. On the Pricing of Corporate Debt: the Risk Structure of Interest Rates [J]. *Journal of Finance*, 1974, 29(2): 449-470.
- [4] Guan Xin, Wang Zheng. Comparison of financial early warning models based on Logistic regression and BP neural network[J]. *Statistics and Decision Making*, 2016, 17: 179-181.
- [5] Yan Juanjuan, Sun Hongmei, Liu Jinhua. Support vector machine's early warning model for financial crisis of listed companies[J]. *Statistics and Decision Making*, 2006, 12: 158-159.

- [6] Wu Shinong, Chen Zhiyu, Wu Yuhui. Can machine learning early warning models predict bond default risk more effectively? *Working Paper*, School of Management, Xiamen University, August 2021.